Exh. 1002

Proceedings of the Second International Conference on *Autonomous Agents*

Minneapolis/St. Paul, MN USA
May 9-13, 1998

Edited by Katia P. Sycara and Michael Wooldridge

Sponsored by ACM SIGART.

Co-sponsored by ACM/SIGGRAPH and ACM/SIGCHI, with support from Microsoft, Nokia, Sun Microsystems, British Telecom, Mitsubishi Electric ITA (MERL and Horizon Systems Laboratory), SONY, AT&T, and the University of Minnesota. Held in cooperation with AAAI.

The Association for Computing Machinery
1515 Broadway
New York New York 10036

ACM ISBN: 0-89791-983-1

Additional copies may be ordered prepaid from:

ACM Order Department
PO Box 12114
Church Street Station
New York, NY 10257

Phone: 1-800-342-6626
(US and Canada)
+1-212-626-0500
(all other countries)
Fax: +1-212-944-1318
E-mail: acmpubs@acm.org

ACM Order Number: 605981
Printed in the USA

CONTENTS

PAPER SESSION 5: WWW AGENTS I

Monday, May 11 2:00 pm - 3:30 pm

116 **CiteSeer: An Autonomous Web Agent for Automatic Retrieval and Identification of Interesting Publications**
Kurt D. Bollacker, *University of Texas at Austin and NEC Research Institute*
Steve Lawrence, *NEC Research Institute*
C. Lee Giles, *NEC Research Institute and University of Maryland*

124 **Query Restart Strategies for Web Agents**
Prasad Chalasani, *Carnegie Mellon University*
Somesh Jha, *Carnegie Mellon University*
Onn Shehory, *Carnegie Mellon University*
Katia Sycara, *Carnegie Mellon University*

132 **WebMate: A Personal Agent for Browsing and Searching**
Liren Chen, *The Robotics Institute, Carnegie Mellon University*
Katia Sycara, *The Robotics Institute, Carnegie Mellon University*

PANEL SESSION

Monday, May 11 2:00 pm - 3:30 pm
Session Chair: Andrew Stern, *PF. Magic*

140 **AUTONOMOUS AGENTS IN ENTERTAINMENT**
Joe Bates, *Carnegie Mellon University*
Bruce Blumberg, *MIT Media Lab*
Athomas Goldberg, *NYU Media Research Laboratory*
Barbara Hayes-Roth, *Stanford University*
Andrew Stem, *PF. Magic*

PAPER SESSION 6: AGENTS FOR INFORMATION MANAGEMENT I

Monday, May 11 4:00 pm - 5:30 pm

141 **Concept Features in Re: Ageot, an Intelligent Email Agent**
Gary Boone, *Georgia Institute of Technology*

149 **Agents for Collecting Application Usage Data Over the Internet**
David M. Hilbert , *University of California, Irvine*
David F. Redmiles, *University of California, Irvine*

157 **Adaptive Information Agents in Distributed Textual Environments**
Filippo Menczer, *University of California, San Diego*
Richard K. Belew, *University of California, San Diego*

PAPER SESSION 7: SYNTHETIC AGENTS

Monday, May 11 4:00 pm - 5:30 pm

165 **A Social-Psychological Model for Synthetic Actors**
Daniel Rousseau, *Machina Sapiens*
Barbara Hayes-Roth, *Stanford University*

173 **Agents in Aonotated Worlds**
Patrick Doyle, *Stanford University*
Barbara Hayes-Roth, *Stanford University*

181 ***Story-morphing* in the Affective Reasoning Paradigm: Generating Stories Semi-Automatically for Use with "Emotionally Intelligent" Multimedia Agents**
Clark Elliott; *DePaul University*
Jacek Brzezinski, *DePaul University*
Sanjay Sheth, *DePaul University*
Robert Salvatoriello, *DePaul University*

PANEL SESSION

Monday, May 11 4:00 pm - 5:30 pm
Session Chair: Hyacinth S. Nwana, *BT Labs*

189 **AGENT-MEDIATED ELECTRONIC COMMERCE: ISSUES, CHALLENGES AND SOME VIEWPOINTS**
Hyacinth S. Nwana, *BT Labs*
Jeff Rosenschein, *Hebrew University*
Tuomas Sandholm, *Washington University*
Carles Sierra, *IIIA*
Pattie Maes, *MIT Media Laboratory*
Rob Guttmann, *MIT Media Laboratory*

PAPER SESSION 8: MOBILE SOFTWARE AGENTS

Tuesday, May 12 10:30 am - 12:30 pm

197 **Market-Based Resource Control for Mobile Agents**
Jonathan Bredin, *Dartmouth College*
David Kotz, *Dartmouth College*
Daniela Rus, *Dartmouth College*

205 **Mobile Agents for Enabling Mobile User Aware Applications**
Akhil Sahai, *INRIA-IRISA*
Christine Morin, *INRIA-IRISA*

212 **A Secure Marketplace for Mobile Java Agents**
Kay Neuenhofen, *Sun Microsystems*
Matthew Thompson, *Sun Microsystems*

219 **Mobile Agents on the Digital Battlefield**
Martin O. Hofmann, *Lockheed Martin ATL*
Amy McGovern, *University of Massachusetts*
Kenneth R. Whitebread, *Lockheed Martin ATL*

PAPER SESSION 9: LEARNING AND ADAPTATION I

Tuesday, May 12 10:30 am - 12:30 pm

226 A Multiagent Perspective of Parallel and Distributed Machine Learning
Gerhard Weiß, *Technische Universität München*

231 Learning Situation-Dependent Costs: Improving Planning from Probabilistic Robot Execution
Karen Zita Haigh, *Carnegie Mellon University*
Manuela M. Veloso, *Carnegie Mellon University*

239 Online Learning about Other Agents in a Dynamic Multiagent System
Junling Hu, *University of Michigan*
Michael P. Wellman, *University of Michigan*

247 Learning What Is Relevant to the Effects of Actions for a Mobile Robot
Matthew D. Schmill, *University of Massachusetts*
Michael T. Rosenstein, *University of Massachusetts*
Paul R. Cohen, *University of Massachusetts*
Paul Utgoff, *University of Massachusetts*

PAPER SESSION 10: LIFELIKE AGENTS

Tuesday, May 12 10:30 am - 12:30 pm

254 Behavioural Self-Organization in Lifelike Agents
Jiming Liu, *Hong Kong Baptist University*
Hong Qin, *Hong Kong Baptist University*

261 Integrating Reactive and Scripted Behaviors in a Life-Like Presentation Agent
Elisabeth André, *German Research Center for Artificial Intelligence (DFKI)*
Thomas Rist, *German Research Center for Artificial Intelligence (DFKI)*
Jochen Müller, *German Research Center for Artificial Intelligence (DFKI)*

269 BodyChat: Autonomous Communicative Behaviors in Avatars
Hannes Högni Vilhjálmsson, *MIT Media Laboratory*
Justine Cassell, *MIT Media Laboratory*

277 Actor-Based Computing: Vision Forestalled, Vision Fulfilled
Arthur Allen, *Charis Software Systems*

PAPER SESSION 11: COMPUTATIONAL MARKET SYSTEMS

Tuesday, May 12 2:00 pm - 3:30 pm

285 A Market Architecture for Multi-Agent Contracting
John Collins, *University of Minnesota*
Ben Youngdahl, *University of Minnesota*
Scott Jamison, *University of Minnesota*
Bamshad Mobasher, *University of Minnesota*
Maria Gini, *University of Minnesota*

293 Competitive Scenarios for Heterogeneous Trading Agents
Juan A. Rodríguez-Aguilar, *IIIA, CSIC*
Francisco J. Martín, *IIIA, CSIC*
Pablo Noriega, *IIIA, CSIC*
Pere Garcia, *IIIA, CSIC*
Carles Sierra, *IIIA, CSIC*

301 The Michigan Internet AuctionBot: A Configurable Auction Server for Human and Software Agents
Peter R. Wurman, *University of Michigan*
Michael P. Wellman, *University of Michigan*
William E. Walsh, *University of Michigan*

PAPER SESSION 12: AGENT CONTROL ARCHITECTURES I

Tuesday, May 12 2:00 pm - 3:30 pm

309 Increasing Agent Autonomy in Dynamic Environments
Subrata Das, *Charles River Analytics*
Alper Caglayan, *Charles River Analytics*
Paul Gonsalves, Charles River Analytics

317 Movement Control Methods for Complex, Dynamically Simulated Agents: Adonis Dances the Macarena
Maja J. Mataric, *University of Southern California*
Victor B. Zordan, *Georgia Institute of Technology*
Zachary Mason, *Brandeis University*

325 Integrating Active Perception with an Autonomous Robot Architecture
Glenn Wasson, *University of Virgina*
David Kortenkamp, *Metrica, Inc., TRACLabs*
Eric Huber, Metrica, *Inc., TRACLabs*

PAPER SESSION 16: WWW AGENTS II

Wednesday, May 13 10:30 am - 12:00 pm

400 **A Web-based Information System that Reasons with Structured Collections of Text**
William W. Cohen, *AT&T Labs-Research*

408 **WebACE: A Web Agent for Document Categorization and Exploration**
Eui-Hong (Sam) Han, *University of Minnesota*
Daniel Boley, *University of Minnesota*
Maria Gini, *University of Minnesota*
Robert Gross, *University of Minnesota*
Kyle Hastings, *University of Minnesota*
George Karypis, *University of Minnesota*
Vipin Kumar, *University of Minnesota*
Bamshad Mobasher, *University of Minnesota*
Jerome Moore, *University of Minnesota*

416 **SHOPPER'S EYE: Using Location-based Filtering for a Shopping Agent in the Physical World**
Andrew E. Fano, *Andersen Consulting*

PAPER SESSION 17: LEARNING AND ADAPTATION II

Wednesday, May 13 1:30 pm - 2:30 pm

422 **A History-Based Approach for Adaptive Robot Behavior in Dynamic Environments**
François Michaud, *Université de Sherbrooke*
Maja J. Mataric, *University of Southern California*

430 **Effects of Misconception on Reciprocative Agents**
Sandip Sen, *The University of Tulsa*
Anish Biswas, *The University of Tulsa*

PAPER SESSION 18: APPLICATIONS

Wednesday, May 13 1:30 pm - 2:30 pm

436 **SANI: A Seamless and Non-Intrusive Framework and Agent for Creating Intelligent Interactive Homes**
Sandeep Chatterjee, *Massachusetts Institute of Technology*

441 **Reactive Agents for Adaptive Image Analysis**
Jiming Liu, *Hong Kong Baptist University*

POSTER SESSION

Wednesday, May 13 1:30 pm - 3:00 pm

449 **Artificial Autonomous Agents with Artificial Emotions**
Luís Miguel Botelho, *ISCTE*
Helder Coelho, *University of Lisbon*

451 **Mobility and Management Support for Mobile Agents**
Wen-Shyen E. Chen, *National Chung-Hsing University*
S. T. Su, *National Chung-Hsing University*
Yao-Nan Lien, *National Chengchi University*
H. T. Shu, *Institute for Information Industry*
Huiling Liu, *Institute for Information Industry*

453 **Dynamic Software Agents for Business Intelligence Applications**
Qiming Chen, *HP Laboratories*
Parvathi Chundi, *HP Laboratories*
Umeshwar Dayal, *HP Laboratories*
Meichun Hsu, *HP Laboratories*

455 **RoboTA: An Agent Colony Architecture for Supporting Education**
Kenneth D. Forbus, *Northwestern University*
Sven E. Kuehne, *Northwestern University*

457 **A Formal Treatment of Distributed Matchmaking**
Somesh Jha, *Carnegie Mellon University*
Prasad Chalasani, *Carnegie Mellon University*
Onn Shehory, *Carnegie Mellon University*
Katia Sycara, *Carnegie Mellon University*

459 **A Synergy of Agent Components: Social Comparison for Failure Detection**
Gal A. Kaminka, *USC/Information Sciences Institute*
Milind Tambe, *USC/Information Sciences Institute*

461 **Subjective Avatars**
Michael Mateas, *Carnegie Mellon University*

463 **Embodied Cultural Agents and Synthetic Sociology**
Simon Penny, *Carnegie Mellon University*

465 **The Multi-Agent-Based Schedule Calculator (MASC) System**
Aruna Sankaranarayanan, *Fidelity Investments*
Maja Mataric, *University of Southern California*

467 **ULTIMA RATIO – Should Hamlet Kill Claudius?**
Michael Schroeder, *University of Hannover*
Daniela A. Plewe, *Franz-Künstler-Str*
Andreas Raab, *University of Magdeburg*

INTRODUCTION

Autonomous agents are computer systems that are capable of independent action in dynamic, unpredictable environments. Agents are also one of the most important and exciting areas of research and development in computer science today. Agents are currently being applied in domains as diverse as computer games and interactive cinema, information retrieval and filtering, user interface design, and industrial process control. The aim of the Agents '98 conference is to bring together researchers and developers from industry and academia in order to report on the latest scientific and technical advances, discuss and debate the major issues, and showcase the latest systems.

The First International Conference on Autonomous Agents (Agents '97) was held in Marina del Rey, California, in February 1997. It was attended by nearly 500 people, and received media coverage from such varied and widely-respected organizations as *Wired* magazine, the *New York Times*, and CNN. It was generally reckoned to have created something of a stir far beyond the audience that the organizers originally expected. All this made Agents '97 a hard act to follow — but we believe that we have succeeded in Agents '98.

It is only a year since the first Autonomous Agents conference, and yet in that time, agent technology has come a long way. At Agents '97, delegates were talking about the *possibility* of commercializing agent technology; of using agents in "real" systems. In just one year, we might have expected to see a few tentative efforts in this direction. But to our pleasure and surprise, we have seen agent technology adopted not just by a few research projects, but by nearly all major players in the commercial software marketplace. Agents are now an everyday component of software, with agent-enabled features rapidly becoming accepted as the norm, rather than as the exception.

Autonomous Agents '98 is a vivid illustration of the latest developments in agent technology. Like its predecessor, it is focused around three main strands:

- Software agents, which are situated in a software environment, and typically act as "expert assistants" to users carrying out some task.
- Robotic agents, which are physically embodied autonomous robots, sensing and acting in the everyday physical world.
- Synthetic agents, which inhabit shared virtual environments, often in the form of computer games, virtual theater, or interactive cinema.

Nearly 180 technical papers were submitted to the conference, and all were rigorously reviewed by the program committee. Of these submissions, only 57 were accepted as full technical papers. This high rejection rate is more a reflection of the care and thought that the program committee and area chairs put into the review and selection process than the standard of papers submitted. The overall outcome of the review process is a selection of papers, videos, and software and hardware demonstrations that showcase the very best of agent technology today.

We are confident that Agents '98 will confirm the Autonomous Agents series of conferences as a key forum for presenting work in the applications of agent technology.

Acknowledgments

We would like to take this opportunity to thank everyone involved with the organisation of Agents '98. First, we would like to thank Tim Finin and the area chairs, for their hard work and first-rate scientific evaluation of papers and videos submitted. Dan Weld must be singled out for handling no less than 125 papers on software agents — somewhat more than anyone anticipated! Maja Mataric and Clark Elliott did excellent jobs of handling papers and videos in the robotics and synthetic agents/agents for entertainment areas. Keith Decker handled publicity for the conference, which involved (amongst other things) changes to the conference WWW site seemingly every hour. Milind Tambe handled the finances for the conference, and Maria Gini did a superb job of local organization. Mike Huhns handled the organization of workshops, an innovation of this year's conference, which added significantly to the richness of the conference program. Anand Rao was tutorial chair, and Henry Kautz and Robin Murphy handled the demonstration sessions, of software agents and robotic agents respectively. Afsaneh Haddadi managed the poster sessions for the conference, and David Musliner the exhibits. Bamshad Mobasher handled registrations on behalf of the conference. The program committee did a typically thorough and conscientious job of reviewing a very large number of papers.

We would like to thank the staff of AAAI, and in particular Carol Hamilton and Keri Vasser, for handling the submissions to the conference, putting together the proceedings so professionally, and generally giving excellent advice on organizational matters of all kinds. From ACM, we would like to extend our gratitude to Alisa Rivkin for her help with the proceedings. We would also like to thank ACM SIGART and ACM in general for their continued support, which has been of enormous benefit in so quickly establishing Autonomous Agents a major event in the conference calendar, and in addition our other sponsors, without whose support Agents '98 would have been significantly less interesting. The enthusiastic support of so many sponsors is a good indicator of how seriously the world is taking agent technology.

Finally, we would like to extend our thanks to Lewis Johnson for his sound and timely advice, and the Autonomous Agents steering committee for their helpful suggestions at many points throughout the long (and often exhausting) process of conference organization.

Katia P. Sycara (Carnegie Mellon University, USA) and

Michael Wooldridge (Queen Mary & Westfield College, UK)

March 1998

WebMate : A Personal Agent for Browsing and Searching*

Liren Chen and Katia Sycara
The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA. 15213
lchen@cs.cmu.edu, katia@cs.cmu.edu

Abstract

The World-Wide Web is developing very fast. Currently, finding useful information on the Web is a time consuming process. In this paper, we present WebMate, an agent that helps users to effectively browse and search the Web. WebMate extends the state of the art in Web-based information retrieval in many ways. First, it uses multiple TF-IDF vectors to keep track of user interests in different domains. These domains are automatically learned by WebMate. Second, WebMate uses the Trigger Pair Model to automatically extract keywords for refining document search. Third, during search, the user can provide multiple pages as similarity/relevance guidance for the search. The system extracts and combines relevant keywords from these relevant pages and uses them for keyword refinement. Using these techniques, WebMate provides effective browsing and searching help and also compiles and sends to users personal newspaper by automatically spiding news sources. We have experimentally evaluated the performance of the system.

Area: Software Agents

Keywords: Information Agents, Instructability, Knowledge acquisition and accumulation, long-term adaptation and learning, user modeling

1 Introduction

The Web is full of information and resources. People have at least three ways to find information they need: (1) by browsing (following hyper-links that seem of interest to them), (2) by sending a query to a search engine, such as Altavista, (3) by following existing categories in search engines, such as Yahoo or Lycos. The problem is that people have to spend a lot of time and effort to navigate but may not find interesting personalized information. However, it is difficult to find the wanted information because a user can't accurately express what he wants and search engines don't adapt their search strategies according to different users. Moreover, the problem is exacerbated because the information sources have high "noise", i.e. most of the pages are irrelevant to a particular user's interests. Intelligent software agents are being developed to deal with these issues.

Intelligent agents are programs that act on behalf of their human users to perform laborious information-gathering tasks [1] and they are one of the "hot" topics in Information Systems R&D at the moment. The last ten years have seen a marked interest in agent-oriented technology, spanning applications as diverse as information retrieval, user interface design and network management.

In this paper, we present WebMate, a personal software agent that accompanies a user when he browses and searches and provides intelligent help [1].

For clarity of presentation, the WebMate capabilities will be presented in roughly two categories: (1) learning user interests incrementally and with continuous update and automatically providing documents (e.g. a personalized newspaper) that match the user interests, and (2) helping the user refine search so as to increase retrieval of relevant documents. In section 2, we describe the architecture of the system. The WebMate acts as a proxy and monitors a user's actions. In section 3, we describe the user profile representation and learning algorithm [3, 4]. In addition, we provide experimental results of compiling a personal newspaper. In section 4, we discuss how to use the Trigger Pairs Model to extract relevant words to use as keyword refinements to improve search. We also present utilizing relevance feedback [8] during search to dynamically enhance the search for relevant documents. Finally, related work and our future work are described.

*This research has been supported in part by ARPA contract F33615-93-1-1330, and by ONR Grant N00014-96-1222.

[1]The WebMate system has been operating on Web and has been downloaded by more than 600 users since it was published in the middle of September 1997 (15 days ago). Its URL is http://www.cs.cmu.edu/~softagents/webmate.

Autonomous Agents 98 Minneapolis MN USA
Copyright 1998 0-89791-983-1/98/ 5...$5.00

2 WebMate architecture

WebMate is composed of a stand-alone proxy that can monitor a user's actions to provide information for learning and search refinement, and an applet controller that interacts with a user (See Figure 1).

The stand-alone proxy is an HTTP proxy that sits between a user's web browser and the World-Wide Web. All HTTP transactions pass through WebMate which can monitor a user's browsing and searching activities and learn from them.

The applet controller is the interface between the user and the stand-alone proxy. Through it, the user can express his interests when he browses and provide relevance feedback when he searches. In addition, through the applet controller, the user receives intelligent help from WebMate.

3 Learning profile to compile personal newspaper

3.1 Profile Representation and Learning Algorithm

There are several machine learning approaches that can be used to learn a user profile, such as Bayesian classifier, Nearest Neighbor, PEBLS, Decision Trees, TF-IDF, Neural Nets [4, 5]. In order for a particular technique to be effective, it should match the characteristics of the task and the user.

The filtering task for our agent involves judging whether an article is relevant or irrelevant to the user based on the user profile, in an environment where the prior probability of encountering a relevant document is very low compared to the probability of encountering an irrelevant document. In such an environment, it would be very frustrating and time consuming for a user to interact with an agent that starts with no knowledge but must obtain a set of positive and negative examples from user feedback. When a user browses, he does not want to evaluate all web pages that might contain potentially interesting information. To reduce user evaluation burden, WebMate collects only examples that are interesting to the user (only positive training examples). This kind of interaction presents potential problems since the documents that a user might label as "I like It" might fall into many distinct domains (e.g fishing, computer science, soccer). Those subclasses correspond to the different interests a user has. There have been two methods to address the problem of multiple user interests. The first is to keep a single user profile where the keywords might come from different domains but are "averaged'. This method has the disadvantage that averaging the vectors from the different documents might decrease too much the weights of words that are important for only a few of the interest categories. The second method is to ask the user to explicitly provide labels for the sub-categories of interest. WebMate does not ask the user to label the category that the interesting document is in, but learns the categories automatically.

In contrast to other systems that learn a user profile and use it statically to determine relevant documents, WebMate learns the user profile incrementally and continuously. When a new positive example is known, the system updates the profile. In order to save on storage space, the system doesn't keep any of the previous positive example documents. It only keeps the profile learned from those positive examples. In this way, the system will adapt to the user's evolving and recent interests.

WebMate utilizes TF-IDF method [7] with multiple vectors representation. The basic idea of the algorithm is to represent each document as a vector in a vector space so that documents with similar content have similar vectors. Each dimension of the vector space represents a word and its weight. The values of the vector elements for a document are calculated as a combination of the statistics term frequency $TF(w,d)$ (the number of times word w occurs in document d) and document frequency $DF(w)$ (the number of documents the word w occurs in at least once). From the document frequency the inverse document frequency $IDF(w)$ can be calculated.

$$IDF(w) = \log \frac{|D|}{DF(w)}$$

$|D|$ is the total number of documents. The value $d^{(i)}$ of an element in the vector is then calculated as the product

$$d^{(i)} = TF(w_i, d) \times IDF(w_i)$$

We have developed an algorithm for *multi TF-IDF* vector learning. The algorithm follows.

We assume that a user has at most N domains of interest. [2] Assume the initial profile set is V, $|V| = 0$; the predefined number of TF-IDF vectors in the profile set is N, the preset number of elements of a vector is M. For each positive example (i.e. an HTML documents that the user has marked "I like It"), do:

1. Preprocess: parse HTML page, deleting the *stop* words (or non-informative words) such as "a", "the", "is", "in", etc, stemming the plural noun to its single form and inflexed verb to its original form, extracting the words in *title*(<TITLE>), *head1*(<H1>), *head2*(<H2>), *head3*(<H3>) because they will be given more weights;
2. Extract the TF-IDF vector for this document, let it be V_i;
3. If $|V| < N$ ($|V|$ is the number of vectors in the profile set V), then $V \Leftarrow V \cup V_i$;
4. Otherwise, calculate the cosine similarity between every two TF-IDF vectors including the vectors in the

[2] In the current implementation, N is heuristically set to 10

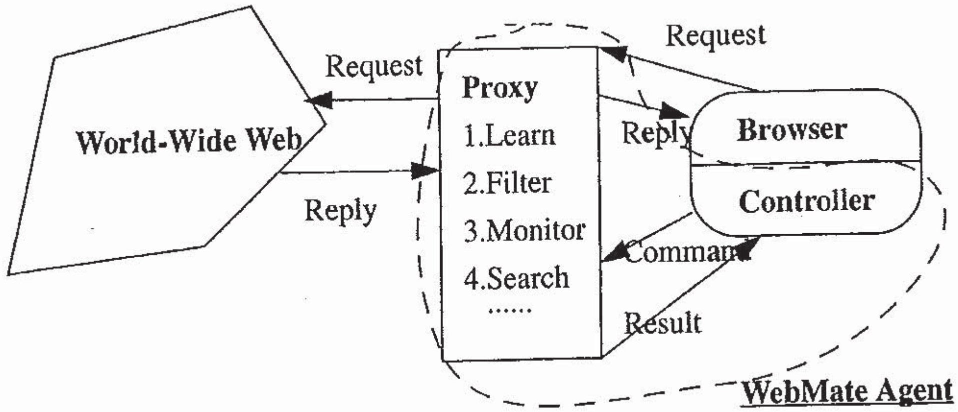


Figure 1: System Architecture

profile set V and the new document vector V_i. Assume the profile set V is $\{V_1, V_2, \ldots, V_n\}(n = N)$.

$$Sim(V_j, V_k) = \frac{V_j \bullet V_k}{|V_j| \times |V_k|} \quad j, k \in \{1, 2, \ldots, n, i\}$$

5. Combine the two vectors V_l and V_m with the greatest similarity..

$$V_l = V_l + V_m \quad (l, m) = \arg\max_{(x,y)}(Sim(V_x, V_y))$$

6. Sort the weights in the new vector V_k in decreasing order and keep the highest M elements.

This algorithm is run whenever a user marks a document as "I like it". Thus, the user profile is incrementally, unobtrusively and continuously updated.

3.2 Compiling personal newspaper

We utilize the approach of learning user profile to compile a personal newspaper [9, 10, 11]. We do this in two ways.

One way is to automatically spide a list of URLs that the user wants monitored. An example of such a URL is one that consists of many news headlines like the home page of the NewsLinx Company[3]. WebMate (1) parses the html page , (2) extracts the links of each headline, (3) fetches those pages, (4) constructs the TF-IDF vector for each of those pages (using as additional heuristics that words in title, and headings are given additional weights), and (5) calculates the similarity with the current profile. If the similarity is greater than some threshold, it recommends the page to the user, and sorts all the recommended pages in decreasing order of similarity to form the personal newspaper. All operations are often performed in the middle of the night when the network traffic is low. In the morning, the user can read the recommended personal newspaper.

If the user does not provide any URLs that he would like to be the information sources, WebMate constructs a query[4] using the top several words in the current profile and sends it to popular search engines (e.g. Altavista, Yahoo). If the result is needed immediately, the results returned by the search engines are directly used as the recommended web pages. Otherwise, the system fetches the pages corresponding to each and every URL in the results. It then calculates the similarity of the profile and these web pages and recommends the pages whose similarity is greater than some threshold presenting the results in descending order of relevance.

3.3 Experiments

In our experiments, the system monitors about 14 news sites that contain articles about high technology including LAN time news[4], Media Central [5], PC magazine online [6], etc. We recorded the personal newspaper and evaluated whether a piece of news is interesting to us (Table 1). The first column is the date of the personal news, the second column is the percentage accuracy of how many pieces of news are interesting in the top 10 returned by WebMate, the third column is the percentage accuracy in the top 20. In order to evaluate the learning approach, the percentage accuracy in the whole recommended news (the number of interesting news articles divided by the total number of news articles in the newspaper) is given in the fourth column.

From Table 1, we see that the average accuracy (relevance rate) that the recommended news is relevant to our

[3] http://www.newslinx.com/

[4] http://www.lantimes.com/

[5] http://www.mediacentral.com/Magazines/MediaDaily/Archive

[6] http://www8.zdnet.com/pcmag/

Date	Accuracy in top 10	Accuracy in top 20	Accuracy in whole
Sep.16	70%	60%	17/55=31%
Sep.17	40%	35%	11/42=26%
Sep.18	50%	35%	9/33=27%
Sep.19	60%	65%	18/76=24%
Sep.20	50%	40%	9/29=31%
Sep.22	40%	40%	12/49=25%
Sep.23	50%	50%	18/78=23%
Sep.24	60%	56%	10/18=56%
Average	52%	49%	30.4%

Table 1: Experiment Results

interests is between 50% and 60% in the top 10 news articles . Generally the system will spide more than 500 pieces of news for a day. In the whole recommended news, the average accuracy is about 30%. But if the news are randomly chosen from 500 pieces of news in which we assume there are 100 interesting news to us (this is based on our observation that for a typical news site such as LinkExchange, there are about 10 out of 50 pieces of news that are interesting to us in any given day), the default accuracy in the whole news is about 20%. So a 50% to 60% accuracy, achieved by WebMate, represents a two to three-fold accuracy increase.

There are several factors that lower the accuracy of the system. First, it is difficult to determine which links are the headlines of the news and which links are irrelevant stuff such as advertisements. We are currently working on heuristics to filter out advertisements. So, currently, all the links in the page are used to calculate the similarity, not just the links of the news headlines. Second, while calculating the TF-IDF vectors, the irrelevant stuff around the news affects the accuracy of the TF-IDF.

4 Search refinement by keywords expansion and relevance feedback

4.1 Trigger Pairs Model to extract relevant words

Single keywords are usually ambiguous, or too general. Moreover, they can occur in vast quantities of documents, thus making the search return hundreds of hits, most of which are irrelevant to the intended user query. Giving additional keywords can refine search providing considerable improvement in the retrieval results. Good refinement words must have meanings that help disambiguate or make more specific the original search word. For example, the word "stock" has more than 10 definition in the WordNet[7] including "the capital raised by a corporation through the issue of shares entitling holders to partial ownership", "gun-stock", "inventory", "stock certificate", etc. Providing the refinement words that correspond to each one of those meanings, would help a search engine, for example, to prune out documents where the word is used with any of its other meanings. There are three ways to expand the query: manual query expansion, semi-manual query expansion, and automatic query expansion [12]. No matter which method is used, the key point is to get the best refinement words. In manual query expansion, although the user knows the intended meaning of the keyword she is using, she may not be able to provide the best refinement words. "Best" here means refinement words that most frequently co-occur with the word in its intended meaning in large number of documents. In other words, one of the characteristics of good refinement words is that they be domain specific. In this section we present the method for automatically finding appropriate keywords to constrain and refine search for relevant documents.

We use the Trigger Pairs Model [13, 14]. If a word S is significantly correlated with another word T, then (S, T) is considered a "trigger pair", with S being the trigger and T the triggered word. When S occurs in the document, it triggers T, causing its probability estimate to change. That is, when we see the word S appearing at some point in a text, we expect the word T to appear somewhere after S with some confidence[8]. The mutual information (MI) that considers the words order is a measure of the correlation and used to extract trigger pairs from large corpus. The mutual information is given by the following formula:

$$\mathcal{MI}(s,t) = \mathcal{P}(s,t)\log\frac{\mathcal{P}(s,t)}{\mathcal{P}(s)\mathcal{P}(t)}$$

To evaluate the method, we used the Broadcast News Corpus of $140M$ words and set the maximum distance between S and T to 500. Some randomly selected trigger pairs which are sorted in decreasing order of the mutual information are shown.

product ← {maker, company, corporation, industry, incorporate, sale, computer, market, business, sell, machine, consumer, share, software, manufacture, electronic, base, million, manufacturer}

car ← {motor, auto, model, maker, vehicle, ford, buick, honda, inventory, assembly, chevrolet, sale, nissan, incentif, pontiac, plant, toyota, dealer, chrysler}

interest ← {rate, bank, loan, point, dollar, credit, bond, percent, investment, market, reserve, term, debt, investor, billion, exchange, higher, treasury, lower}

fare ← {airline, maxsaver, carrier, discount, air, coach, flight, traveler, travel, continental, unrestrict, ticket, texas, northwest, pettee, match}

music ← {musical, symphony, orchestra, composer, song, concert, tune, concerto, sound, musician, classical, album, violin, violinist, jazz, audience, conductor, play, audio, rock, cello, perform, dance}

[7] http://www.cogsci.princeton.edu/~wn/

[8] In the Trigger Pairs Model, (S, T) is different from (T, S), so the Trigger Pairs Model is different from the method of using co-occurrence of two words that is generally used in other keywords expansion experiments[12]

pork ← {meat, hog, slaughter, livestock, mercantile, cattle}
plead ← {guilty, sentence, insider, indictment, indict, ivan, charge, attorney, fraud, boesky, lasker, criminal, pleas, investigation, plea, court, prosecutor, prison, felony, defendant, cooperate, palmieri}

We also extracted trigger pairs from the Wall Street Journal Corpus of 1M words. We found that the trigger pairs are domain specific. For example, the triggers to "Stock" in news and media domain (Broadcast News Corpus, 140M tokens) are {company, bond, buy, business, bank, dow, earning, composite, cent, analyst, big, chrysler, investor, cash, average, economy, close, capital, chip, ...}. However, in business and Economic (Wall Street Journal Corpus, 1M tokens) the triggers are {share, investor, index, exchange, price, dow, market, buy, point, jone, trade, trader, average, cent, industrial, gain, shareholder, company, board, ...}

4.2 Keywords Expansion Algorithm

The trigger pair method can provide several candidate refinement keywords. An additional question is, how many and which ones to use under any given circumstances. extract relevant words from large corpus. For a search with only one keyword, the top several triggers to the keyword are used to expand the search. But for a search with more than 2 keywords, the choice becomes more complicated. We use the following algorithm for keywords expansion based on the trigger pairs:

Let us assume that the keywords are $K_1, K_2, \ldots, K_m$, and the the expected number of refinement words is N. Initialize $n = m$, $\mathcal{S}$ is the empty set.

1. $\mathcal{S}_1 = \{s_{11}, s_{12}, \ldots, s_{1i}\} \rightarrow K_1$, $\mathcal{S}_1$ is the triggers set to K_1. $s_{11}, s_{12}, \ldots, s_{1i}$ are sorted in decreasing order of the mutual information.

   $\mathcal{S}_2 = \{s_{21}s_{22}, \ldots, s_{2j}\} \rightarrow K_2$, $\mathcal{S}_2$ is the triggers set to K_2

   ...

   $\mathcal{S}_m = \{s_{m1}, s_{m2}, \ldots, s_{mk}\} \rightarrow K_m$, $\mathcal{S}_m$ is the triggers set to K_m

2. $\mathcal{S} = \mathcal{S} \cup (\forall(\mathcal{S}_p, \mathcal{S}_q, \ldots, \mathcal{S}_r)(\mathcal{S}_p \cap \mathcal{S}_q \cap \ldots \cap \mathcal{S}_r))$, and $(\mathcal{S}_p, \mathcal{S}_q, \ldots, \mathcal{S}_r)$ is one of the combinations of n sets out of m. The words in the $\mathcal{S}$ are sorted in decreasing order of mutual information.

3. If $|\mathcal{S}| \geq N$, let the top N words in the $\mathcal{S}$ be the refinement words and stop.

4. otherwise, let $n \Leftarrow n - 1$, goto 2.

This method can improve the recall rate of the search. For example, if a system uses TF-IDF to extract informative words to index documents, some K_i itself might be ignored because of its low weight. However, some words in $\mathcal{S}_i$ could be selected thus helping to recall documents where the ignored $K_(i)$ appears thus improving recall rate.

This method also provides disambiguation information for ambiguous query words. For example, K_1 = *charge* and $\mathcal{S}_1$ =*{federal, investigation, attorney, plead, indict, allege, fraud, guilty, indictment, jury, prosecutor, court, case, criminal, law, grand, commission, insider, conspiracy, ...}*, K_2 =*fee* and $\mathcal{S}_2$ = *{pay, dollar, million, bank, service, tax, raise, federal, bill, require, percent, charge, paid, law, client, loan, money, legal, payment, ...}*, then $\mathcal{K} = \{K_1, K_2\}$ = *{Charge, Fee}* and $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$ = *{million, pay, dollar, tax, service, federal, client, law, loan, legal, payment, court, suit, file, cost, case, company, firm, ...}*. So triggers, such as million, pay, dollar, tax and service, help confine and disambiguate the meaning of the word "charge".

4.3 Examples on keywords expansion

In this section, we present a typical example of how our refinement method indeed helps improve search results. Suppose the user is interested in documents where the word "stock" appears in its financial meaning. Inputting simply the keyword "stock" to Lycos and Altavista returns the following results.

From Lycos:

1) YOSEMITE STOCK PHOTOS, ROCK CLIMBING, Daniela Masetti PHOTOS
2) YOSEMITE STOCK PHOTOS, ROCK CLIMBING PHOTOS
3) YOSEMITE STOCK PHOTOS, FISHING PHOTO
*4) Stock information Java Applet
5) STOCK GRAPHICS & PHOTOS
*6) American Stock Transfer & Trust Home Page
*7) STOCK CHARTS
*8) GROWTH STOCK ADVISOR FULL DISCLAIMER
*9) Stock information Java Applet
10) Ocean Stock

Only 5 hits are relevant to the financial meaning of "stock" in the top 10.

From Altavista:

1. E. coli Genetic Stock Center
2. Michael Paras Photography: Photographs, Photography, stock photos,stock photo
*3. iGOLF Features - Stocks & Industry - Stock Report: Tuesday,September 5, 1995
4. Cedar Stock Resort Trinity Center Marina
*5. Stock 4 Art: HOME PAGE!
6. NET INFO - Luc Sala - Myster - stock footage
*7. The Official Vancouver Stock Exchange
*8. Stock Club
*9. NIAGARA MOHAWK DECLARES PREFERRED STOCK DIVIDEND
*10. The Italian Stock Exchange

There are 6 hits that are relevant to the financial meaning of the "stock" in the top 10.

At this time, it is difficult for a user to figure out what words should be used to expand or refine the current search. So the trigger pairs can be used to expand the current search. The triggers to "stock" are {share, investor, index, exchange, price, dow, market, buy, point, jone, trade, trader, average, cent, industrial, gain, shareholder, company, board, ...}. If we use the first word "share" in the ranked triggers list to expand the keyword "stock" and send {stock share} to the above two search engines, the following results get returned.

From Lycos:

*1) Share, Stock or CD Secured Loans

*2) Share / Stock Option Scheme Administration

*3) Allfinanz: Stock, Share Dealers

*4) One Share of Stock, Inc. - Ordering Info

*5) One Share of Stock - Product Line

*6) Akiko New Zealand: Stock And Share Market Links (12-Sep-1995)

*7) Akiko New Zealand: Stock And Share Market Links (12-Sep-1995)

*8) Money: $50 can buy share of stock in a company

*9) ONE SHARE OF STOCK - Order Form

*10) One Share of Stock, Inc. - Company Info

Those results are all relevent to the financial meaning of the word "stock".

From Altavista:

*1. South Africa: Stock market: Share price index (dissemination formats)

*2. Denmark: Stock market: Share price index (base page)

*3. ONE SHARE OF STOCK, INC.

*4. Chile: Stock market: Share price index (base page)

*5. Accounting financial software share stock market money portfolio bank mutual f

*6. Singapore: Stock market: Share price index (dissemination formats)

*7. Mexico: Stock market: Share price index (base page)

*8. Netherlands: Stock market: Share price index (base page)

*9. Ireland: Stock market: Share price index (dissemination formats)

*10. Japan: Stock market: Share price index (base page)

Those results are all relevent to the financial meaning of the word "stock".

We can see the results are better than before. We can also refine the search "stock share" if the results are not satisfactory. The intersection of the triggers sets of "stock" and "share" is {stake, outstanding, company, common, quarter, convertible, shareholder, cent, takeover, earning, exchange, incorporate, acquire, million, composite, dividend, percent, point}. Again we can use the words in this set to continue to expand the keywords "stock" and "share" by choosing one or more of them.

4.4 Relevance feedback

One of the most important ways in which current information retrieval technology supports refining searches is relevance feedback. Relevance feedback is a process where users identify relevant documents in an initial list of retrieved documents, and the system then creates a new query based on those sample relevant documents [14]. The idea is that since the newly formed query is based on documents that are similar to the desired relevant documents, the returned documents will indeed be similar. The central problems in relevance feedback are selecting "features" (words, phrases) from relevant documents and calculating weights for these features in the context of a new query [8].

In WebMate agent, the *context* of the search keywords in the "relevant" web pages is used to refine the search because we think that if a user tells the system some page is relevant to his searh, the context of the search keywords is more informative than the content of the page.

Given a relevant page, the system first looks for the keywords (assume K_i is one of the keywords) and context of the keywords (assume the context of the keyword K_i is $\ldots W_{-5}W_{-4}W_{-3}W_{-2}W_{-1}K_iW_1W_2W_3W_4W_5\ldots$). For each keyword $K_{(i)}$, the system then extracts the chunks of 5 words $W_{-5}W_{-4}W_{-3}W_{-2}W_{-1}$ before K_i and the chunks of 5 words $W_1W_2W_3W_4W_5$ after K_i until all the keywords in the query are processed.

Then, a bag of chunks are collected and passed to the processes of deleting the stop words and calculating the frequency. After that, the top several frequent words are used to expand the current search keywords.

For example, the following text is part of the overview of our Intelligent Agents project at CMU[9]. Suppose a user gives this text as a relevance feedback to the search keywords "intelligent agent".

> Intelligent Software Agents
>
> The voluminous and readily available information on the Internet has given rise to exploration of Intelligent Agent technology for accessing, filtering, evaluating and integrating information.
>
> In contrast to most current research that has investigated single-agent approaches, we are developing a collection of multiple agents that team up on demand—depending on the user, task, and situation—to access, filter and integrate information in support of user tasks. We are investigating techniques for developing distributed adaptive collections of information agents that coordinate to retrieve, filter and fuse information relevant to the user, task and situation, as well as anticipate user's information needs.
>
> Approach is based on:
>
> adaptable user and task models
>
> flexible organizational structuring
>
> a reusable agent architecture

[9] The URL of our project is: http//www.cs.cmu.edu/~softagents.

Underlying Technology
Our intra-agent architecture and inter-agent organization is based on the RETSINA multiagent reusable infrastructure that we are developing.

Using our method, the refinement words extracted from the text are {software, structure, reusable, architecture, technology, organizational, network, schedule, research, rise}. Most of the refinement words reflect well the characteristic of the project. But, if instead of using the context method, we considered the whole content of the page when calculating the frequency, then the expanding words would be {software, information, task, area, application, technology, user, current, develop, underlying}. Obviously, the context of the search keywords can reflect the relevance better than the whole content of the web page.

Subsequently, we used the top 5 words {software structure reusable architecture technology} to expand the search "intelligent agent". These are the results returned by Lycos. The content of links marked with "*" are similar to the content of the page given as the "relevant" feedback.

*1) The Agent Building Shell: Programming Cooperative Enterprise Agents
(http://www.ie.utoronto.ca/EIL/ABS-page/ABS-overvie)
*2) The Agent Building Shell: Programming Cooperative Enterprise Agents
(http://www.ie.utoronto.ca/EIL/ABS-page/ABS-overvie)
*3) An Architecture for Supporting Quasi-agent Entities in the WWW
(http://www.cs.umbc.edu/˜cikm/iia/submitted/viewing)
4) Knowledge Sharing Papers
(http://hpp.stanford.edu/knowledge-sharing/papers/R)
5) Knowledge Sharing Papers
(http://hpp.stanford.edu/knowledge-sharing/papers/i)
6) Knowledge Sharing Papers
(http://ksl.stanford.edu/knowledge-sharing/papers/i)
*7) The Agent Building Shell: Programming Cooperative
(http://www.ie.utoronto.ca/EIL/ABS-page/ABS-intro.h)
*8) Special Issue AI in Medicine Editorial Special Issue Artificial Intelligence in Medicine "Architectures for Intelligent Systems Based on Reusable Components"
(http://www.swi.psy.uva.nl/usr/Schreiber/papers/Mu)
*9) CS 791A – Agent Architectures for Information Gathering
(http://centaurus.cs.umass.edu/ig-seminar.html)
*10) Interaction Protocols for Software Agents on the World Wide Web
(http://rbse.jsc.nasa.gov/eichmann/www-s96/interact)

5 Related work

WebWatcher [10][16] is a tour guide for the web. It learns from experiences of *multiple users* to improve its advice-giving skills. Letizia [17] can recommend nearby pages by doing lookahead search. Syskill & Webert [4] is a software agent that learns to rate pages on the Web, deciding which pages might interest a user. Lira [3] works offline and returns a set of pages that match the user's interest. Daily Briefing [11] allows you to use Autonomy Intelligent Agents as Newshounds to sniff out stories and compile a personal daily newspaper with stories, features and articles selected from the Internet to match your requirements. WBI [12] is a personal web agent designed to personalize your web browsing. Metabot [13] is a Java-based, client-server application for searching the web by performing a simultaneous query on multiple web search services. CoolURL [14] is an exploratory technology that enables users to use agent technology to recommend cool URLs to a community of users. Beehive [18] is a distributed system for social sharing and filtering of information. Firefly [15] uses software agents that automate the process of retrieving data from the Web based on what they know about their owner's tastes and interests. Their core technology is the social filtering (or colaborative filtering). WiseWire [16] uses advanced neural net technology and adaptive collaborative filtering to filter all types of digital content that is personally relevant to you.

6 Summary and Future Research

WebMate is a personal agent running on the end user machine. It accompanies users from page to page to provide assistance. It can learn the user profile and compile personal newspaper, help the user improve the search by keyword expansion and relevance feedback, and aid the user in other ways such as alias, reference, prefetch, and monitor bookmarks or web pages for changes.

Currently in WebMate, only words are used to represent a user's profile. We feel that new machine learning algothrims for classifying the new web pages are necessary to improve the accuracy of the recommendation. We are currently implementing phrases, bigram [13] of words and plan to explore the trigger pairs or relevant words to improve the learning. In addition, we are implementing heuristics to filter out advertisements and irrelevant content around web pages comtaining news.

[10] http://www.cs.cmu.edu/Groups/webwatcher/
[11] http://www.agentware.com/main/dailyme.html
[12] http://www.networking.ibm.com/iag/iaghome.html
[13] http://metabot.kinetoscope.com/docs/docs.html
[14] http://support.intel.com/oem-developer/internet/coolurl/COOL.FAQ.HTM
[15] http://www.firefly.com/
[16] http://www.wisewire.com/

References

[1] Katia Sycara, Anandeep Pannu, Mike Williamson, Dajun Zeng, Keih Decker. *1996, Distributed Intelligent Agents*. Published in IEEE Expert, Intelligent Systems & their applications, Dec, 1996

[2] Shaw Green, Leon Hurst, Brenda Nangle, Padraig Cunninggham, Fergal Somers, Richard Evans. *1997, Software Agents: A review*. http://www.cs.tcd.id/Brenda.Nangle/iag.html

[3] Marko Balabanovic, Yoav Shaham. *1995, Learning Information Retrieval Agents: Experiments with Automated Web Browsing*. Proceedings of the AAAI Spring Symposium Series on Information Gathering from Heterogeneous, Distributed Environments: 13-18.

[4] M. Pazzani, J. Muramatsu, D. Billsus. *1996, Syskill & webert: Identifying interesting web sites*. In AAAI conference, Portland, 1996

[5] Pannu, A. and Sycara, K. *1996, A Personal Text Filtering Agent*. Proceedings of the AAAI Stanford Spring Symposium on Machine Learning and Information Access, Stanford, CA, March 25-27, 1996.

[6] K.Lang. *1995, NewsWeeder: Learning to filter Netnews*. Proceedings of Machine Learning, Morgan Kaufman, San Francisco, 1995

[7] G.Salton and M.J.McGill. *1983, Introduction to Modern Information Retrieval*. McGraw-Hill, New York, 1983

[8] Gerard Salton, Chris Buckley. *1988, Improving Retrieval Performance by Relevance Feedback*. Cornell University, 88-898

[9] Marbo Balabonovic, Yoav Sholiam. *1997, Combining Content-Based and Collaborative Recommendation*. Communications of the ACM, March, 1997

[10] Peter W. Foltz, Susan T. Dumais. *1992, Personalized Information Delivery: An Analysis of Information Filtering Methods*. Published in Comminucations of the ACM, 35(12), 51-60, 1992

[11] Paul Resnick. *1997, Filtering Information on the Internet*. http://www.sciam.com/0397issue/0397resnick.html

[12] Chengfeng Han, Hideo Fujii, W. Bruce Croft. *1994 Automatic Query Expansion for Japanese Text Retrieval*. UMass Technical Report

[13] Ronald Rosenfeld. *1994, Adaptive Statistical Language Modeling: A Maximum Entropy Approach*. Carnegie Mellon University, Ph.D. Thesis

[14] Susan Gauch, Robert P. Futrelle. *Experiments in Automatic Word Class and Word Sense Identification for Information Retrieval*. Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval

[15] Kathleen Webster, Kathryn Paul. *1996, Beyond Surfing: Tools and Techniques for Searching the Web*. Jan. 1996, Information Technology

[16] Thorsten Joachims, Dayne Freitag, Tom Mitchell. *1997, WebWatcher: A Tour Guide for the World Wide Web*. Proceedings of IJCAI97, August 1997.

[17] H.Lieberman. *1995, Letizia: An agent that assists web browsing*. In International Joint Conference of Artificial Intelligence, Montreal, Aug. 1995

[18] Bernardo A. Huberman, Michael Kaminsky. *1996, Beehive: A System for Cooperative Filtering and Sharing of Information*

[19] URL: *Collection of Bots in the Web*, http://www.botspot.com/

Exh. 1003

IJCAI-95

Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence

Volume I

Montréal, Québec, Canada

August 20–25, 1995

Sponsored by the

International Joint Conferences on Artificial Intelligence, Inc. (IJCAII)
American Association for Artificial Intelligence
Canadian Society for Computational Studies of Intelligence
Société Canadienne pour l'étude de l'intelligence par ordinateur

Edited by Chris S. Mellish

Distributed by
Morgan Kaufmann Publishers, Inc.
2929 Campus Drive, Suite 260
San Mateo, California 94403

Printed in the United States of America.

Production management, design, and composition by
Professional Book Center, Denver, Colorado.

Susann LuperFoy
1320 Schuman Court
Vienna, Virginia 22182

Brief Contents

Volume 1

Volume 2

Contents

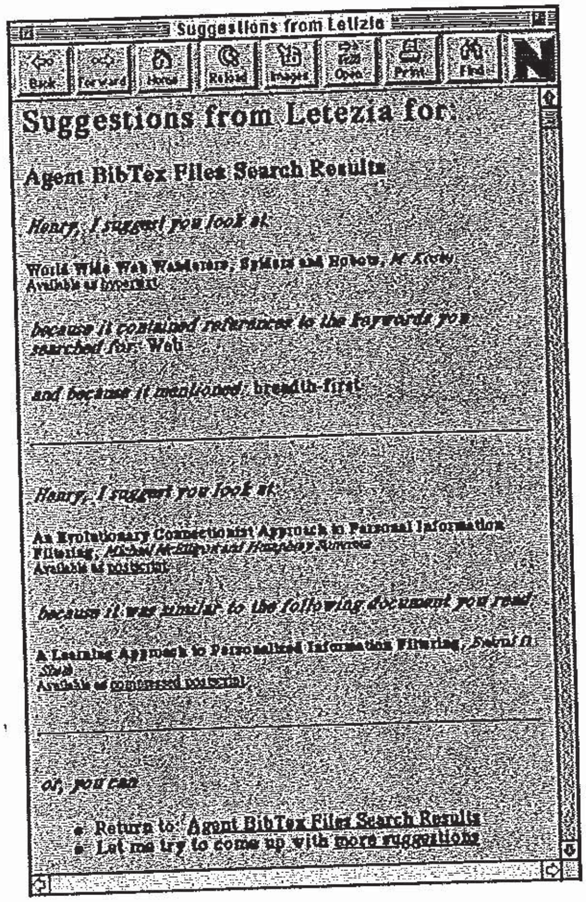
VOLUME 1

ACTION AND PERCEPTION

AUTOMATED REASONING

Letizia: An Agent That Assists Web Browsing

Henry Lieberman
Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA, USA
lieber@media.mit.edu



Abstract

Letizia is a user interface agent that assists a user browsing the World Wide Web. As the user operates a conventional Web browser such as Netscape, the agent tracks user behavior and attempts to anticipate items of interest by doing concurrent, autonomous exploration of links from the user's current position. The agent automates a browsing strategy consisting of a best-first search augmented by heuristics inferring user interest from browsing behavior.

1 Introduction

"Letizia Álvarez de Toledo has observed that this vast library is useless: rigorously speaking, a *single volume* would be sufficient, a volume of ordinary format, printed in nine or ten point type, containing an infinite number of infinitely thin leaves."

- Jorge Luis Borges, *The Library of Babel*

The recent explosive growth of the World Wide Web and other on-line information sources has made critical the need for some sort of intelligent assistance to a user who is browsing for interesting information.

Past solutions have included automated searching programs such as WAIS or Web crawlers that respond to explicit user queries. Among the problems of such solutions are that the user must explicitly decide to invoke them, interrupting the normal browsing process, and the user must remain idle waiting for the search results.

This paper introduces an agent, *Letizia*, which operates in tandem with a conventional Web browser such as Mosaic or Netscape. The agent tracks the user's browsing behavior -- following links, initiating searches, requests for help -- and tries to anticipate what items may be of interest to the user. It uses a simple set of heuristics to model what the user's browsing behavior might be. Upon request, it can display a page containing its current recommendations, which the user can choose either to follow or to return to the conventional browsing activity.

2 Interleaving browsing with automated search

The model adopted by Letizia is that the search for information is a cooperative venture between the human user and an intelligent software agent. Letizia and the user both browse the same search space of linked Web documents, looking for "interesting" ones. No goals are predefined in advance. The difference between the user's search and Letizia's is that the user's search has a reliable static evaluation function, but that Letizia can explore search alternatives faster than the user can. Letizia uses the past behavior of the user to anticipate a rough approximation of the user's interests.

Critical to Letizia's design is its control structure, in which the user can manually browse documents and conduct searches, without interruption from Letizia. Letizia's role during user interaction is merely to observe and make inferences from observation of the user's actions that will be relevant to future requests.

In parallel with the user's browsing, Letizia conducts a resource-limited search to anticipate the possible future needs of the user. At any time, the user may request a set of recommendations from Letizia based on the current state of the user's browsing and Letizia's search. Such recommendations are dynamically recomputed when anything changes or at the user's request.

Letizia is in the tradition of *behavior-based* interface agents [Maes 94], [Lashkari, Metral, and Maes 94]. Rather than rely on a preprogrammed knowledge representation structure to make decisions, the knowledge about the domain is incrementally acquired as a result of inferences from the user's concrete actions.

Letizia adopts a strategy that is midway between the conventional perspectives of *information retrieval* and *information filtering* [Sheth and Maes 93]. Information retrieval suggests the image of a user actively querying a base of [mostly irrelevant] knowledge in the hopes of extracting a small amount of relevant material. Information filtering paints the user as the passive target of a stream of [mostly relevant] material, where the task is to remove or de-emphasize less relevant material. Letizia can interleave both retrieval and filtering behavior initiated either by the user or by the agent.

3 Modeling the user's browsing process

The user's browsing process is typically to examine the current HTML document in the Web browser, decide which, if any, links to follow, or to return to a document previously encountered in the history, or to return to a document explicitly recorded in a *hot list*, or to add the current document to the hot list.

The goal of the Letizia agent is to automatically perform some of the exploration that the user would have done while the user is browsing these or other documents, and to evaluate the results from what it can determine to be the user's perspective. Upon request, Letizia provides recommendations for further action on the user's part, usually in the form of following links to other documents.

Letizia's leverage comes from overlapping search and evaluation with the "idle time" during which the user is reading a document. Since the user is almost always a better judge of the relevance of a document than the system, it is usually not worth making the user wait for the result of an automated retrieval if that would interrupt the browsing process. The best use of Letizia's recommendations is when the user is unsure of what to do next. Letizia never takes control of the user interface, but just provides suggestions.

Because Letizia can assume to be operating in a situation where the user has invited its assistance, its simulation of the user's intent need not be extremely accurate for it to be useful. Its guesses only need be better than no guess at all, and so even weak heuristics can be employed.

4 Inferences from the user's browsing behavior

Observation of the user's browsing behavior can tell the system much about the user's interests. Each of these heuristics is weak by itself, but each can contribute to a judgment about the document's interest.

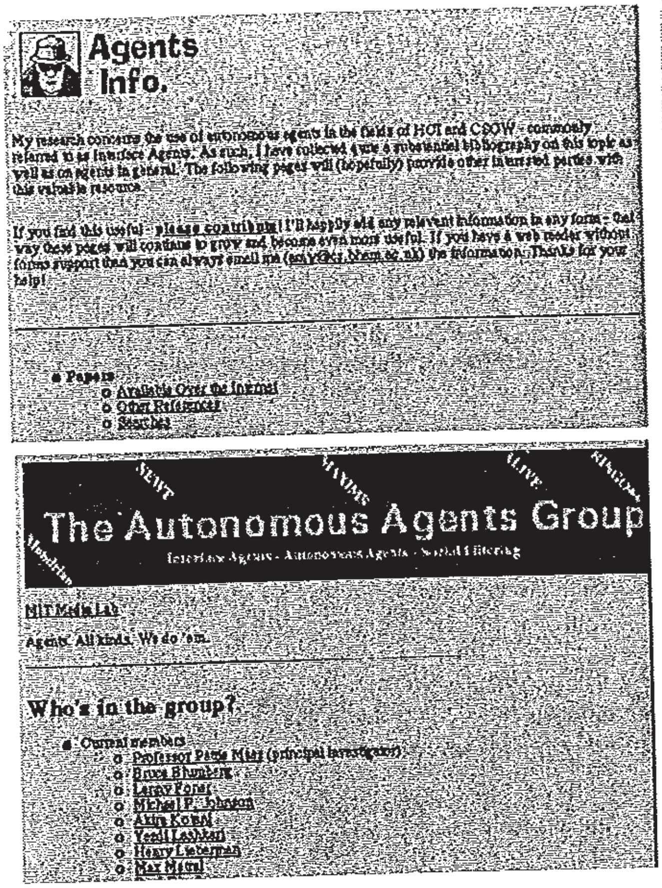
One of the strongest behaviors is for the user to save a reference to a document, explicitly indicating interest. Following a link can indicate one of several things. First, the decision to follow a link can indicate interest in the topic of the link. However, because the user does not know what is referenced by the link at the time the decision to follow it has been made, that indication of interest is tentative, at best. If the user returns immediately without having either saved the target document, or followed further links, an indication of disinterest can be assumed. Letizia saves the user considerable time that would be wasted exploring those "dead-end" links.

Following a link is, however, a good indicator of interest in the document *containing* the link. Pages that contain lots of links that the user finds worth following are interesting. Repeatedly returning to a document also connotes interest, as would spending a lot of time browsing it [relative to its length], if we tracked dwell time.

Since there is a tendency to browse links in a top-to-bottom, left-to-right manner, a link that has been "passed over" can be assumed to be less interesting. A link is passed over if it remains unchosen while the user chooses other links that appear later in the document. Later choice of that link can reverse the indication.

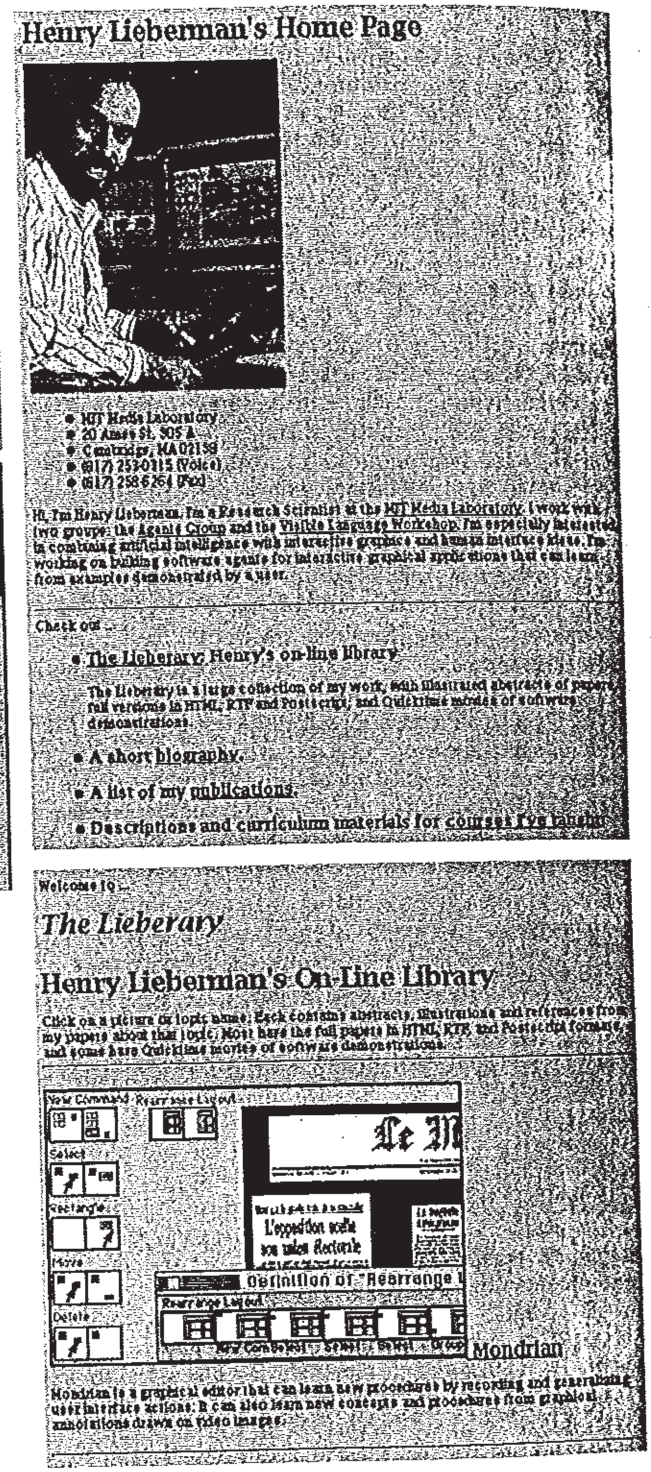
Letizia does not have natural language understanding capability, so its content model of a document is simply as a list of keywords. Partial natural language capabilities that can extract some grammatical and semantic information quickly, even though they do not perform full natural language understanding [Lehnert 93] could greatly improve its accuracy.

Letizia uses an extensible object-oriented architecture to facilitate the incorporation of new heuristics to determine interest in a document, dependent on the user's actions, history, and the current interactive context as well as the content of the document.



User browses many pages having to do with "Agents". System infers interest in the topic "Agent".

An important aspect of Letizia's judgment of "interest" in a document is that it is not trying to determine some measure of how interesting the document is in the abstract, but instead, a *preference* ordering of interest among a set of links. If almost every link is found to have high interest, then an agent that recommends them all isn't much help, and if very few links are interesting, then the agent's recommendation isn't of much consequence. At each moment, the primary problem the user is faced with in the browser interface is "which link should I choose next?", And so it is Letizia's job to recommend which of the several possibilities available is most likely to satisfy the user. Letizia sets as its goal to recommend a certain *percentage* [settable by the user] of the links currently available.



Later, the user independently browses a personal Web page, with a publications list. Letizia recommends articles having to do with "Agents".

5 An example

In the example, the user starts out by browsing home pages for various general topics such as Artificial Intelligence. Our user is particularly interested in topics involving Agents, so he or she zeros in on pages that treat that topic, such as the general Agent Info page, above. Many pages will have the word Agent in the name, the user may search for the word Agent in a search page, etc. and so the system can infer an interest in the topic of Agents from the browsing behavior.

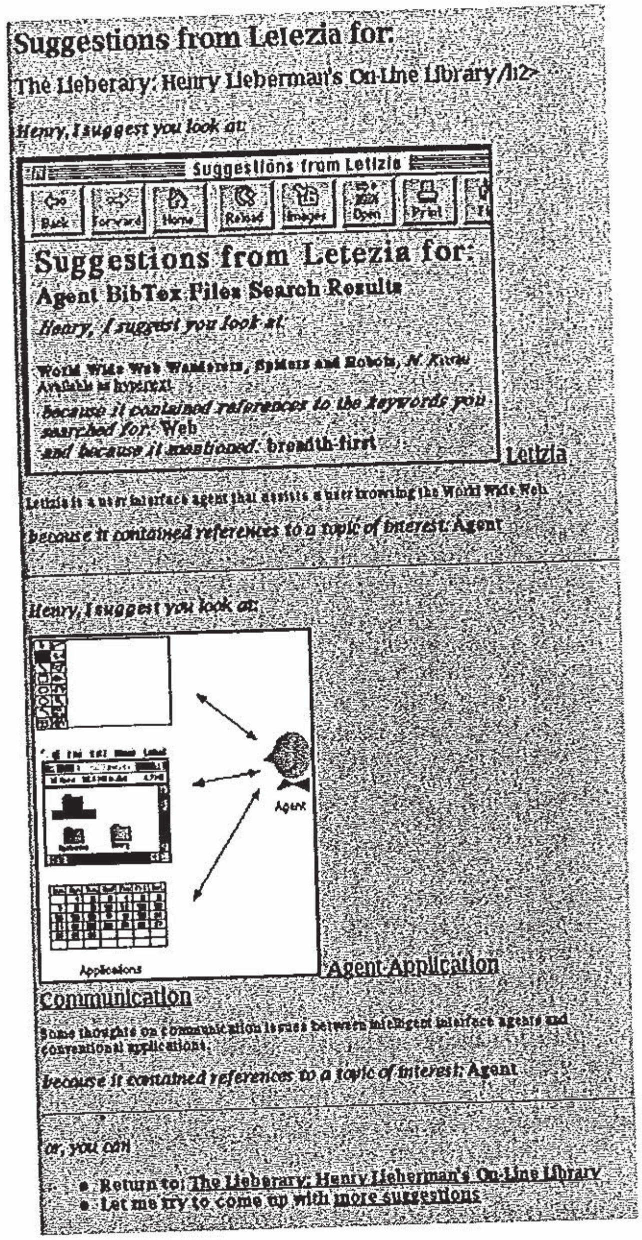
At a later time, the user is browsing personal home pages, perhaps reached through an entirely different route. A personal home page for an author may contain a list of that author's publications. As the user is browsing through some of the publications, Letizia can concurrently be scanning the list of publications to find which ones may have relevance to a topic for which interest was previously inferred, in this case the topic Agents. Those papers in the publication list dealing with agents are suggested by Letizia.

Letizia can also explain why it has chosen that document. In many instances, this represents not the only reason for having chosen it, but it selects one of the stronger reasons to establish plausibility. In this case, it noticed a keyword from a previous exploration, and in the other case, a comparison was made to a document that also appeared in the list returned by the bibliography search.

6 Persistence of interest

One of the most compelling reasons to adopt a Letizia-like agent is the phenomenon of *persistence of interest*. When the user indicates interest by following a link or performing a search on a keyword, their interest in that topic rarely ends with the returning of results for that particular search.

Though the user typically continues to be interested in the topic, he or she often cannot take the time to restate interest at every opportunity, when another link or search opportunity arises with the same or related subject. Thus the agent serves the role of remembering and looking out for interests that were expressed with past actions.



Persistence of interest is also valuable in capturing users' preferred personal strategies for finding information. Many Web nodes have both subject-oriented and person-oriented indices. The Web page for a university or company department typically contains links to the major topics of the department's activity, and also links to the home pages of the department's personnel. A particular piece of work may be linked to by both the subject and the author.

Some users may habitually prefer to trace through personal links rather than subject links, because they may already have friends in the organization or in the field, or just because they may be more socially oriented in general. An agent such as Letizia picks up such preferences, through references to links labeled as "People", or through noticing particular names that may appear again and again in different, though related, contexts.

Indications of interest probably ought to have a factor of decaying over time so that the agent does not get clogged with searching for interests that may indeed have fallen from the user's attention. Some actions may have been highly dependent upon the local context, and should be forgotten unless they are reinforced by more recent action. Another heuristic for forgetting is to discount suggestions that were formulated very far in "distance" from the present position, measured in number of web links from the original point of discovery.

Further, persistence of interest is important in uncovering *serendipitous connections*, which is a major goal of information browsing. While searching for one topic, one might accidentally uncover information of tremendous interest on another, seemingly unrelated, topic. This happens surprisingly often, partly because seemingly unrelated topics are often related through non-obvious connections. An important role for the agent to play is in constantly being available to notice such connections and bring them to the user's attention.

7 Search strategies

The interface structure of many Web browsers encourages depth first search, since every time one descends a level the choices at the next lower level are immediately displayed. One must return to the containing document to explore brother links at the same level, a two-step process in the interface. When the user is exploring in a relatively undirected fashion, the tendency is to continue to explore downward links in a depth-first fashion. After a while, the user finds him or herself very deep in a stack of previously chosen documents, and [especially in the absence of much visual representation of the context] this leads to a "lost in hyperspace" feeling.

The depth-first orientation is unfortunate, as much information of interest to users is typically embedded rather shallowly in the Web hierarchy. Letizia compensates for this by employing a breadth-first search. It achieves utility in part by reminding users of neighboring links that might escape notice. It makes user exploration more efficient by automatically eliding many of the "dead-end" links that waste users' time.

The depth of Letizia's search is also limited in practice by the effects of user interaction. Web pages tend to be of relatively similar size in terms of amount of text and number of links per page, and users tend to move from one Web node to another at relatively constant intervals. Each user movement immediately refocuses the search, which prevents it from getting too far afield.

The search is still potentially combinatorially explosive, so we put a resource limitation on search activity. This limit is expressed as a maximum number of accesses to non-local Web nodes per minute. After that number is reached, Letizia remains quiescent until the next user-initiated interaction.

Letizia will not initiate further searches when it reaches a page that contains a search form, even though it could benefit enormously by doing so, in part because there is as yet no agreed-upon Web convention for time-bounding the search effort. Letizia will, however, recommend that a user go to a page containing a search form.

In practice, the pacing of user interaction and Letizia's internal processing time tends to keep resource consumption manageable. Like all autonomous Web searching "robots", there exists the potential for overloading the net with robot-generated communication activity. We intend to adhere to conventions for "robot exclusion" and other "robot ethics" principles as they are agreed upon by the network community.

8 Related work

Work on intelligent agents for information browsing is still in its infancy. The closest work to this is [Armstrong, et. al. 95], especially in the interface aspects of annotating documents that are being browsed independently by the user. Letizia differs in that it does not require the user to state a goal at the outset, instead trying to infer "goals" implicitly from the user's browsing behavior. Also quite relevant is [Balabonovic and Shoham 95], which requires the user to explicitly evaluate pages. Again, we try to infer evaluations from user actions. Both explicit statements of goals and explicit evaluations of the results of browsing actions do have the effect of speeding up the learning algorithm and making it more predictable, at the cost of additional user interaction.

[Etzioni and Weld 94], [Knoblock and Arens 93], and [Perkowitz and Etzioni 95] are examples of a knowledge-intensive approach, where the agent is pre-programmed with an extensive model of what resources are available on the network and how to access them. The knowledge-based approach is complementary to the relatively pure behavior-based approach here, and they could be used together.

Automated "Web crawlers" [Koster 94] have neither the knowledge-based approach nor the interactive learning approach,. They use more conventional search and indexing techniques. They tend to assume a more conventional question-and-answer interface mode, where the user delegates a task to the agent, and then waits for the result. They don't have any provision for making use of concurrent browsing activity or learning from the user's browsing behavior.

Laura Robin [Robin 90] explored using an interactive, resource-limited, interest-dependent best-first search in a browser for a linked multimedia environment. Some of the ideas about control structure were also explored in a different context in [Lieberman 89].

9 Implementation

Letizia is implemented in Macintosh Common Lisp. It uses Netscape as a Web browser and user interface. The agent runs as a separate process, and communication between Lisp and Netscape takes place using AppleEvents and AppleScript interprocess communication. Currently, we are severely limited by the extent to which Netscape is programmable via AppleEvents. HTML is parsed using the Zebu parser-generator [Laubsch 94].

Acknowledgments

This work has been supported in part by research grants to the MIT Media Laboratory from Alenia, ARPA/JNIDS, Apple Computer, the National Science Foundation, and other Media Lab sponsors.

References

[Armstrong, Freytag, Joachims and Mitchell 95] Robert Armstrong, Dayne Freitag, Thorsten Joachims and Tom Mitchell, WebWatcher: A Learning Apprentice for the World Wide Web, in *AAAI Spring Symposium on Information Gathering*, Stanford, CA, March 1995.

[Balabonovic and Shoham 95] Marko Balabanovic and Yoav Shoham, Learning Information Retrieval Agents: Experiments with Automated Web Browsing, in *AAAI Spring Symposium on Information Gathering*, Stanford, CA, March 1995.

[Etzioni and Weld 94] Oren Etzioni and Daniel Weld, A Softbot-Based Interface to the Internet, *Communications of the ACM*, July 1994.

[Knoblock and Arens 93] Craig Knoblock and Yigal Arens, An Architecture for Information Retrieval Agents, *AAAI Symposium on Software Agents*, Stanford, CA, March 1993.

[Koster 94] M. Koster, World Wide Web Wanderers, Spiders and Robots, http:// web.nexor.co.uk/ mak/doc/robots/robots.html

[Lashkari, Metral, and Maes 94] Yezdi Lashkari, Max Metral, Pattie Maes, Collaborative Interface Agents, *Conference of the American Association for Artificial Intelligence*, Seattle, August 1994.

[Laubsch 94] Zebu: A Tool for Specifying Reversible LALR(1) Parsers, Hewlett-Packard Laboratories, 1994.

[Lehnert and Sundheim 91] Wendy Lehnert and B. Sundheim, A Performance Evaluation of Text-Analysis Technologies," *AI Magazine*, p. 81-94., 1991

[Lieberman 89] Henry Lieberman, Parallelism in Interpreters for Knowledge Representation Systems, in *Concepts and Characteristics of Knowledge-Based Systems*, M. Tokoro, Y. Anzai, A. Yonezawa, eds., North-Holland, 1989.

[Maes 94] Pattie Maes, Agents that Reduce Work and Information Overload, *Communications of the ACM*, July 1994.

[Perkowitz and Etzioni 95] Mike Perkowitz and Oren Etzioni, Category Translation: Learning to Understand Information on the Internet, *International Joint Conference on Artificial Intelligence*, Montréal, August 1995.

[Robin 90] Robin, Laura, Personalizing Hypermedia: The Role of Adaptive Multimedia Scripts, MS Thesis, Massachusetts Institute of Technology, 1990.

[Sheth and Maes 93] Beerud Sheth and Pattie Maes, Evolving Agents for Personalized Information Filtering, *IEEE Conference on Artificial Intelligence for Applications*, 1993.

Exh. 1004

AF $

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Gary Odom

Application No.: 09/796,235

Filed: 02/28/2001

For: Automatic directory supplementation

Examiner: William D. Hutton, Jr.

Art Unit: 2179

Date: February 14, 2005

Mail Stop AF
Commissioner for Patents
Box 1450
Alexandria, VA 22313-1450

NOTICE OF APPEAL FROM THE EXAMINER TO THE BOARD OF PATENT APPEALS AND INTERFERENCES

Gary Odom hereby appeals to the Board from the decision of Examiner William D. Hutton, Jr. mailed January 26, 2005, finally rejecting claims 9-30.

If an extension of time is required for filing this Notice of Appeal, please consider this a petition therefor.

A triplicate copy of this Notice of Appeal is enclosed.

The $250.00 fee per 37 C.F.R. § 1.17 (b) for filing this Notice of Appeal is enclosed as a credit card form. Please charge any additional fees that may be required in connection with filing this Notice of Appeal and any extension of time, or credit any overpayment, to the credit card on the enclosed credit card form.

Respectfully,

[signature]

Gary Odom
15505 SW Bulrush Lane, Tigard, OR 97223
telephone: (503) 524-8371
fax: (775) 942-8525

02/22/2005 MAHMED1 00000048 09796235
01 FC:2401 250.00 OP

Application No.: 09/796,235
Filed: 02/28/2001
Group Art Unit: 2179

In the United States Patent and Trademark Office

Before The Board of Patent Appeals and Interferences

Regarding the application:

Title: Automatic directory supplementation

Number: 09/796,235

Priority: 02/28/2001

Examiner: William Hutton, Jr.

Art Unit: 2179

Mail Stop Appeal Brief-Patents
Commissioner for Patents
Box 1450
Alexandria, VA 22313-1450

Brief For Appellant

This is an appeal from the Examiner's January 26, 2005 final rejection.

1. Real Party in Interest

Gary Odom, appellant, is the real party in interest.

2. Related Appeals and Interferences

There are no related appeals or interferences.

3. Status of Claims

Appeal is sought for rejection of claims 9-24, 27-29. Claims 25-26, and 30 are herein canceled. Claim 31 is objected to as being dependent upon a rejected base claim.

4. Status of Amendments

No amendment has been filed subsequent to final rejection.

5. Summary of Invention

09/796,235 describes an autonomous search mechanism, solving the problem of finding similar documents to ones already known without any user effort whatsoever. The only precondition to initiating the claimed process is user placement of one or more documents in a file system directory as reference material for guiding the search.

09/796,235 is fairly characterized as lazy because time is not of the essence. A user doesn't initiate search: the process works in the background, without arousing expectation of quick results.

As an exemplary use-case scenario, a user browses the web, saving topically-related document links in the same web-favorites folder. Once this precondition is met, the claimed invention software kicks in: deriving keywords from the saved documents, thus discerning the topic of interest, then searching for other related documents, resulting in supplementing the directory with newly-found documents - hence the title of 09/796,235: "automatic directory supplementation".

6. Issues

There was but one overall issue in Examiner's January 26, 2005 final rejection: 35 U.S.C. §103 combination reference anticipation by prior art.

Appellant respectfully contends:

Essential features of the prior art itself were mischaracterized as bases for rejection.

The references, even combined, fail to anticipate all limitations of the claims.

Used as bases for rejection, the necessary combination of references, or applying specific features of one reference with another, comprise a non-obvious combination. The cited prior art references themselves provide no suggestion of combination. Respectfully, Examiner applied impermissible hindsight, without regard to prior art teaching or motivation.

With all due respect, there appears a lapse in considering the claims and prior art holistically, instead treating claim limitations and prior art reference features as dissectible components, without proper regard for context.

7. GROUPING OF CLAIMS

On the whole, from a viewpoint of patentability, of claims standing or falling together, there is but one group.

8. ARGUMENT

Statutory and case law bases for determining whether a preamble limits a claim

MPEP 2111.02 discusses preamble statements limiting structure or intended use. The meaning MPEP 2111.02 and case law are plain and clear that a preamble may limit claim scope. Examiner cited the same quotation. Preamble claim limitation may of course be supported by example within the claim body.

> MPEP 2111.02 - Any terminology in the preamble that limits the structure of the claimed invention must be treated as a claim limitation. See, e.g., Corning Glass Works v. Sumitomo Elec. U.S.A., Inc., 868 F.2d 1251, 1257, 9 USPQ2d 1962, 1966 (Fed. Cir. 1989).
>
> "[A] claim preamble has the import that the claim as a whole suggests for it." Bell Communications Research, Inc. v. Vitalink Communications Corp., 55 F.3d 615, 820, 34 USPQ2d 1816, 1820 (Fed. Cir. 1995). "If the claim preamble, when read in tha context of the entire claim, recites limitations of the claim, or, if the claim preamble is `necessary to give life, meaning, and vitality' to the claim, then the claim preamble should be construed as if in the balance of the claim." Pitney Bowes, Inc. v. Hewlett-Packard Co., 182 F.3d 1298, 1305, 51 USPQ2d 1161, 1165-66 (Fed. Cir. 1999). See also Kropa v. Robie, 187 F.2d 150, 152, 88 USPQ 478, 481 (CCPA 1951)

Specific arguments related to rejection and preamble limitation are discussed in the below section titled: "Unanticipated limitations for all claims".

Statutory and case law bases for 35 U.S.C. §103 rejections based upon prior art combinations

The consistency of the below quotations edify criteria for obviousness rejection via 35 U.S.C. §103 using a combination of references.

I. The prior art references themselves must suggest combination. Failing explicit self-suggestion, the prior art must provide the motivation for obviousness in combination. Such motivation may be found by considering the references holistically. If the purpose / problem being solved ("nature of the problem"), function and structure of the prior art references are aligned, one may reasonably conclude combination of the references obvious, as no differences exist in the principles of operation between the references. The burden of meeting this criterion by logical exposition belongs to the Examiner.

3. To combine references without evidentiary support by the prior art constitutes impermissible hindsight. Combination of prior art with different principles of operation is impermissible. An Examiner cannot simply assert 'well within the ordinary skill of the art at the time the claimed invention was made'.

4. To be construed anticipatory, the prior art must teach or at least suggest all claim limitations, whether such limitations appears in the preamble or body of a claim.

5. The final test is comparing the claimed invention as a whole to a prior art reference. Claim limitations are not puzzle pieces to be matched to atomized prior art reference suggestions, and thus examined out of context. As with obviousness in combining prior art references, only if the prior art aligns with the claimed invention in principles of operation may a prior art reference be considered anticipatory.

MPEP 2143 -To establish a prima facie case of obviousness, three basic criteria must be met. First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art reference (or references when combined) must teach or suggest all the claim limitations. The teaching or suggestion to make the claimed combination and the reasonable expectation of success must both be found in the prior art, not in applicant's disclosure. In re Vaeck, 947 F.2d 488, 20 USPQ2d 1438 (Fed. Cir. 1991).

The initial burden is on the examiner to provide some suggestion of the desirability of doing what the inventor has done. "To support the conclusion that the claimed invention is directed to obvious subject matter, either the references must expressly or impliedly suggest the claimed invention or the examiner must present a convincing line of reasoning as to why the artisan would have found the claimed invention to have been obvious in light of the teachings of the references." Ex parte Clapp, 227 USPQ 972, 973 (Bd. Pat. App. & Inter. 1985).

When applying 35 U.S.C. 103, the following tenets of patent law must be adhered to:

(A) The claimed invention must be considered as a whole;

(B) The references must be considered as a whole and must suggest the desirability and thus the obviousness of making the combination;

(C) The references must be viewed without the benefit of impermissible hindsight vision afforded by the claimed invention; and

(D) Reasoneble expectation of success is the standard with which obviousness is determined.

Hodosh v. Block Drug Co., Inc., 786 F.2d 1136, 1143 n.5, 229 USPQ 182, 187 n.5 (Fed. Cir. 1986).

Obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention where there is some teaching, suggestion, or motivation to do so found either explicitly or implicitly in the references themselves or in the knowledge generally available to one of ordinary skill in the art. "The test for an implicit showing is what the combined teachings, knowledge of one of ordinary skill in the art, and the nature of the problem to be solved as a whole would have suggested to those of ordinary skill in the art." *In re Kotzab,* 217 F.3d 1365, 1370, 55 USPQ2d 1313, 1317 (Fed. Cir. 2000). See also *In re Lee,* 277 F.3d 1338, 1342-44, 61 USPQ2d 1430, 1433-34 (Fed. Cir. 2002) (discussing the importance of relying on objective evidence and making specific factual findings with respect to the motivation to combine references); *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988); *In re Jones*, 958 F.2d 347, 21 USPQ2d 1941 (Fed. Cir. 1992).

The mere fact that references can be combined or modified does not render the resultant combination obvious unless the prior art also suggests the desirability of the combination. In re Mills, 916 F.2d 680, 16 USPQ2d 1430 (Fed. Cir. 1990)

A statament that modifications of the prior art to meet tha claimad invention would have been " 'well within the ordinary skill of tha art at tha time tha claimed invantion was mada' " because tha references relied upon teach that all aspects of the claimed invention ware individually known in tha art is not sufficient to astablish a prima facia case of obviousnass without some objective reason to combina the teachings of tha referances. Ex parta Levengood, 28 USPQ2d 1300 (Bd. Pat. App. & Intar. 1993). See also In re Kotzab, 217 F.3d 1365, 1371, 55 USPQ2d 1313, 1318 (Fed. Cir. 2000)

MPEP 2141.01(a) - Whila Patent Office classification of references and tha cross-referances in tha official search notes of the class definitions are some evidence of "nonanalogy" or "analogy" respectively, the court has found "the similarities and differences in structure and function of the inventions to carry far greater weight." In ra Ellis, 476 F.2d 1370, 1372, 177 USPQ 526, 527 (CCPA 1973)

To establish prima facie obviousness of a claimed invention, all the claim limitations must be taught or suggasted by tha prior art. In re Royka, 490 F.2d 981, 180 USPQ 580 (CCPA 1974). "All words in a claim must ba considared in judging the patentability of that cleim against tha prior art." In re Wilson, 424 F.2d 1382, 1385, 165 USPQ 494, 496 (CCPA 1970). If an independent claim is nonobvious under 35 U.S.C. 103, then any claim dapending therefrom is nonobvious. In ra Fine, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988).

If the proposad modification or combination of the prior art would change the principle of oparation of the prior art invention being modified, then tha teachings of the referances are not sufficient to render the claims prima facie obvious. In re Ratti, 270 F.2d 810, 123 USPQ 349 (CCPA 1959) (Claims ware directed to an oil seal comprising a bora engaging portion with outwardly biased resilient spring fingers inserted in a resiliant sealing membar. The primary refarence relied upon in a rejection based on a combination of raferences disclosed an oil saal wherain tha bore engaging portion was rainforced by a cylindrical sheat metal casing. Patentee taught the davice requirad rigidity for operation, whereas the claimed invention required resiliancy. The court raversed tha rejection holding tha "suggested combination of references would require a substantial reconstruction and redesign of the elements shown in [the primary reference] as well as a change in the basic principle under which the [primary reference] construction was designed to operate." 270 F.2d at 813, 123 USPQ at 352.).

Distilling an invention down to the "gist" or "thrust" of an invention disregards the requirement of analyzing the subject matter "as a whole." W.L. Gore & Associates, Inc. v. Garlock, Inc., 721 F.2d 1540, 220 USPQ 303 (Fed. Cir. 1983), cert. denied, 469 U.S. 851 (1984)

In determining the differences between the prior art and the claims, the question under 35 U.S.C. 103 is not whether the differences themselves would have been obvious, but whether the claimed invention as a whole would heve been obvious. Stratoflex, Inc. v. Aeroquip Corp., 713 F.2d 1530, 218 USPQ 871 (Fed. Cir. 1983); Schenck v. Nortron Corp., 713 F.2d 782, 218 USPQ 698 (Fed. Cir. 1983)

A prior art reference must be considered in its entirety, i.e., as a whole, including portions that would lead away from the claimed invention. W.L. Gore & Associates, Inc. v. Garlock, Inc., 721 F.2d 1540, 220 USPQ 303 (Fed. Cir. 1983), cert. denied, 469 U.S. 851 (1984)

Claim Rejections - 35 U.S.C. §103

There are four aspects to considering the claim rejections: 1) understanding the nature of the prior art references; 2) considering the appropriateness of combining prior art references or specific features thereof; 3) assessing the anticipatory power of the prior art used for rejection, particularly what remains unanticipated; 4) examining the specific logic for rejection on a claim-by-claim basis.

Prior art references used for 35 U.S.C. §103 rejections

One cannot appreciate a prior art reference as anticipatory without understanding it holistically: the nature of the problem being solved and solution provided, namely function and structure. Similarly, one cannot consider the appropriateness of combination without checking alignment of principles of operation.

5,598,557 (Doner) - Getting highly relevant results from a coherent database

> ...searching and retrieving files in a database without a user being required to provide keywords or query terms. A user first selects and opens a reference file... Relevant files are prioritized and displayed to the user in groups... The groups of retrieved files are displayed in associating with the subject word they are relevant to. (abstract)

Doner required user selection of a topically-coherent target database for searching.

> To conduct a search, a user first specifies a particular database. Databases are usually organized so that files stored on a particular database share a common attribute. For example, an attorney might utilize a database containing cases from a particular jurisdiction; a doctor might consult a database containing files of patient histories; a marketing manager might access a database containing product reviews for spotting market trends; etc. The database can be an already existing database or a newly created database. (4:65-5:5)

Doner's database is indexed for rapid searching, a typical technique.

> Finally, the processed information is indexed and saved to the database, step 207.

In the most relevant embodiment to the claimed invention, Doner allowed user-specified search based upon a user-selected reference file, in lieu of directly inputting search terms (the other option for specifying search parameters):

> Once a database has been selected, the user can select a weighted keyword search, a weighted Boolean search, or a document agent search. (5:21-23)
>
> Alternatively, a user can opt for a Document Agent Search, which allows the user to initiate a search for documents which are similar to a reference document selected by the user. First, the user selects and opens a reference document. Next, the user selects the Document Agent Search option from the Search pull-down menu. (6:14-18)

Doner did mention networking: "Finally, computer system 100 can be a terminal in a computer network (i.e., a LAN)" (4:60-62), suggesting that the target database may be on a networked computer.

Doner displayed results. Doner did not, as Examiner asserted, anticipate augmenting a directory as claimed.

2003/0195877 (Ford) - Finding products for sale

Ford aimed at e-commerce shopping convenience by finding and displaying all products for sale based upon user-input search terms.

> One problem currently encountered by online merchants is the inability to effectively present groups of related products that span the predefined categories. [0004]

Ford solved the problem of trying to provide inclusive results by accessing multiple databases.

> The web site includes a query server that processes queries by searching a number of databases. [0027]

Ford's technology did not search the Internet per se, but instead an indexed database of data gleaned from a spider crawl. This approach is ubiquitous with so-called Internet search sites/engines that offer a user quick search results.

> The Product Spider database 147 is generated through the use of a web crawler 160 that crawls web sites on the Internet 120 while storing copies of located web pages. The output of the web crawler 160 is input to a product score generator 162 that assigns a numerical score ("product score") to each web page based upon the likelihood that the page offers a product for sale for either online or offline purchase. [0034]

Ford did not search documents as claimed, but pre-digested database index records, the same as Doner.

> As noted above, the Product Spider database 147 is indexed by keyword 166. Each keyword in the database is associated with one or more web pages for which the indexer 164 has determined an association. [0037]

Ford's explanation of the derivation of the databases, including the Product Spider database, is at [0030]-[0031] and [0034]-[0037].

As Ford was concerned with the web environment, particularly product searching, searching is necessarily user-interactive. The user inputs both search terms, and sets the scope of the search (search location(s)).

Alternatively, users may search for products using a search engine interface 220. Users can perform searches with the search engine interface 220 by typing in the desired information (referred to herein as a "query") into a query window 230 and then clicking on a search initiation button 240. The user may control the scope of the search with a pulldown window 250 containing multiple categories. The search may be limited to any one category through selection of that category from the pulldown menu 250. Alternatively, the user may conduct a broad-based search through selection of an "All Products" option 260. [0040]

When the user submits a query from the search engine interface 220 of FIG. 2 to the web site 130, the query server 140 applies the query to the database, or databases, corresponding to the search scope selected by the user. [0046]

Given the utility of Ford's interactive product searching, where keywords are few, one would never think having to create a reference document to initiate a search. Ford certainly didn't.

Ford's real problem is not making search easy for the user (it already is), but being properly inclusive: namely, showing all products for sale, but not referencing sources that don't offer the desired product for sale.

6,353,822 (Lieberman) - recommending web pages via user profiling

Liebermen profiled a user's interests by tracking web page selection and consumption (reading time spent) while browsing the Internet. Recommendations of other web pages were made by a contemporaneous background search, using search terms from the profile.

The present invention operates in tandem with a conventional document-retrieval facility, such as a web browser, by tracking the choices made by the user in retrieving and viewing items (such as web pages)--i.e., which links are followed, when searches are initiated, requests for help, etc.--and, based thereon, identifying additional items likely to be of interest to the user. In other words, the invention browses the same search space as the user, but faster and guided by the user's past behavior. (3:52-60)

Liberman's technology searched the Internet for documents, similarly as the claimed technology. Neither Liebermen nor the claimed technology offers the same as the quick-response search engine Ford employed. Creating Ford's Product Spider database is a huge undertaking, requiring massive storage, not something a client computer would do, as opposed to the technology of Lieberman or that claimed herein.

Prior Art Combination

The prior art references used by the Examiner for rejection do not themselves suggest combination. Examiner provided no logical motivation for combining the specific features used for rejection by using the prior art as a touchstone of rationale.

Doner and Ford - Doner's reference document with Ford

For claims 9-17, 21-24, 27-29, Examiner combined Doner and Ford for rejection. Specifically, Examiner wanted to combine a specific feature of Doner's with Ford: allowing a user to select a reference document as a basis for search, in lieu of directly inputting search terms.

Search specification using Ford's process is quite simple: a specific product, so Doner's technique of simplifying search by using a reference document would be inappropriate in combination with Ford. There is no reason to think that a user would find it harder to type in "lawnmower" than select a reference document containing the same word; quite the contrary. Besides lack of self-suggestion within the prior art, not only is there no motivation to combine Doner's reference document with Ford's disclosed process, as Examiner contended, but the idea is counter-intuitive, and hence that specific feature combination constitutes impermissible hindsight.

Ford and Doner combined fail to anticipate other crucial claim limitations, as described below in the section titled: "Unanticipated limitations for all claims".

Lieberman with either Doner or Ford

The background of the 09/796235 specification briefly mentions search engines. The specification glossed over the different construction of search engines and search sites, as that technology itself was already well known to those skilled in the art. With all due respect, now facing rejection over confusion, some elucidation is required.

Lieberman performed ad hoc Internet document searching based upon a user profile of previously tracked input.

> Module 102 conducts the actual searches for candidate web items through web interface 50, which performs the mechanical tasks of accessing network 31 and retrieving items. (6:5-8)

With regard of ad hoc document searching, Lieberman and the claimed technology are equivalent. Lieberman used a database to store found documents, an unnecessary elaboration in using 09/796235 technology, but Lieberman's searching was of documents on the web. Lieberman and the claimed technology could easily tap into commercial search engines/sites, such as Google, for results, as suggested in the background of the 09/796235 specification, or in Lieberman 8:4-16.

Significantly different, Ford and Doner performed user-interactive database searches, relying upon user input for both search parameters and search scope/location.

Doner's anticipated a database that is self-constructed.

> The database can be an already existing database or a newly created database. FIG. 2 is a flowchart illustrating the steps for creating a new database. Computer files containing useful information can be imported by copying it over to the database, step 201. Moreover, data in the form of documents, reports, magazine and newspaper articles, can be entered either manually by means of a keyboard, step 202, or they can be entered by using an optical scanner, step 203. Moreover, the data can already exist on the computer system. The user can specify zones of a scanned image or file which is of particular significance for further processing, step 204. Textual portions of a scanned bit-map image or file can be recognized and converted into ASCII code data, step 205. The ASCII code data can then be edited, step 206. Finally, the processed information is indexed and saved to the database, step 207. (5:5-20)

In contrast to Doner, a different approach is Ford's Product Spider database, which resembles commercial search sites such as Google, A9, Alta Vista, Yahoo, and others. Here, a web crawler collates pages (or, at the least, page references) into a database, as well as creating an index record of keywords for each page. A user search doesn't actually go the web, but instead to the index of database records that comprise page links and their associated keywords.

> The Product Spider database 147 includes information about independent web sites, unaffiliated with the host web site 130, that have been identified as offering products for sale. This database is particularly useful in that it allows the host web-site

130 to help a consumer find product offerings for products that are not sold by the host web site 130 or by affiliated on-line merchants. [0030]

The nature of the problem, function and structure of Lieberman's ad hoc web document searching and the claimed invention differs markedly from the database index searching of Doner or Ford.

Examiner provided no explanation of logical connectivity between these references that could be grounded within the prior art itself, so as to make a Lieberman-Doner/Ford combination proper under the 35 U.S.C. §103 guidelines, applicable to the claimed invention.

Appellant respectfully traverses rejection of claims 18-22, and 27, combining Lieberman with Ford or Doner, as constituting impermissible hindsight.

Unanticipated limitations for all claims

"...without user input"

Respectfully, Examiner disavowed plain-meaning claim language in the preamble applicable to all claims: "augmenting a directory without user input". Examiner considered claim 9 as exemplary.

> Stating that a search and retrieval computer system "augments a directory" "without user input" could be interpreted in many ways. Search and retrieving computer files have many steps, including entering search criteria, search locations and the minutia performed by the computer to determine whether a computer file meets the search criteria and is retrieved. (01/26/2005 office action, pp. 27-28)

Examiner's "many ways" of interpreting "without user input" comes down to two aspects of potential user input:

1. search parameters/terms/criteria, and
2. search location(s).

Examiner's mention of " the minutia performed by the computer" is irrelevant to user input.

With all due respect, in context, Examiner's argument of vagueness with regard to "without user input" was an insupportable straw man.

So, Examiner overtly disagreed, but tacitly concurred with appellant, that, in context, the two limitations applicable to the meaning of "without user input" comprise:

1. no user input of search parameters;
2. no user input of search locations.

That is exactly what appellant had explained in his 08/27/2004 reply to the first office action rejection.

Appellant had amended claim 9 to explicitly point out "without user input of search location" as a claim limitation in the body of the claim. While on the one hand complaining about the preamble "without user input" limiting the scope of a claim, on the other hand, Examiner on page 28 of his 01/26/2005 office action inexplicably inferred a nefarious intent to stating an aspect of this limitation, "without user input of a search location", within the body of the claim.

None of the cited prior art references meet both aspects of the limitation "without user input". Particularly, Ford and Doner take user input of both search parameters and location. Doner in one embodiment allows user selection of a reference document in lieu of inputting search terms, but that still constitutes user input, albeit indirect input of search terms.

Lieberman created a user profile based upon tracking user input as a means for building search parameters. Relative to the claimed invention, Lieberman's was an active and tedious process of data collation from user input.

By contrast, the claimed invention relies solely upon documents in a directory, without relying upon user input. Yes, a user must first put the documents in the directory, but that is a precondition; user input is not required for the claimed process to work, unlike Lieberman. That cannot be said for Doner, Ford, Lieberman, search engines, or any other cited art used as a basis of rejection.

"augmenting a directory" (all claims)

Doner, Ford, and Lieberman all display results interactively. No cited prior art teaches augmenting a directory with found relevant references as a process termination as claimed.

Respectfully, Examiner's mistaken attributions with regard to the cited prior art adding results to a file directory are traversed.

"documents in storage"

The nature of "documents" as used in context throughout the claims and 09/796,235 specification is consistent and singular. In the claims, the same type of document is used for deriving search parameters, search, and results references, hence the same term: "document". Documents are individual files in storage, to which a reference may be made and used for access, as in a file pointer or hyperlink or URL (universal resource locator). Technically, from an access perspective, a document is always file system pointer/reference, as the file system may maintain a document in fragments on physical storage, collating the fragments and delivering the contents only upon request by software yielding a file pointer.

Ford and Doner searched databases, not documents as claimed.

Documents in a file system storage are not the same as database records.

One simply could not describe a technology that relies upon a database and not use the word "database". The word database does not appear in the 09/796,235 specification.

With all due respect, Examiner oversimplified Ford's disclosed process. Ford used a spider to create an indexed database of spider-found web pages, the fruit of a "search engine". Ford's user-interactive search was conducted on a database index, not the documents themselves. User interactivity would be severely compromised if a user had to await the results of a broad search of Internet documents in real-time. All known quick-response Internet search engines, Ford's included, take a moving snapshot of the Internet, predigesting web pages into a indexed database, then search the index upon user request input. The explanation and quotations provided describing Ford verify this process, and hence refute Examiner's assertion. So, Ford did not anticipate the plain literal meaning of the claim limitation "searching a plurality of documents". Examiner's contention that Ford searched documents is respectfully traversed.

Doner's search parameter reference document, which may be the same type of document as claimed, is not the same structure as the database record index searched, or a search result database record (albeit derived from a document).

Yes, Lieberman searched documents through the Internet, and displayed document references, but Lieberman relied upon a user-input derived profile for search parameters, not documents as claimed.

Owing to different natures of the problem, functions and structures, as argued foregoing, Lieberman cannot be combined with Doner or Ford without impermissible hindsight, as the prior itself provides no teaching, suggestion, or motivation.

Specific Claim Rejections

The foregoing arguments about unanticipated limitations applies in respectfully traversing rejection of all claims.

The foregoing arguments about the impermissible prior art combination of Lieberman with Doner or Ford apply to claim 18 and its dependents (claims 19-22).

Besides whatever specific arguments are presented below, all dependent claims rely one or more unanticipated limitations within their respective base claim for novelty.

Claim 9

Examiner: "Doner discloses a method for augmenting a directory without user input". This assertion is respectfully traversed. First, Doner displayed results; Doner never suggested augmenting a directory. Second, Doner required user input, both in input of search parameters (either directory by inputting search terms, or indirectly by selecting a reference document from which search terms are derived), and in selection of a database to search.

Examiner conceded that "Doner fails to disclose: searching a plurality of documents in storage in at least one computer *without user input of a search location.*"

Examiner: "Ford teaches a method for augmenting a directory." This assertion is respectfully traversed. Ford displayed results; Ford never suggested augmenting a directory.

Perhaps mistakenly mixing up Ford with Doner, Examiner contended that Ford taught the limitation of accessing a first document comprising context from which keywords are derived, an assertion respectfully traversed. Ford did not teach this limitation. The evidence Examiner presented with regard to Ford did not address this issue, and there is no such evidence to be found within Ford.

Doner taught using a reference document for keyword extraction as a prelude for search, though this required user selection of the document, and thus failed to meet the limitation in the preamble of accessing a first document *without user input.*

Further, as aforementioned, it would have been non-obvious to apply Doner's reference document approach for search parameters to Ford.

Examiner contended that Ford taught the limitation of "searching a plurality of documents in storage in at least one computer without user input of a search location". Respectfully traversed, Ford neither searched documents, nor did so without user input of a search location. Ford searched databases, not documents, as did Doner. Further, Ford stated, as quoted in full above, that searches are performed "corresponding to the search scope selected by the user".

Claim 12

Examiner contended that Doner disclosed the limitation in claim 12 of accessing a plurality of documents for derivation of keywords for search. The contention is respectfully traversed.

In one embodiment, Doner disclosed user selection of a single reference document as a base for keyword derivation. Doner made no suggestion of user selection of multiple such documents. Further, just having user selection fails to anticipate the limitation of operation without user input.

Claim 15

Examiner contended that Doner disclosed checking enable of directory augmentation. Examiner explained that "Doner discloses this limitation in that the system determine (sic) whether the database includes relevant documents".

Examiner's assertion is respectfully traversed. With all due respect, Examiner misconstrued Doner in light of the claim limitations. First, Doner did not anticipate augmenting a directory. Second, Doner 6:13-65 stated nothing with regard to checking an enablement option as to whether to search for results.

Claim 17

Again, Examiner contended that Ford taught augmenting a directory, an assertion respectfully traversed.

With all due respect, Examiner was grossly mistaken in referring to the Internet as a "dynamic database". The dispersed Internet is no database. Respectfully, this statement belies understanding the technical nature of databases or the Internet, and calls into question Examiner's objectivity and/or technical competence.

Databases comprise records adherent to a particular data structure, which are commonly indexed, as with Ford and Doner, and thus organized. The Internet is a network affording access to documents lacking any organization with regard to consistency in data structure. HTML, the common format of web pages, provides a page layout formatting mechanism; not at all the same as a data structure imposed by a database.

Claim 18

Based upon foregoing argument, appellant respectfully traverses Examiner's assertion that "Doner discloses an apparatus for augmenting a directory without user input".

Database index records taught by Doner and Ford are not document files. Respectfully, Examiner repeatedly confused the two as being the same. Please see the above section about "documents in storage".

As argued foregoing, respectfully traversing Examiner's assertion, no cited prior art, Doner especially, suggested adding a document reference to a search results directory.

Respectfully, in rejecting claim 18, Examiner combined Doner with Lieberman, along with Internet search engines, in a gumbo of impermissible hindsight.

Claim 23

Again, Examiner contended that Ford and Doner taught augmenting a directory, an assertion respectfully traversed. No cited prior art, neither Doner nor Ford particularly, suggested adding a document reference to a search results directory.

With all due respect, as described foregoing, Examiner mischaracterized Ford with regard to Ford searching the Internet per se. A web crawler does not search in the literal sense, and anyway Ford did not disclose crawling upon user invocation. A web crawler is a collation mechanism for database storage, and the database subsequently searched, as Ford disclosed.

Claim 28

Examiner contended that Ford's web crawl constituted a search for documents as claimed. With all due respect, Appellant posits this as a failure of appreciation regarding specific processes.

Ford performed a two-step process: first, a web crawl to populate a database, resulting in creating indexed records from the web pages gleaned in the crawl, where each record comprised keywords extracted from a web page, along with a link to the page; second, an actual user search was of the database records index, not of documents in storage across a network as claimed.

Claim Objections

Appellant takes no umbrage to amending the claims to meet Examiner's claim objections; in fact, thanks Examiner for such careful attention to detail; but poses the following comment.

"How to Write a Patent Application" by Jeffrey Sheldon is a well respected tome on the subject. "this book is highly recommended..." gushed the National Council of Intellectual Property Law Associations Newsletter. One chapter of Sheldon's book covers claim language.

Claim 9 & 23 - replacing "such that" with "wherein". According to Sheldon, "such that" indicates achieving a functional relationship, whereas "wherein" "is used to modify or qualify a previously introduced element". "such that" in claim 9 & 23 was used to establish a functional relationship: that searching by keyword (second element) was functionally related to searching documents (first element). Appellant considers "wherein" less specific, and thus less appropriate, in the particular instances where used.

Appellant does not request an oral hearing.

The $250.00 fee per 37 C.F.R. § 1.17 (c) for filing this appeal brief is enclosed as a separate credit card form. Please charge any additional fees that may be required in connection with filing this appeal brief and any extension of time, or credit any overpayment, to the credit card on the enclosed credit card form. Thank you.

Respectfully submitted,

Gary Odom
15505 SW Bulrush Lane, Tigard, OR 97223

telephone: (503) 524-8371
fax: (775) 942-8525
date: February 14, 2005

9. APPENDIX

1-8. (canceled)

9. (previously presented) A method for augmenting a directory without user input comprising the following steps:

accessing at least a first document via a first directory,

said first document comprising at least in part topical textual content;

deriving at least one first keyword indicative of at least one topical content within said first document;

searching a plurality of documents in storage in at least one computer without user input of a search location,

such that searching for documents related by said keyword to said first document,

thereby retrieving a second document;

determining relevance of said second document to at least said first keyword;

adding a reference to said second document in a results directory.

10. (previously presented) The method according to claim 9, wherein said storage is at least in part on a different computer than the computer storing said first directory.

11. (previously presented) The method according to claim 9, wherein deriving a plurality of keywords and determining relevance to a plurality of keywords.

12. (previously presented) The method according to claim 9, wherein accessing a plurality of documents in said first directory.

13. (previously presented) The method according to claim 9, with the additional steps of deriving a plurality of keywords and ranking at least two said keywords.

14. (previously presented) The method according to claim 9, with the additional step of signifying the relevancy of said second document to documents in the first directory when displaying said results directory.

15. (previously presented) The method according to claim 9, with the additional step of checking enablement of said augmentation.

16. (previously presented) The method according to claim 9, with the additional step of comparing the relevance of said second document to a preset threshold for determining said augmentation.

17. (previously presented) The method according to claim 9, wherein said results directory is said first directory.

18. (previously presented) An apparatus for augmenting a directory without user input, said apparatus comprising:

means for accessing at least a first document via a first directory, said first document comprising at least in part topical textual content;

means for deriving at least one first keyword indicative of at least one topical content within said first document;

means for searching documents in storage in at least one computer,

wherein at least some said documents are independent and not organized in relation to one another,

wherein said search means comprising searching for documents related by said keyword to said first document;

means for retrieving a second document resultant from said search means;

means for determining relevance of said second document to at least said first keyword;

means for adding a reference to said second document in a results directory;

means for displaying said directories.

19. (previously presented) The apparatus according to claim 18, wherein said storage comprises a plurality of computers connected to at least one network.

20. (previously presented) The apparatus according to claim 18, with additional means for deriving a plurality of keywords

and means for determining relevance of said second document to a plurality of keywords.

21. (previously presented) The apparatus according to claim 18, with additional means for comparing the relevance of said second document to a preset threshold for determining said augmentation.

22. (previously presented) The method according to claim 18, wherein said results directory is said first directory.

23. (previously presented) A method for augmenting a directory without user input, said method comprising the following steps:

accessing a plurality of grouped documents;

deriving a plurality of keywords indicative of the aggregate content of said grouped documents;

prioritizing the relative relevance of said keywords;

storing said keywords and said relevance prioritization;

searching a plurality of documents in storage in at least one computer,

such that searching for documents related by at least one said keyword to said stored keywords,

whereby retrieving a second document;

determining relevance of said second document to said plurality of stored keywords;

adding a reference to said second document in a results directory.

24. (previously presented) The method according to claim 23, with the additional step of comparing the relevance of said second document to a preset threshold for determining said augmentation.

25. (canceled)

26. (canceled)

27. (previously presented) The method according to claim 9, wherein said storage comprises a plurality of computers connected to at least one network.

28. (previously presented) The method according to claim 23, wherein said storage comprises a plurality of computers connected to at least one network.

29. (previously presented) The method according to claim 9, with the additional step of displaying said results directory.

30. (canceled)

31. (previously presented) The apparatus according to claim 18, with additional means for not adding a reference to a retrieved document to said results directory if said retrieved document had previously been deleted from said results directory.

EXHIBIT D

Search the web Search

IRON:DOME

rapid-response patent interception™

Press Releases

Iron Dome petitions Patent Office for cancellation of web crawler patent

Rockville, MD (April 22, 2014) -- Iron Dome LLC is requesting the U.S. Patent Office to revoke a patent owned by Chinook Licensing that is being used to sue Facebook, Match.com, Hulu and others for having websites that make recommendations to users about items they may like. Iron Dome announces the filing today of an Inter Partes Review (IPR2014-00674) at the U.S. Patent Office requesting the cancellation of Patent No. 7,047,482. "No one should have to surrender to these lawsuits exploiting defective patents," remarked Steven Yu, M.D., J.D., principal of Iron Dome.

###

Iron Dome petitions Patent Office for cancellation of cameraphone patent

Rockville, MD (February 18, 2014) -- Iron Dome LLC, a subsidiary of RozMed LLC, is requesting the U.S. Patent Office to revoke one of e-Watch's patents that is being used to sue Apple, Blackberry, Samsung, and numerous others for their globally-popular smartphones that transmit photo images over wireless cellular networks. Iron Dome announces the filing today of an Inter Partes Review (IPR2014-00439) at the U.S. Patent Office requesting the cancellation of Patent No. 7,365,871. "Smartphone makers are forced to defend against these lawsuits exploiting defective patents, and that drives up costs so that all of us pay the price," remarked Steven Yu, M.D., J.D., principal of Iron Dome.

CIVIL COVER SHEET

The JS-44 civil cover sheet and the information contained herein neither replace nor supplement the filing and service of pleadings or other papers as required by law, except as provided by local rules of court. This form, approved by the Judicial Conference of the United States in September 1974, is required for the use of the Clerk of Court for the purpose of initiating the civil docket sheet. (SEE INSTRUCTIONS ON THE REVERSE OF THE FORM.)

I.(a) PLAINTIFFS	DEFENDANTS
CHINOOK LICENSING DE, LLC	ROZMED LLC, ET AL.
(b) County Of Residence Of First Listed Plaintiff: New Castle County, Delaware	County Of Residence Of First Listed Defendant: Fairfax County, Virginia
(c) Attorneys (Firm Name, Address And Telephone Number) Richard D. Kirk (No. 0922) Stephen Brauerman (No. 4952) Bayard, P.A. 222 Delaware Avenue, Suite 900 Wilmington, DE 19899-5130 (302) 655-5000	Attorneys (If Known)

II. BASIS OF JURISDICTION (PLACE AN "X" IN ONE BOX ONLY)

- ☐ 1 U.S. Government Plaintiff
- ☐ 3 Federal Question (U.S. Government Not a Party)
- ☐ 2 U.S. Government Defendant
- ☒ 4 Diversity (Indicate Citizenship of Parties in Item III)

III. CITIZENSHIP OF PRINCIPAL PARTIES (Place An 'X' In One Box For Plaintiff And One Box For Defendant) (For Diversity Cases Only)

	PTF	DEF		PTF	DEF
Citizen of This State	☐ 1	☐ 1	Incorporated *or* Principal Place of Business in this State	☒ 4	☐ 4
Citizen of Another State	☐ 2	☐ 2	Incorporated *and* Principal Place of Business in Another State	☐ 5	☒ 5
Citizen or Subject of a Foreign Country	☐ 3	☐ 3	Foreign Nation	☐ 6	☐ 6

IV. NATURE OF SUIT *PLACE AN "X" IN ONE BOX ONLY)*

CONTRACT
- ☐ 110 Insurance
- ☐ 120 Marine
- ☐ 130 Miller Act
- ☐ 140 Negotiable Instrument
- ☐ 150 Recovery of Overpayment & Enforcement of Judgment
- ☐ 151 Medicare Act
- ☐ 152 Recovery of Defaulted Student Loans (Excl. Veterans)
- ☐ 153 Recovery of Overpayment of Veteran's Benefits
- ☐ 160 Stockholders' Suits
- ☒ 190 Other Contract
- ☐ 195 Contract Property Liability

REAL PROPERTY
- ☐ 210 Land Condemnation
- ☐ 220 Foreclosure
- ☐ 230 Rent Lease & Ejectment
- ☐ 240 Torts to Land
- ☐ 245 Tort Product Liability
- ☐ 290 All Other Real Property

TORTS

PERSONAL INJURY
- ☐ 310 Airplane
- ☐ 315 Airplane Product Liability
- ☐ 320 Assault, Libel & Slander
- ☐ 330 Federal Employers' Liability
- ☐ 340 Marine
- ☐ 345 Marine Product Liability
- ☐ 350 Motor Vehicle
- ☐ 355 Motor Vehicle Product Liability
- ☐ 360 Other Personal Injury

PERSONAL INJURY
- ☐ 362 Personal Injury Med. Malpractice
- ☐ 365 Personal Injury Product Liability
- ☐ 368 Asbestos Personal Injury Product Liability

PERSONAL PROPERTY
- ☐ 370 Other Fraud
- ☐ 371 Truth in Lending
- ☐ 380 Other Personal Property Damage
- ☐ 385 Property Damage Product Liability

CIVIL RIGHTS
- ☐ 441 Voting
- ☐ 442 Employment
- ☐ 443 Housing/ Accommodations
- ☐ 444 Welfare
- ☐ 440 Other Civil Rights

PRISONER PETITIONS
- ☐ 510 Motions to Vacate Sentence

HABEUS CORPUS:
- ☐ 530 General
- ☐ 535 Death Penalty
- ☐ 540 Mandamus & Other
- ☐ 550 Civil Rights
- ☐ 555 Prison Condition

FORFEITURE/PENALTY
- ☐ 610 Agriculture
- ☐ 620 Other Food & Drug
- ☐ 625 Drug Related Seizure of Property 21 USC 881
- ☐ 630 Liquor Laws
- ☐ 640 RR & Truck
- ☐ 650 Airline Regs
- ☐ 660 Occupational Safety/Health
- ☐ 690 Other

LABOR
- ☐ 710 Fair Labor Standards Act
- ☐ 720 Labor/Mgmt Relations
- ☐ 730 Labor/Mgmt Reporting & Disclosure Act
- ☐ 740 Railway Labor Act
- ☐ 790 Other Labor Litigation
- ☐ 791 Empl Ref Inc Security Act

BANKRUPTCY
- ☐ 422 Appeal 28 USC 158
- ☐ 423 Withdrawal 28 USC 157

PROPERTY RIGHTS
- ☐ 820 Copyrights
- ☒ 830 Patent
- ☐ 840 Trademark

SOCIAL SECURITY
- ☐ 861 HIA (1395ff)
- ☐ 862 Black Lung (923)
- ☐ 863 DIWC/DIWW (405(g))
- ☐ 864 SSID Title XVI
- ☐ 865 RSI (405(g))

FEDERAL TAX SUITS
- ☐ 870 Taxes (U.S. Plaintiff or Defendant)
- ☐ 871 IRS Third Party 26 USC 7609

OTHER STATUTES
- ☐ 400 State Reapportionment
- ☐ 410 Antitrust
- ☐ 420 Banks and Banking
- ☐ 450 Commerce/ICC Rates/etc.
- ☐ 460 Deportation
- ☐ 470 Racketeer Influenced and Corrupt Organizations
- ☐ 810 Selective Service
- ☐ 850 Securities/Commodities/ Exchange
- ☐ 875 Customer Challenge 12 USC 3410
- ☐ 891 Agricultural Acts
- ☐ 892 Economic Stabilization Act
- ☐ 893 Environmental Matters
- ☐ 894 Energy Allocation Act
- ☐ 895 Freedom of I Information Act
- ☐ 900 Appeal of Fee Determination Under Equal Access to Justice
- ☐ 950 Constitutionality of State Statutes
- ☐ 890 Other Statutory Actions

V. ORIGIN

- ☒ 1 Original Proceeding
- ☐ 2 Removed from State Court
- ☐ 3 Remanded from State Court
- ☐ 4 Reinstated or Reopened
- ☐ 5 Transferred from another district (specify)
- ☐ 6 Multidistrict Litigation
- ☐ 7 Appeal to District Judge from Magistrate Judgment

VI. CAUSE OF ACTION

(Cite The U.S. Civil Statute Under Which You Are Filing And Write Brief Statement Of Cause. Do Not Cite Jurisdictional Statutes Unless Diversity)

Action for patent infringement under 35 U.S.C. § 101, et seq.

Injunctive and declaratory relief and for damages for patent infringement

VII. REQUESTED IN COMPLAINT

CHECK IF THIS IS A CLASS ACTION ☐ UNDER F.R.C..P. 23

DEMAND $

CHECK YES only if demanded in complaint
JURY DEMAND: ☒ YES ☐ NO

VIII. RELATED CASE(S) (See instructions)

See addendum attached hereto.

JUDGE

DOCKET NUMBERS

DATE: MAY 12, 2014

SIGNATURE OF ATTORNEY OF RECORD: /S/ STEPHEN B. BRAUERMAN (SB4952)

FOR OFFICE USE ONLY

RECEIPT # ________ AMOUNT ____________ APPLYING IFP ____ JUDGE ____________________ MAG. JUDGE ____________________

Addendum to Civil Cover Sheet

RELATED CASES	JUDGE	DOCKET NUMBERS
Chinook Licensing DE, LLC v. Scribd, Inc.	Judge Leonard P. Stark	C.A. No. 13-2078-LPS
Chinook Licensing DE, LLC v. StumbleUpon, Inc.	Judge Leonard P. Stark	C.A. No. 13-2079-LPS
Chinook Licensing DE, LLC v. Hulu, LLC	Judge Leonard P. Stark	C.A. No. 14-074-LPS
Chinook Licensing DE, LLC v. RozMed LLC, et al.	Unassigned	Filed on May 12, 2014