WebSifter: An Ontological Web-Mining Agent for E-Business

Anthony Scime and Larry Kerschberg
Department of Computer Science and Department of Information and Software Engineering
E-Center for E-Business (http://eceb.gmu.edu/)
SUNY-Brockport and George Mason University
scime@brockport.edu and kersch@gmu.edu

Key words: Domain knowledge, ontology, taxonomy, personalization, search engine.

Abstract: The World Wide Web provides access to a great deal of information on a vast array of subjects. A user can begin a search for information by selecting a Web page and following the embedded links from page to page looking for clues to the desired information. An alternative method is to use one of the Web-based search engines to select the Web pages that refer to the general subject of the information desired. In either case, a vast amount of information is retrieved. The quantity can be overwhelming, and much of the information may be marginally relevant or completely irrelevant to the user's request.

We present a methodology, architecture, and proof-of-concept prototype for query construction and results analysis that provides the user with a ranking of choices based on the user's determination of importance. The user initially designs the query with assistance from the user's profile, a thesaurus, and previously constructed queries acting as a taxonomy of the information requirements. After the query has returned its results, decision analytic methods and information source reliability information are used in conjunction with the expanded taxonomy to rank the solution candidates.

1. INTRODUCTION

On the World Wide Web so much information is now available that people must sift and winnow out the meaningful information from the plethora of accessible information. The Web is a collection of servers on the Internet, which support access by means of the hypertext transfer protocol (HTTP). The servers are identified by their domain name, which is the first part of a web page's Universal Resource Locator (URL). The web pages are stored on the server as files. The URL is the complete directory structure, or logical path, from the Web server root, the domain name, to the page's file name. Web pages, collected together around a theme, create a web site. The web site is dedicated typically to one specific topic.

Within a web site, the pages are organized in two ways simultaneously. The pages are contained in a hierarchical directory structure designed by the site webmaster. The URL reflects this directory structure. Secondly, author defined hypertext links within the pages link the site's pages in a network (web) structure. The web structure may extend beyond the site's server as links are made to pages on other servers.

On the Web, a user initiates a search by using a search engine to find documents that refer to the desired subject. This requires the user to define his information needs as a collection of keywords. Typically, the search engines return the URLs, together with a brief summary of potentially useful information. These results may be ranked by the search engine according to its assessment and ranking of the document's relevance to the web-searcher's query and intent. Unfortunately, because of the limited ability of Web search engines to capture and interpret the user's information "intent", many of the retrieved results may be irrelevant.

To improve upon the search results obtained from Web queries, we have developed a methodology, associated algorithms and a proof-of-concept prototype that results in a user-determined and experience-driven ranking of Web pages. Using a mechanism by which a user can define the information intent as an *ontology* of search keywords. This personal ontology is complemented with a standard thesaurus to accommodate possible differences between the user's terminology and the search engine keywords. Additionally, the search engine's resulting page selections are matched to a collection of page addresses previously reviewed for relevance and reliability. This process provides a rating of pages based on the user's preferences. A combination of decision analysis and database management approaches provides ranking as a function of both the page's availability and the user's decision criteria. We show that this approach improves upon the results provided by existing search engines, by additionally filtering and ranking results based upon one user's perception of his needs.

2. RELATED WORK

Systems have been designed to extract relevant information from unstructured sources such as the Web. The PHOAKS (People Helping One Another Know Stuff) system searches Usenet FAQ's to identify a consensus of Web sites valid for a domain [TERV97]. Specialized search engines and indexes have been developed for many domains [SELB95]. Search engines have been developed to combine the efforts of other engines [SELB95] and select the best search engine for a domain [HOWE97]. However, these approaches do not consider experience in previous searches.

User preferences have been addressed by establishing profiles. Agents search out Web sites on user stated interests [ACKE97, MAES94] or through the joint interests of a group of users [BALA97, MAES94]. These approaches do not consider other users' experiences with specific sites.

Some Web search engines find information by categorizing the pages in their indexes. One of the first to create a structure as part of their Web index is Yahoo!. Yahoo! has developed a classification hierarchy, which is designed to help users find information faster. This hierarchy acts as a taxonomy of the search engine index. The categories, however, need to meet the specific search requirements and then be populated with the appropriate Web pages.

Studies have shown that categories can be matched with Web pages using the very brief description associated with each Web page searched [LABR99]. However, these studies did not consider the possibility that pages could have been placed in semantically similar categories or that pages might be suitable for multiple categories. It should be possible to use a specially constructed search taxonomy of categories for a specific search problem. DynaCat provides knowledge-based, dynamic categorization of search results in the medical domain [PRAT99]. This system uses an established taxonomy as a starting point, but does not allow the user to define the taxonomy according to their understanding of their search problem. It does not attempt to evaluate the taxonomy to identify the document that provides a comprehensive solution to the problem.

Footprints is a system that provides an interaction history of past Web browsing to users [WEXE99]. New users use Footprints to assist in browsing the Web by seeing where others have 'travelled' from the current page. It allows for the evaluation of page quality using the previous user comments, but the user must evaluate these comments by reading them.

Search engine selection and query execution decisions are made by the Value-driven Information Gathering (VDIG) system [GRAS00]. VDIG users construct decision model candidate solutions. VDIG selects the queries to be made to find evidence in support of the candidates. However, the decision is limited to the candidate solutions identified in the decision model.

This paper presents an agent called WebSifter that permits the user to create categories and sub-categories in a taxonomy representing their individual information problem. As a user is developing the search taxonomy she consults the taxonomy store for suggestions based on the work of previous web-searchers. During search taxonomy construction the thesaurus is consulted to find semantically similar categories. The resulting taxonomy is populated with the appropriate Web pages found by multiple search engines. A page reliability store keeps track of Web pages visited and rated by previous web-searchers. These quality ratings are automatically used in the WebSifter ranking process. WebSifter's ranking process gives more weight to Web pages matching multiple categories within the taxonomy and considers both the organization of the taxonomy and the relative weights of categories in the ranking process. This combination of decision-analytic and Web-retrieval processes results in the top ranked web page being the most comprehensive solution to the information problem from the available web pages found by multiple search engines.

3. DECISION ANALYSIS TECHNIQUES

3.1 Multiattribute Utility Technology

Multiattribute Utility Technology (MAUT) is a method used to select the best decision from a set of alternatives. The decision criteria are well defined, and are applied to each alternative independently of the other alternatives. MAUT involves the decomposition of the concepts contained in the criteria into component parts. This decomposition process creates a hierarchical taxonomic structure from the criteria concept, or root node, through component nodes to the leaf nodes [ADEL92].

Following completion of the taxonomy, each node is then visited and the weight of the node is determined. Node weight is relative to the other nodes at the same level in the hierarchy. The weight of the node is a measure of the attribute's, or component's, worth to the overall problem and it's solution. The value of a node is determined by a utility function, which adds the weighted values of a node's children to obtain its value.

Formally, this is expressed as:

$$u_i = \sum_{j=1}^{n} w_{ij} u_{ij} \qquad [1]$$

where u_i is the utility at node *i*;
w_{ij} is the relative weight of node *j*, a child of *i* (there are *n* children of *i*);
u_{ij} is the utility value at node *j*, a child of *i*.

This summation method is applied from the leaf nodes up through each node in the hierarchy. The summation method permits high weights on some nodes to compensate for low weights on other nodes. Nodes that are not important can be weighted at 0 to remove them from consideration in the overall utility of the taxonomy branch.

A method is needed to find the utility value of each leaf node attribute for each alternative. Complete evaluation of the taxonomy results in a best alternative from among the candidates.

3.2 Repertory Grid

The repertory grid is a process that compares alternatives to decision elements [SHAW87]. It measures the relationship of each alternative between ideal and actual states for selected attributes common to the alternatives. The attributes are combined to determine the overall most favorable alternative.

First the alternatives, or the possible choices for the decision, are identified. These are the candidate alternative solutions to a problem, from which the most favorable is selected. A process outside the repertory grid analysis completes the selection of the alternatives. For example, the keyword search of the Web results in pages, each having the keyword as an attribute.

Decision constructs are selected. These are the important considerations for deciding which single alternative is selected over the others. For example, three objects (*x*, *y*, *z*) may have as attributes *C1* and *C2*. These alternatives can be evaluated among themselves based upon their relationship along constructs composed of values for *C1* and *C2*. The construct poles represent the extremes of the construct; the best (*C1'*) and worst (*C1''*) possible attribute values. Best and worst values are dependent on the nature of the attribute. The construct line is scaled in a manner representative of the possible attribute values.

An analytical technique is applied to combine the constructs so as to achieve a single value for each alternative. One such technique is to multiply the attribute values of each construct for each alternative.

Formally, this is expressed as:

$$p_a = \prod_{n=1}^{m} v_{aCn} \qquad [2]$$

where

p_a is the product of alternative a;

v_{aCn} is the value of alternative a on construct Cn;

m represents the number of constructs.

Using this technique, the user's selection of the scale for the constructs influences the relative importance of the constructs to each other. A mapping of the attribute value to the construct numeric value is necessary to complete the computation. The largest resulting product identifies the overall most favorable alternative from among the candidate alternatives.

3.3 Web Page Example

Web pages may be placed on two different repertory grid constructs. They may be evaluated based on the page's structure and position in the host server's file structure. They may also be placed on a construct measuring the page's content of keywords. These constructs may then be combined to create the repertory grid.

Users seek information in their searches. Some pages provide this information directly, while other pages may act only as directories that point to additional pages. This leads to a construct line Directory Hit – Direct Hit to which Web pages may be classified, as depicted in Figure 1.

Evaluation of a page's type may be determined by a syntactic review of the structure of the page's URL and title. The URL generally provides an indication of the page's relative importance in the hierarchy of the page's parents. The scale is ordered 1 to 5. The Web pages are placed on the scale relative to the closeness the page exhibits in the construct, independent of the keywords that returned the pages.

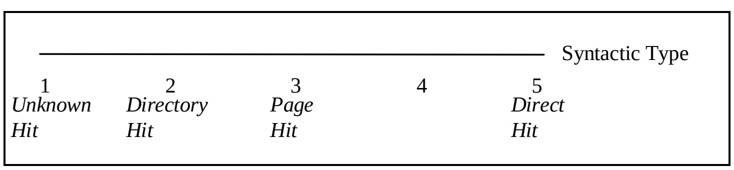

Figure 1. Web Page Type Construct

We define criteria for Direct Hit, Page Hit, and Directory Hit, and we develop heuristics to classify a Web page along the construct (Table 1).

A second construct line can be devised based on the relationship between a search's keywords and the description of the web page in the search engine's index. The construct line is No Description/Keyword Match – Perfect Description/Keyword Match (Figure 2).

Table 1. Syntactic Evaluation Rules

Syntactic Rules	Example
Direct Hit	
1. If there is a word in the page title that is also a string in the URL before the domain name (i.e., .com, .gov, .mil, .us, .ca, .au, etc.). This rule does not apply if the word is "www".	Ergonomic Office Chairs from Harter http://www.harter.com/
2. If the domain name is the last group of characters in the URL.	Boring Business Systems Home Page http://www.boring.com/
3. If "home" is a word in the title or a string in the URL. Example: The word "home" appears in the title, or as part of the URL.	Heritage Office Furnishings, Victoria, BC. Office furniture http://www.coastnet.com/home/heritage/
Page Hit	
1. If there is a word in the title that is also a string in the URL after the domain name.	Bargain Pencils custom printed for promotional use http://www.extra-mile.com/pencil/bargain.htm
2. If the URL does not end in ".htm" or ".html".	Free Outlook Express Stationery by CloudEight Stationery, Ma... http://thundercloud.net/stationery/
3. If a numeric digit is between 5 and 12 places from the end of the URL.	PENCILS - MECHANICAL http://www.workstuff.com/ksu0458.htm
4. If the string "pg" occurs in the URL. This is an indication the Web page is part of a larger Web site.	Promotional products t-shirts caps pens mugs premiums http://www.aridzone.com.au/pgarments.html
Directory Hit	
If one of the words "directory", "add," "ads," "classified," "sponsors," "members," "mall," "index," or "menu" appears as a string in the URL, as a word in the title, or as a word in the summary.	Pens, Fine pens, collectors pens, refills, ink, fountain pens, roller. http://www.joon.com/index.html

This semantic score is a measure of the number of nodes along the taxonomic query hierarchy, which match a page description, and the more nodes in common, the greater the page's value.

No Description/Keyword Match	1	2	3	4	5	6	7	8	9	10	*Perfect Description/Keyword Match*

Figure 2. Description/Keyword Construct

To mitigate preference being given to longer branches, dimensional analysis is used to convert the number of nodes into an ordinal ranking [MART95]. Since more nodes are considered a better score, the relative ranking is determined by rescaling the values of each sub-criterion on a scale of 1 to 10, with 10 being the best score.

$$r_{alternative} = 10 - \{[(v_{alternative} - v_{max}) / (- v_{max})] * 10\} \qquad [3]$$

where:

$r_{alternative}$ is the rank of the page,

$v_{alternative}$ is the number of nodes in common with the page,

v_{max} is the number of nodes in the branch.

3.4 The Decision-Making Process

The specification of objectives, parameters, and probabilities; the retrieval and management of data; and the generation of decision alternatives are important steps in the decision-making process [MEAD84]. The selection of search keywords and the creation of a problem hierarchy provide a specification of the problem’s objectives and parameters from the user’s viewpoint. This specification is based on the goals and biases of the decision-maker as well as the nature and environment of the problem.

Using web-mining the decision-maker retrieves the information available concerning a problem and its possible solutions. This retrieval is accomplished by the execution of specification keyword queries against a search engine index. The returned pages are evaluated and ranked according to the decision-maker’s understanding of, and biases toward, the problem.

4. ARCHITECTURE FOR WEBSIFTER AGENT

Figure 3 shows the architecture for the WebSifter agent and its integration into a World Wide Web distributed data source configuration. The system performs fourteen steps from elicitation of the problem taxonomy to the final ranking of Web pages.

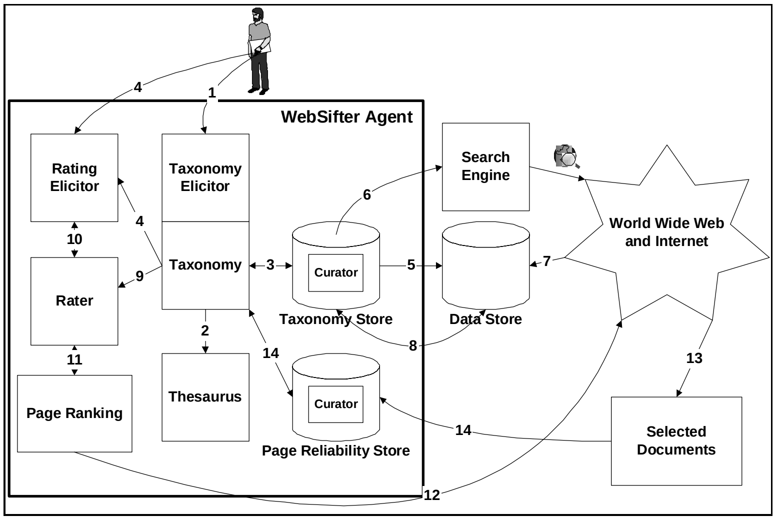

Figure 3. WebSifter Agent System Architecture

1. **Elicitation**: The user interacts with the system to begin building the problem taxonomy.
2. **Thesaurus Review**: The terms in the user's taxonomy are compared with the system's thesaurus. The user selects appropriate synonyms to add to the taxonomy.
3. **Taxonomy Review**: The taxonomy is compared with taxonomies stored as a result of previous searches successfully completed by the system. The user modifies the taxonomy by adding or pruning branches.
4. **Elicitation of Node Weights:** The user is asked to assess the importance of each of the taxonomy nodes. This importance is expressed as an integer weight relationship between the node and its parent node.
5. **Establish Data Store Structure**: The completed taxonomy becomes the data structure for the Web pages that will be retrieved by the search engine. As the Web page meta-data is retrieved, it is organized according to the stored taxonomy data structure.
6. **Search the Web**: A Web search agent uses the taxonomy leaf nodes as keywords. The agent searches multiple search engine indices performing the queries. The queries are completed and results returned.
7. **Populate the Data Store**: The results of the Web searches are used to populate the taxonomy data store. This extends the taxonomy to include the returned URLs to pages. Pages that are attributed to more than one taxonomy node have a greater potential of belonging to the solution.
8. **Prune the Taxonomy**: Some query keyword search terms may not retrieve documents in which case the taxonomy branches are pruned.

9. **Assess Page Relevance**: The pages are analyzed to determine the page's overall value in response to the problem. This analysis contains two parts. The syntactic evaluation to determine the relationship of the page to other pages on the Web. The semantic evaluation considers the page's value in response to the user's problem as stated in the taxonomy.
10. **Rate the Pages**: The node weights elicited from the user are applied to the pages. The combining of the node weights, the reliability confidence factors, and the evaluation score results in a utility rating for each page.
11. **Rank the Pages**: Based on the final utility rating score, the pages are rank ordered according to the user's preferences stated in the taxonomy.
12. **Selection of Desired Page**: From the ordered list, the user selects a page to visit. The first page listed provides the most valuable information.
13. **Page Returned**: The selected page returns a document from the Web for the user's review.
14. **Assign Page Reliability**: The page reliability is a function of its semantic relevance to the problem and the reliability of the Web site from which the page was obtained. The reliability of the Web site is captured in a reliability store and is based in user feedback of site reliability.

4.1 The Search and WebSifter Process

A modified repertory grid technique is used to evaluate the taxonomy, based on the alternatives found and their relationship to the problem. The repertory grid provides both problem-independent and problem-dependent dimensions. The problem-independent dimension is based on a page's internal characteristics and location in the World Wide Web. The problem-dependent dimension is derived from the user's representation of the search requirements as a taxonomy. The repertory grid values are then applied to the taxonomy and evaluated using Multiattribute Utility Technology (MAUT) to rank the available solutions.

Conceptually, the agent works as shown in Figure 4. First, a taxonomy is constructed consisting of nodes (N) and weights (W). The taxonomy has a root node (NR), two intermediate nodes (N1 and N2), and six keyword nodes (KW1 - KW6), and the last keyword node has two synonyms (KW6-syn1 and KW6-syn2). The taxonomy is developed by the user as a representation of the problem to be solved. Each of the nodes (NR, N_i, and KW_j) has an associated weight.

The Web is a collection of sites containing pages (P). The pages are catalogued in search engine indices. A search using each of the keywords or keyword synonyms in the taxonomy is conducted. For example, KW3 retrieves pages P5 through P9. Some pages in the search engine index (P1 - P4) do not match any of the keywords in the taxonomy. Other pages are on the World Wide Web but are not catalogued (P19 and P20). Those pages not found are not considered further in the analysis.

Each returned page has an assigned reliability determined by the agent's reliability store, or it is assigned a default value if not in the store. For example, P5's reliability is R5, P7's R7, etc. The documents also undergo a repertory grid analysis as discussed previously (RG5 denotes the repertory grid analysis for page P5). The page rank is then computed as the product of the taxonomy weights from the root to the keyword, the reliability value, and the repertory grid analysis. For P5 the rank is W1 * W13 * R5 * RG5.

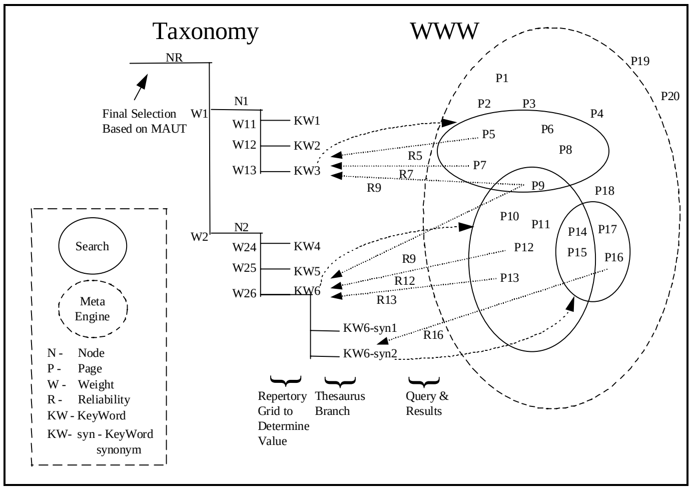

Figure 4. Retrieval and Analysis Architecture

When a keyword has synonyms, a search is conducted on each synonym as well as on the keyword. The synonyms are rated in the same way as the keywords. The synonyms do not have a weight assigned; they use the corresponding keyword's weight.

Pages may be found by more than one search. If a page is found by a keyword and its synonyms, or by two synonyms of the same keyword, the final scores may be different. The repertory grid semantic score varies with the keyword synonym. In this case, the highest score for this page is accepted. If a page is found as a result of different branches on the taxonomy (the case of P9), the scores from each branch will be different. Such a page has more value because it meets more criteria assigned in the taxonomy. It is assigned a higher final score by adding the scores of the branches. The calculation of all the final scores results in a ranking of the pages found and the identification of the top page.

4.2 Illustrative Example

Consider, a search to find available office equipment companies. The first step is to construct the taxonomy, which has a root node, "office equipment," several intermediate nodes, and several leaf nodes. The root node is office equipment. Specifically, office equipment suppliers are sought who provide various types of office equipment. The intermediate nodes and leaf nodes specify categories and specific types of office equipment. Figure 5 provides the complete taxonomy with user assigned branch weights. We may want to further restrict the search to companies located in Florida.

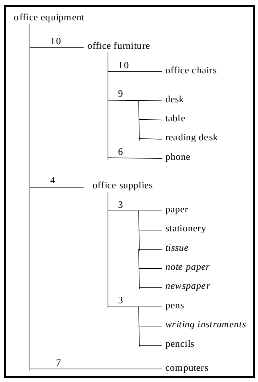

Figure 5. Office Equipment Taxonomy

In this search a page for OfficeDirect.com was found. The company's home page is titled "Office Chairs Local Office Furniture Listings By State Dealers By Area Code Online Office Furniture Locator Find" with a URL of http://www.chairsearch.com/. This page will be referred to as Chairsearch in the remainder of this paper. Using the syntactic rules in section 3 this page is a "direct hit" by direct hit rule 1, a "direct hit" by direct hit rule 2, and a "page hit" by page hit rule 2. By averaging the individual scores, this page scores 4.333 on the syntactic page type index. Using the dimensional analysis described in section 3 on the Chairsearch page, as returned from the office chairs search, a semantic score of 4.0 is calculated.

To determine a page's overall relevance to a user's search request, the product of the syntactic and semantic scores is computed. The resulting repertory grid score for the Chairsearch, the product of the syntactic score (4.3333) and the semantic score (4.0), is 17.333.

After the repertory grid score has been determined, the page's final score is determined based on the node weights, repertory grid score, and reliability factor. This results in a final score for the Chairsearch example of 867.

In the Florida office equipment search, the syntactic and semantic scores combined with the taxonomic evaluation create a ranking that reorders and combines the search engine results (Table 2). Searches were done using as keywords the nodes of the branches of the Office Equipment taxonomy; that is for office chairs: Florida AND business AND "office equipment" AND "office furniture" AND "office chairs". The original rank is recorded in Table 2 for the top ten WebSifter results.

Table 2. Florida - Business - Office Equipment Results

Original Rank	**Branch Leaf Node**	**New Rank**	**Total Score**	**Name**	**URL**
1	Office chairs	1	867	Office Chairs Local Office Furniture Listings By State Dealers By Area Code Online Office Furniture Locator Find	http://www.chairsearch.com/
1 1	Phone Computers	2	859	Boring Business Systems Home Page	http://www.boring.com/
7	Phone	3	780	BBB Serving Central Florida Member Directory O	http://www.orlando.bbb.org/members/O1.html
4	Phone	4	660	SBA: FLORIDA Business Cards -- Part 14	http://www.sba.gov/buscard/states/busfl14.html
27 5	Desk Office Chairs	5	633	HyperMart Business Directory	http://hypermart.net/dir/General_Business/Consumer_Services.html
2 16 15	Office chairs Phone Desk	6	600	Mead-Hatcher, Inc. Florida Distributor/Dealer List	http://www.meadhatcher.com/fl.php3
21 4 2	Desk Office chairs Desk (table)	8	570	Yahoo! Business and Economy>Companies> Office Supplies and Services>Furniture	http://www.yahoo.co.uk/Business_and_Economy/Companies/Office_Supplies_and_Services/Furniture
39 22	Phone Desk (table)	8	570	Printing & Office Equipment Marketplace	http://www.equipmentsource.com/printing.html
6	Desk	10	540	Advanced Business Services - Titusville, FL	http://www.nbbd.com/abs/
13	Desk	10	540	Chapter 4 - Handbook on Business Procedures	http://www.admin.ufl.edu/division/fa/HB_C04P3

5. CONCLUSIONS

Our prototype WebSifter Agent ranks the candidate solutions to a query based on the user query specification as well as the results returned by the data sources. This agent would coexist with other agents and knowledge rovers [BROD99; KERS99; KERS97a; KERS97b; MAES97; MAES94] in a federation of heterogeneous distributed databases. The knowledge rovers consult the search engines on the Web. After the knowledge rovers have returned results, the WebSifter reviews them.

The WebSifter agent is intelligent in the application of the user's current and previous preferences. The agent possesses a profile of previous query structures developed by users. These are used to develop a solution taxonomy. The agent also gains intelligence regarding the reliability of the data source through user feedback, which it applies to future decisions.

The lowest level nodes of the taxonomy are evaluated using the relationship between sibling pages on the taxonomy leaf node. Pages are evaluated against their siblings using the repertory grid technique.

A unique contribution of this paper is the syntactic and semantic rating of returned pages. A method was developed to classify Web pages independent of the page subject. The second part of rating pages is a semantic analysis of the usefulness of a page in terms of its response to the query. The two rating methods are brought together with the problem taxonomy to evaluate and rank the list of possible choices returned by a query.

ACKNOWLEDGEMENTS

This research was sponsored, in part, by a grant from the Virginia Center for Innovative Technology to the E-Center for E-Business at George Mason University.

REFERENCES

[ACKE97] Ackerman, M., Billsus, D., Gaffney, S., Hettich, S., Khoo, G., Kim, D. J., Klefstad, R., Lowe, C., Ludeman, A., Muramatsu, J., Omori, K., Pazzani, M. J., Semler, D., Starr, B. and Yap, P.; "Learning Probabilistic User Profiles Applications for Finding Interesting Web Sites, Notifying Users of Relevant Changes to Web Pages, and Locating Grant Opportunities;" AI Magazine 18(2), 1997; pp. 47 - 56.

[ADEL92] Adelman, L.; Evaluating Decision Support and Expert Systems; New York; John Wiley & Sons; 1992.

[BALA97] Balabanovic, M. and Shoham, Y.; "Fab: Content-Based, Collaborative Recommendation;" Communications of the ACM 40(3), 1997; pp 66 - 72.

[BROD99] Brodsky, A., Kerschberg, L. and Varas, S.; "Resource Management in Agent-based Distributed Environments," in Cooperative Information Agents III, vol. 1652, Lecture Notes in Artificial Intelligence, M. Klusch, O. Shehory, and G. Weiss, Eds. Berlin, et al.: Springer-Verlag, 1999, pp. 50-74.

[GRAS00] Grass, J. and Ziberstein, S.; "A Value-Driven System for Autonomous Information Gathering;" Journal of Intelligent Information Systems; Vol 14, No. 1, March 2000, pp. 5 – 27.

[HOWE97] Howe, A., and Dreilinger, D.; "Savvy Search A Metasearch Engine That Learns Which Search Engines to Query;" AI Magazine 18(2), 1997; pp. 19 - 25.

[KERS97a] Kerschberg, L.; "Knowledge Rovers: Cooperative Intelligent Agent Support for Enterprise Information Architectures," in Cooperative Information Agents, vol. 1202, Lecture Notes in Artificial Intelligence, P. Kandzia and M. Klusch, Eds. Berlin: Springer-Verlag, 1997, pp. 79-100.

[KERS97b] Kerschberg, L.; "The Role of Intelligent Agents in Advanced Information Systems," in Advances in Databases, vol. 1271, Lecture Notes in Computer Science, C. Small, P. Douglas, R. Johnson, P. King, and N. Martin, Eds. London: Springer-Verlag, 1997, pp. 1-22.

[KERS99] Kerschberg, L. and Banerjee, S.; "An Agency-based Framework for Electronic Business," in Cooperative Information Agents III, vol. 1652, Lecture Notes in Artificial Intelligence, M. Klusch, O. Shehory, and G. Weiss, Eds. Berlin, et al.: Springer-Verlag, 1999, pp. 254-279.

[LABR99] Labrou, Y. and Finin, T.; "Yahoo! as an Ontology - Using Yahoo! Categories to Describe Documents;" Proceedings of the Eight International Conference of Information Knowledge Management (CIKM '99), Kansas City, Missouri, 1999; pp. 180 - 187.

[MAES94] Maes, Pattie; "Agents that Reduce Work and Information Overload;" Communications of the ACM 37, 7 Jul. 1994; pp. 30 – 40.

[MAES97] Maes, Pattie; "Intelligent Software;" Proceedings of the 1997 International Conference on Intelligent User Interfaces; 1997; pp. 41 – 4.

[MART95] Martin, M.; Analysis and Design of Business Information Systems; 2nd Ed.; Prentice-Hall; Englewood Cliffs; 1995.

[MEAD84] Meador, C.L., Guyote, M. J., Keen, P.G.W.; "Setting Priorities for DSS Development;" MIS Quarterly; Vol 8, No. 2; June 1984; 117 – 129.

[PRAT99] Pratt, Wanda, Hearst, Marti, and Fagan, Lawerence; "A Knowledge-Based Approach to Organizing Retrieved Documents;" AAAI-99: Proceedings of the Sixteenth National Conference on Artificial Intelligence. July 1999.

[SELB95] Selberg, E. and Etzioni, O. (1995). "Multi-Service Search and Comparison Using the MetaCrawler;" Proceedings of the 4th International World Wide Web Conference, Boston, Massachusetts, 195 - 208.

[SHAW87] Shaw, M. L. G., and Gaines, B. R.; "KITTEN: Knowledge Initiation and Transfer Tools for Experts and Novices;" Int. J. Man-Machine Studies; September 1987.

[TERV97] Terveen, L., Hill, W., Amento, B., McDonald, D. and Creter J.; "A System for Sharing Recommendations;" Communications of the ACM 40(3), 1997; pp. 59 - 62.

[WEXE99] Wexelblat, Alan and Maes, Pattie; "Footprints: History-Rich Tools For Information Foraging;" Proceeding of the CHI 99 Conference on Human Factors in Computing Systems: The CHI is the Limit, 1999, pp. 270 – 277.