

# Learning from the World Wide Web: Using Organizational Profiles in Information Searches

Anthony Scime  
State University of New York  
College at Brockport

[ascime@brockport.edu](mailto:ascime@brockport.edu)

## Abstract

*The World Wide Web provides access to a great deal of information on a vast array of subjects. Individuals search the Web as part of their organizational responsibilities to keep informed and to help solve business problems. These searches for information are accomplished by selecting a Web search engine and developing a keyword query. Keyword queries tend to return many irrelevant hits and typically each search is conducted independently of queries performed at an earlier date or queries performed by others. The searches have no organizational context beyond the personal knowledge of the searcher. This lack of organizational context limits the ability of the search to use the experience of other searches and does not contribute to organizational learning. Presented in this paper is an architecture for query construction and results refinement using organizational experience and knowledge. A searcher initially designs a Web query with assistance from an organizational thesaurus and previously constructed queries. After the query has returned results, organizational historical knowledge on source quality assists in determining sufficient quality results to assist in problem solution.*

**Keywords:** information retrieval, organizational learning, knowledge management

## Introduction

Decentralized and individual solutions for unstructured problems are typical in many organizations. However, finding solutions often involves a duplication of effort in the solution discovery process. In addition, the solution discovery process seldom uses previous experience. Formalizing decentralized and individualized methods that provide support from current organizational knowledge about similar unstructured problems reduces the duplication of effort and creates a learning environment. Such a learning environment requires a supporting infrastructure that allows individuals to extract learned information from organizational repositories. Repository information also disseminates problem knowledge between closely related problem domains (Henninger et al, 1995).

In the learning organization, the repositories provide knowledge and grow as new problems are solved. The use of prior

knowledge is a beginning point for the collection of new information. Effort is saved by not collecting 'known' information but focusing on the retrieval of value-added information.

Populating the repository needs to be continuous and quick in order to promote a learning environment. In addition, newly retrieved and previously stored information must be continually assessed for usefulness in solving the current problem (Harvey et al, 1997). With a repository focused on learning support, some structure is provided for the unstructured problems.

## Background

With most knowledge management systems, it is the responsibility of the searcher to know about, access and retrieve the documents needed to complete a task. The information in the documents is generally confined to the document - there is no exchange of knowledge between documents other than by the searcher's processing. In some cases, database structures have been overlaid on document retrieval/knowledge management systems to provide a knowledge base within an organization (Liongosari et al, 1999). This knowledge base provides a source for obtaining organizational knowledge. However, knowledge can only be discovered if the organization has specifically built the data structure to accommodate the searcher's need.

---

Material published as part of this journal, either on-line or in print, is copyrighted by the publisher of Informing Science. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact [Editor@inform.nu](mailto:Editor@inform.nu) to request redistribution permission.

There are systems that 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 (Terveen et al, 1997). Specialized search engines and indexes have been developed for many domains (Selberg and Etzioni, 1995). Search engines have been developed to combine the efforts of other engines (Selberg and Etzioni, 1995) and select the best search engine for a domain (Howe and Dreilinger, 1997). However, these approaches do not consider the user's experience in previous searches.

User preferences have been addressed by establishing profiles. Agents search out Web sites on user stated interests (Ackerman et al, 1997) or through the joint interests of a group of users (Balabanovit and Shoham, 1997). These approaches do not consider the organization's environment or other users' experiences with specific Web 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! (<http://www.yahoo.com>). Yahoo! has developed a hierarchy of documents, which is designed to help users find information faster. This hierarchy acts as a taxonomy of the search engine index. Yahoo! helps by directing the searcher through the levels of the taxonomy. The searcher sees document titles and summaries, which may be indexed by keywords that may not have been used in the original search. Such an organization of the contents of the Web into categories could help in an organization's search. 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 (Labrou and Finin, 1999). But, these studies did not consider the possibility that pages could have been placed in semantically similar categories or that pages may be suitable for multiple categories. No consideration was given to the structural organization of the categories. It should be possible to use a specially constructed search taxonomy of categories for a specific search problem. Web pages, based on their summaries, can then be placed on the taxonomy.

A problem taxonomy can assist in Web query construction. There is a specific manner in which expert Web searchers go about searching using browsing and search engines. After making an initial search using a single keyword and one search engine, forty-seven percent of the time, the expert searcher begins to browse (Hoelscher and Strube, 1999). Browsing is done by going from page to page using the links in the pages. This time-consuming approach points out the need for a search method before the search engine is used.

In addition, other factors may impact the search process. Query design can influence the results. Domain experience can improve search results by the proper selection of keywords and phrases. An organizational knowledge repository may assist in the design of the search queries. The formation of queries can be determined by the selection of keywords previously used and stored in the knowledge repository.

## Organizational Searches

Current Web search methods do not directly consider the organizational context of the search. In an organizational setting, individuals perform searches of the Web to obtain information to assist in decision-making and problem solving. The search process requires information collection and evaluation. Collection and evaluation activities need to be conducted within the context of the individual's organization and related to organizational problems and preferences. In order to provide for organizational learning and improve upon the search results obtained from Web queries, a learning search architecture is needed. This architecture would develop queries, which fit within the organization's context of the problem-solving/information-collection process as well as the organization's experience with individual sites.

Conceptually, any search is a problem. The problem solution technique is to develop an information query and analyze the quality of the results. The searcher formulates the query against available data sources using the appropriate approaches. This query is then processed to yield solution options. The searcher uses some decision method to rank the query results and select an option.

In the Web search process, an individual determines the factors important to the search problem. These factors are based on the searcher's biases as well as the nature and organizational environment of the problem. These factors are expressed as keywords. Most searchers limit the keyword search to one or two keywords and search with a single search engine (Hoelscher and Strube, 1999). The search typically ranks the resulting hits. Ranking brings to the top of the results list the Web page that most closely matches the searcher's understanding of and biases toward the problem. However, the limited keywords and resulting ranking process restrict the search results; thus eliminating the discovery of related information.

A common problem-solving process is to break a problem into smaller parts, find information about each part and then combine the information into a solution (Adelman, 1992). By using Multiattribute Utility Technology (MAUT), a search problem can be decomposed into its component parts. This process creates a taxonomy that defines the search problem and facilitates the solution. The downside to using this approach is that the creation of a problem's taxonomy can be a time consum-

ing process. Selection of branch subtopics and sub-subtopics requires a certain level of knowledge in the problem domain.

The use of a taxonomy creates a better understanding of the problem thus resulting in more appropriate solutions being found. Partial solutions are fitted in the taxonomy framework, which defines the partial solution's relationship to the problem. The taxonomy can reduce the number of candidate solutions by considering overlaps between branches in the solution set. The effort spent in taxonomy creation is saved by not having to evaluate as many possible solutions.

However, the creation of a problem's taxonomy can be a time consuming process. Selection of branch subtopics and sub-subtopics requires a certain level of knowledge in the problem domain. Significant construction effort is saved if taxonomy pieces can be reused from previous problems. This also reduces the level of domain knowledge required.

After the searches are conducted, the search results are graded based on previous organizational experience. A grade is assigned to each page (e.g., A for a perfect match) based on its appropriateness to the current selection criteria. Some of the pages found during the current search may belong to sites found by previous searches. The organization's repository would already contain grades for these pages and this information would assist in the ranking process. (Ranking is the placing of Web pages in order from the highest to lowest grade.) With the vast amount of available information and the limited time to come to a solution, it is critical that the number of pages be limited to those useful in formulating a decision.

## Organization Search Assistant (OSA)

An organization faced with unstructured problems can improve the process by using an Organization Search Assistant (OSA) to help employees. The Organization Search Assistant methodology extends the problem taxonomy by applying the organization's previous knowledge of the search domain to the current search taxonomy. As new taxonomy nodes are identified and added, a general organizational thesaurus suggests additional nodes. Query taxonomies, developed previously by the organization, provide ideas in the development of the problem search taxonomy.

The Organizational Search Assistant finds related information in organizational repositories and then searches the Web. It ranks the Web pages returned based on intelligence concerning the quality of the Web sites obtained through organizational user feedback. These organizational preferences are applied to re-rank the results. OSA's goal is to provide the minimum amount of information needed to reach a decision.

## Organization Search Assistant System Architecture

The OSA System Architecture (Figure 1) shows the Organization Search Assistant and its role in an organization repository and a World Wide Web search. A query undergoes 14 steps from elicitation of the domain taxonomy through helping learn about the problem domain and helping to make a decision.

**Elicitation:** The searcher is questioned by the system to build the problem domain taxonomy (step 1). This taxonomy is the initial definition of the problem. The initial taxonomy is based on the searcher's most fundamental knowledge of the problem. OSA asks the searcher for a general topic on which to base the search. OSA continues to query the searcher using a depth-first approach for more details; a break down of the meaning of the last term provided. The searcher may stop the initial taxonomy construction at any time. However, the deeper the taxonomy, the greater specificity possible when checking the taxonomy store and when searching the Web.

**Thesaurus Review:** The terms in the problem domain taxonomy are compared with a thesaurus the organization maintains to expand searches (step 2). The thesaurus increases the breath of the initial taxonomy's end nodes. The thesaurus contains a collection of words and synonyms. Initial seeding with common English words such as found in Roget's Thesaurus starts the collection. The organization can add to the thesaurus any words or synonyms specific to its business domain. The OSA learns the terminology of the business as it adds new words or synonyms to the thesaurus.

**Taxonomy and Repository Review:** The taxonomy is compared with stored taxonomies (step 3) to provide the searcher with suggestions based on the organization's view of the domain. These stored taxonomies are the results of previous searches completed by the organization; that is, information learned by the organization. The stored taxonomies provide the searcher with an additional refinement of the problem definition. Selected portions of the stored taxonomies result in a final taxonomy for the problem as an extension of the initial taxonomy. The final taxonomy sub-topics are defined in a manner that avoids confusion as to the intent of the sub-topic. This preciseness allows for accurate measurement of a document's applicability to the sub-topic.

Each node from the initial taxonomy is used as a search keyword in querying the taxonomy repository. Each query that finds a keyword in the repository returns a taxonomy with the keyword as the root. It also returns a Web page title, uniform resource locator (URL), and previously assigned grade for each returned taxonomy node that has a representative Web page.

The repository can be initialized with taxonomies already in existence, such as the Yahoo! categories. This initialization would not include Web page information until the organization had completed at least one search with a taxonomy segment to allow for grading of the found pages.

The searcher executes two functions at this point. He (or she) visits Web pages that may solve the problem and end the search. A review of previously visited Web pages (step 4) may find some that help in resolving the problem. These pages are returned (step 5) from the Web. The searcher reviews each page (step 6) and he (or she) assigns a grade that is used to recalculate the site's Grade Point Average (GPA) (step 7). The GPA is the average of the page grades for a site, where 'A' is a 4, 'B' a 3, 'C' 2, 'D' 1, and 'F' 0. As OSA is used in an organization, the repository will grow and the likelihood of finding the appropriate pages without searching the Web increases.

If the taxonomy repository Web pages do not satisfy the searcher's needs, the searcher quickly reviews the taxonomy repository taxonomies. Taxonomy segments are selected that make a meaningful addition to the initial domain taxonomy.

**Store Problem Taxonomy:** The final domain taxonomy is stored and becomes available to the organization for future use in taxonomy development (step 8). Because the content of the World Wide Web is constantly changing, Web pages become dated and even disappear. As a result, pages listed in the repository may not be useful over the long term. However, the taxonomy developed as part of the business may still be valid.

From the construction of previous taxonomies the novice searcher is learning about the business. He (or she) is also contributing new knowledge by the storage of his (or her) taxonomies. As time goes by, the stored taxonomies make future searches more comprehensive without adding significant effort on the part of the searcher. In the longer view, increased taxonomies in the repository will result in more accurate and quicker searches.

**Search the Web:** If the taxonomy repository's known Web sites does not provide sufficient information for a decision, the searcher develops a keyword query for each of the leaf-node sub-topics in the taxonomy using keywords, which represents the sub-topic concept (step 9). The queries may be a single keyword, a collection of keywords, a string, or a com-

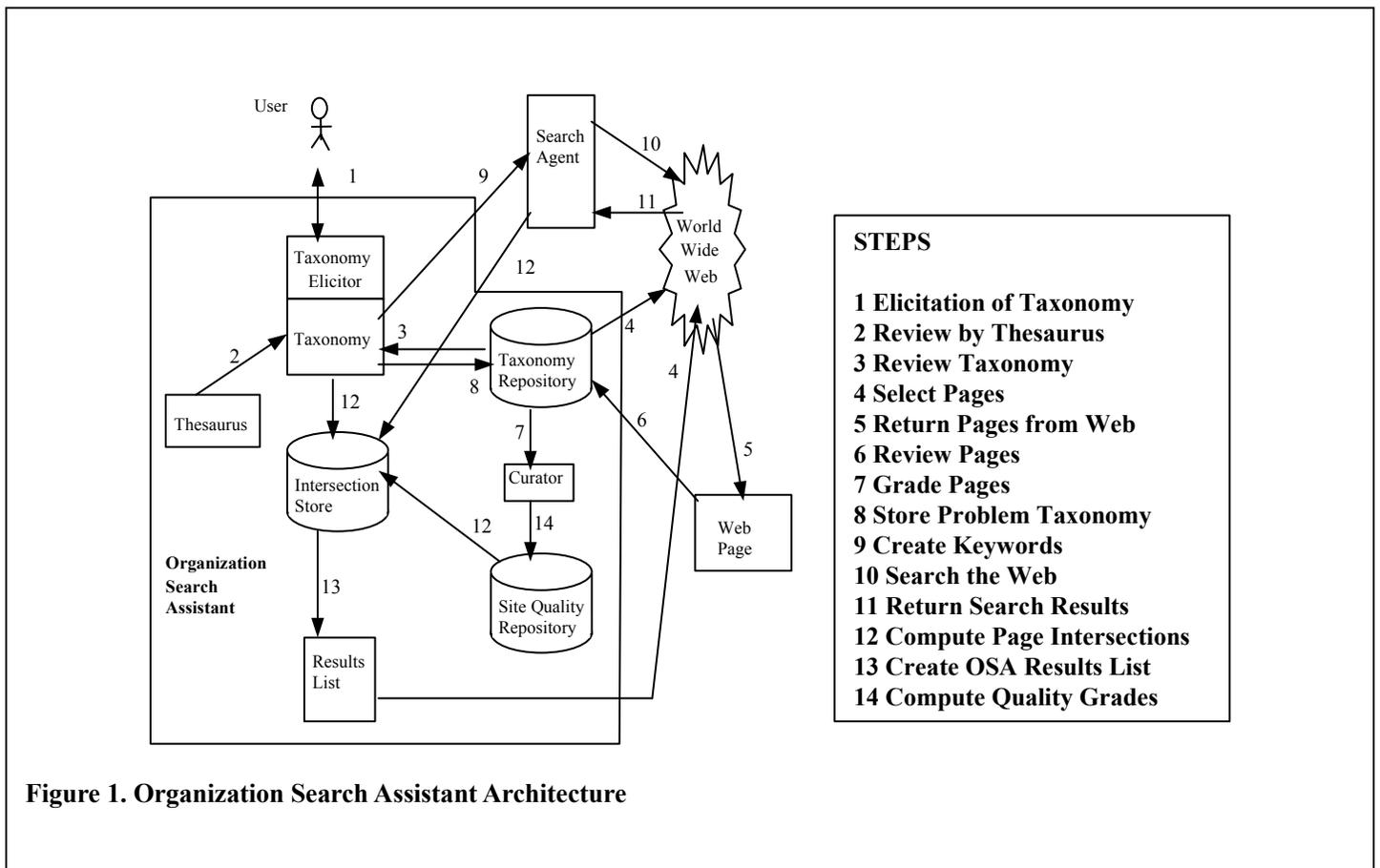


Figure 1. Organization Search Assistant Architecture

combination of keywords and strings. Although a sub-topic

may have a specific meaning in the context of the searcher's need, the use of a keyword or string could lead to the retrieval of many irrelevant sites. Therefore, keywords and strings are constructed to convey the meaning of the taxonomy branch. This increases the specificity of the retrievals.

Because no search engine completely indexes the Web, it is best to use more than one search engine (Selberg and Etzioni, 1995). Queries of the Web are completed using an intelligent search agent. Intelligent search agents are designed to assist users in Web searching and can access multiple Web search engines using the same keywords (step 10). These agents simultaneously use multiple Web search engines in the search, remove duplicates, and store the results on the user's computer. The agent eliminates the need to reformat a search query for different search engines. Copernic99 (<http://www.copernic.com/>) is one such application and is used by OSA. Copernic99 provides access to the search engines AltaVista, Deja.com, Excite, HotBot, Infoseek, Lycos, Magellan, WebCrawler and Yahoo! It returns a ranked list and a relevancy score for each page. The top 10 results from each engine are returned and any duplicates removed.

**Return Search Results:** The results of the Web searches are returned to the search agent and used to populate the taxonomy (step 11). Organizational knowledge about the Web sites is now applied to create the results list identifying those pages from which the searcher is likely to derive the most benefit.

**Create the Results List:** The returned search lists are compared with each other and the site quality repository. OSA computes the intersection of the results sets and the quality repository in the context of the taxonomy (step 12). The OSA results list is generated by ranking the pages (step 13).

The searcher has reached this point because he has not found a sufficient number of satisfactory Web pages already in the taxonomy repository. The Web page URLs returned from the searches are compared to each other. Each time a page was a hit on a search, it is assigned one point. Site quality repository sites are included in this match; providing those pages one additional hit point. This means new pages need to be on the search retrieval list of at least three nodes to overcome the advantage of being previously found. The purpose in providing this advantage is that the sites in the site quality repository have a quality GPA, which assists the searcher in evaluating the page unseen. The page with the greatest number of hit points for each taxonomy node is returned on the OSA results list. Summing the corresponding searches' Copernic99 rankings and taking the lowest total breaks ties. Copernic99 ranking is also used when a branch's results list has no intersections. Pages of sites with a GPA below 2.0 are noted on the OSA results list and an alternate page is also listed.

Pages may be returned which the searcher has already visited because the page was in the taxonomy repository. OSA has determined this is the best page for these taxonomy nodes. If the searcher is not satisfied with the page they may give it a lower grade and select another page from the search engine results for the node in question.

**Pages Selected and Returned:** The searcher selects (step 4), retrieves (step 5) and reviews (step 6) the pages listed on the OSA results list. Finding specific Web pages that will help solve the searcher's problem returns the time spent creating the taxonomy. The immediate return is the time not spent reviewing long search result's lists or navigating from page to page in a site attempting to find part of an answer. In addition, the taxonomy behaves as a framework for the problem and provides for the integration of the partial solutions.

**Assign and Store Quality Grades:** As the pages are reviewed, a quality grade is assigned. The pages are evaluated with respect to the value of the information they contain. The grade is stored in the taxonomy repository with the taxonomy (step 6). The resulting taxonomy graded pages and the re-evaluated pages from the review of previously found pages are used to recalculate the site GPA and are sent to the curator (step 7) for GPA storage (step 14). The site of a page is the domain name portion of the uniform resource locator (URL). Over time, a site's grade point average becomes more refined. Future searches benefit from their predecessor's experience with specific Web sites by estimating the quality of pages at the site.

The organization learns from the Organization Search Assistant's taxonomy and site quality repositories about its types of problems, how the problems are defined, problem solution information, and that information's quality. Knowledge about valuable and not so valuable Web site's is estimated based on the quality feedback used in evaluation of retrieved pages. This Web site grade point average is maintained in the site quality repository and used in future evaluations. As more Web searches are conducted the taxonomy and site quality repositories grow becoming more useful to the organization and its employees.

## Conclusion

An architecture has been presented for assisting in solving unstructured problems by organizationally evaluating and ranking information within the organization and from the Web. A prototype that illustrates this approach can be found in the appendix. Using the Organization Search Assistant, a searcher elicits and refines his or her information needs, determines appropriate organizational search keywords, checks organizational knowledge, and analyzes the results returned by a Web search. This is done using knowledge learned from previous searches, while providing a refinement of that

knowledge and adding to the organization's knowledge repositories.

In the process, the searcher's keywords are intelligently expanded using an organizational thesaurus. As part of search taxonomy development, the searcher calls upon previous organizational taxonomy profiles. These taxonomy profiles assist the searcher in problem decomposition by providing the results of the organization's previous view of similar problems. During the analysis, knowledge is applied from previous experience about site and page quality as a factor in recommending sites to review. In the end, the number of pages for the searcher to review is significantly reduced, yet adequate to assist with the business problem.

## References

- 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. (1997). 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), 47 - 56.
- Adelman, L. (1992). *Evaluating Decision Support and Expert Systems*. New York, NY: John Wiley & Sons.
- Balabanovic, M. and Shoham, Y. (1997). Fab: Content-Based, Collaborative Recommendation. *Communications of the ACM* **40**(3), 66 - 72.
- Harvey, M. G., Palmer, J. and Speier, C. (1997). Intranets and Organizational Learning, in *Proceedings of the 1997 Conference on Computer Personnel Research*, San Francisco, California, 110 - 116.
- Henninger, S., Lappala, K and Raghavendran, A. (1995). An Organizational Learning Approach to Domain Analysis, in *Proceedings of the 17<sup>th</sup> International Conference on Software Engineering*, Seattle, Washington, 95 - 104.
- Hoelscher, C. and Strube, G. (1999). Searching on the Web: Two Types of Expertise, in *Proceedings of SIGIR '99*, Berkeley, California, 305-306.
- Howe, A., and Dreilinger, D. (1997). Savvy Search A Metasearch Engine That Learns Which Search Engines to Query. *AI Magazine* **18**(2), 19 - 25.
- Labrou, Y. and Finin, T. (1999). Yahoo! as an Ontology - Using Yahoo! Categories to Describe Documents, in *Proceedings of the Eight International Conference of Information Knowledge Management (CIKM '99)*, Kansas City, Missouri, 180 - 187.
- Liongosari, E. S., Dempski, K. L., and Swaminathan, K. S. (1999). In Search of A New Generation of Knowledge Management Applications. *SIGGROUP Bulletin* **20**(2), 60 - 63.
- Selberg, E. and Etzioni, O. (1995). Multi-Service Search and Comparison Using the MetaCrawler, in *Proceedings of the 4th International World Wide Web Conference*, Boston, Massachusetts, 195 - 208.
- Terveen, L., Hill, W., Amento, B., McDonald, D. and Creter J. (1997). A System for Sharing Recommendations. *Communications of the ACM* **40**(3), 59 - 62.

## Appendix

### Using the Organizational Search Assistant

The loan department at an investment bank is in the business of helping businesses get started by providing loans. Jack, a loan analyst, receives a business plan for a loan to start a chain of outdoor equipment retail stores in eastern Kentucky and Tennessee. The business plan contains information on the location of the store, the local market for outdoor equipment, the expected sales, potential customers, the store's management and current financial situation. It is Jack's task to evaluate and verify the statements in the business plan. The bank will not loan money to ventures that are not likely to succeed.

Jack, who grew up in the big city, is experienced in finance, accounting, management and sales, but knows nothing about outdoor equipment or the outdoor recreation opportunities in the Mid-Appalachian Mountains. However, the bank has previously made loans to endeavors in that area. Jack checks the bank information repositories and the Web to learn more about outdoor equipment, outdoor recreation, and the Appalachian Mountains.

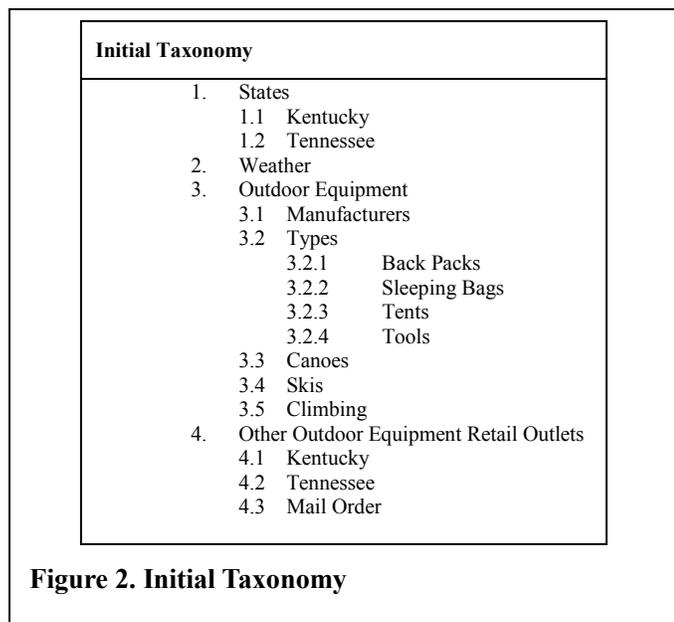
Jack's problem is unstructured; he may find part of the solution in many places. Because the bank is a learning organization, it maintains information repositories for such problems. To find information not in the bank's repositories, Jack turns to the Word Wide Web.

Jack's bank uses an Organization Search Assistant (OSA) to help employees solve unstructured problems. OSA's goal is to provide the minimum amount of information Jack needs to reach a decision. In addition, the OSA helps the bank learn about the problem domain and Jack make a decision on this loan application.

Jack begins with the search problem taxonomy in Figure 2. OSA's thesaurus review then finds that Tools can be expanded to Axes, Saws, Knives, Lanterns, and Stoves. The comparison of this initial taxonomy with stored taxonomies provides Jack with suggestions based on the organization's view of the domain. OSA provides Jack with an already completed taxonomy for each of the two states and a break down of the Mail Order branch. As an example, Figure 3 shows stored taxonomies for Kentucky and Mail Order along with Web page titles, addresses, and grades. Grades for pages are the result of previous reviews by Jack or his co-workers.

Jack reviews the taxonomy repository list and selects Web pages that may help in resolving his problem. As Jack reads the Web pages he assigns a grade that is used to re-calculate

the site's Grade Point Average (GPA). He finds that the page on the Forest (Figure 3, Branch 1.2) no longer exists. Jack further decides that the Appalachian Mountain (Figure 3, Branch 1.3) Web page's grade is low and that it may not be useful. Finally, he decides the Mail Order Web pages (Figure 3) are not specific enough to his problem, but adds the branches to his taxonomy, hoping to find better results.



Tennessee's stored taxonomy is also consulted and found similar to Kentucky's taxonomy. Jack adds branches from Tennessee - Geography, Kentucky - Geography and Mail Order to his taxonomy. These are refinements to his search problem that may yield valuable results. Jack stores this improved taxonomy in the taxonomy repository for future use by himself or his co-workers.

Not yet having sufficient information for a decision, Jack develops a keyword query for each of the leaf-node sub-topics in the taxonomy using keywords, which represent the sub-topic concept (Figure 4). By using a combination of keywords for each sub-topic Jack increases the specificity of the retrievals. For example, the "'Outdoor Equipment' - Catalog' search returns more results specific to catalogs of outdoor equipment (20 of 29 pages) than simply 'Catalog' alone, which would return no addresses for outdoor equipment catalogs.

Jack's searches of the Web by Copernic99 returned the addresses of 1,435 Web pages in 25 results sets. Jack does not have time to review all 1,435 pages of which 1,211 are unique. However, OSA's eliminating duplicate addresses, looking for pages that satisfy multiple searches, and applying

organizational knowledge about the Web sites, results in only 14 pages that need to be reviewed (Figure 5).

OSA's results show that four branches share one Web page - <http://www.outdoorreview.com/techtalk/>. Three branches share a page and a site from the quality repository. The site <http://www.northernmountain.com> was found in the quality repository. Apparently, this site was retrieved as part of another (and unknown) problem. It appears on Jack's results list because the site intersects with three intersecting groups of branches. A page of the site <http://www.gorp.com> was in the taxonomy repository as a reviewed site of Kentucky - Economy - Camping. It is now also used to relate States - Kentucky - Forest and States - Kentucky - River. The site <http://www.buckskin.org>, also in the quality repository, intersects with the 'Weather' branch. This site has a GPA of 1.76, therefore a second page is also provided to Jack.

These 14 pages (along with the pages found in the taxonomy repository) should provide Jack sufficient background to help with the loan decision. These sites were not necessarily selected because they are the best, but because they should be good enough for Jack's purpose and significantly reduce his reviewing effort. Jack now selects, retrieves, and reviews the 14 pages selected by OSA and, along with previous organizational knowledge and information from the business plan, completes his report.

As Jack reviews the pages, he assigns a quality grade (Figure 5). He evaluates the pages with respect to the value of the information they contain. The grade is stored in the taxonomy repository with the taxonomy for future use, and in calculating the GPA. For example, The GPA for <http://www.northernmountain.com> is calculated from the grades for page <http://www.northernmountain.com/catindex.html> and page <http://www.northernmountain.com/links.html>, as these pages are from the same site.

Through the use of OSA, Jack has retrieved sufficient information to make a decision on this particular loan. In addition, he has added to the bank's knowledge base about Kentucky, Tennessee, outdoor equipment and the Appalachian Mountains. He has extended the taxonomy store by his representation of his search problem. He has made decisions about Web page quality and assisted in evaluating Web site quality. All this newly created and discovered knowledge is now available for use by Jack or others in solving different bank problems.

Kentucky	Page Title	Page Address	Grade
1. Geography			
1.1 Land	Land Office	<a href="http://www.sos.state.ky.us/ADMIN/LANDOFF1/landoff2.htm">http://www.sos.state.ky.us/ADMIN/LANDOFF1/landoff2.htm</a>	A
1.2 Forest	Department of Forestry	<a href="http://www.uky.edu/Agriculture/Forestry/annual.htm">http://www.uky.edu/Agriculture/Forestry/annual.htm</a>	A
1.3 Appalachian Mountains	The Commonwealth of Kentucky	<a href="http://www.uky.edu/KentuckyAtlas/kentucky.html">http://www.uky.edu/KentuckyAtlas/kentucky.html</a>	D
1.4 Rivers	Rivers of Kentucky	<a href="http://www.nationalrivers.org/states/ky-menu.htm">http://www.nationalrivers.org/states/ky-menu.htm</a>	C
2. People	The People of Kentucky	<a href="http://www.audubon-area.com/1tsld002.htm">http://www.audubon-area.com/1tsld002.htm</a>	D
3. Economy			
3.1 Natural Resources	Welcome to the Kentucky Natural Resources Geographic Information System Homepage	<a href="http://www.nr.state.ky.us/nrepc/ois/gis/">http://www.nr.state.ky.us/nrepc/ois/gis/</a>	B
3.2 Forestry	Forestry in Kentucky	<a href="http://www.webcom.com/duane/duanefor.html">http://www.webcom.com/duane/duanefor.html</a>	C
3.3 Mining	MOUNTAINTOP MINING	<a href="http://www.mountaintopmining.com/">http://www.mountaintopmining.com/</a>	B
3.4 Tourism			
3.4.1 Amusement Parks	Center for Hospitality, Tourism, and Travel Administration: Hospitality Today	<a href="http://www.sce.nyu.edu/dyncon/hosp/hosp_indu_rec_r.html">http://www.sce.nyu.edu/dyncon/hosp/hosp_indu_rec_r.html</a>	C
3.4.2 State Park	Kentucky State Parks - KY State Parks Homepage	<a href="http://www.kystateparks.com/">http://www.kystateparks.com/</a>	A
3.4.3 National Park	Recreation.Gov	<a href="http://www.recreation.gov/">http://www.recreation.gov/</a>	A
3.4.4 Camping	GORP Kentucky Travel Tourism and Recreation Guide	<a href="http://www.gorp.com/gorp/location/ky/ky.htm">http://www.gorp.com/gorp/location/ky/ky.htm</a>	C
3.4.5 Hiking	Eastern Kentucky Hiking Trails	<a href="http://www.gorss.com/eastern_kentucky_hiking_trails.htm">http://www.gorss.com/eastern_kentucky_hiking_trails.htm</a>	C
3.5 Industry	Kentucky: Economy, Industry, Business and Labor	<a href="http://www.uofl.edu/library/ekstrom/govpubs//states/kentucky/kyecon.html">http://www.uofl.edu/library/ekstrom/govpubs//states/kentucky/kyecon.html</a>	B
4. Government	Alphabetical List of Kentucky State Government Information	<a href="http://www.state.ky.us/govtinfo.htm">http://www.state.ky.us/govtinfo.htm</a>	A
5. Education	Welcome to the Kentucky Department of Education!	<a href="http://www.kde.state.ky.us/">http://www.kde.state.ky.us/</a>	A
6. Health Care	National Hospital : Local Health Care : Kentucky : Hospitals	<a href="http://www.nationalhospital.com/localhealth/kentucky/hospitals.htm">http://www.nationalhospital.com/localhealth/kentucky/hospitals.htm</a>	C
7. Culture			
7.1 Folk Art	Kentucky Folk Art	<a href="http://partners.net/English/publications/Newsletters/June%201998/folkart.htm">http://partners.net/English/publications/Newsletters/June%201998/folkart.htm</a>	A
8. History			
8.1 Early	Early History of Warren County Kentucky	<a href="http://www.wku.edu/Library/onlinexh/rrr1/Pages/Mainpages/earlyhis.html">http://www.wku.edu/Library/onlinexh/rrr1/Pages/Mainpages/earlyhis.html</a>	B
8.2 Statehood	About Kentucky	<a href="http://mrmike.com/relocate/aboutky.htm">http://mrmike.com/relocate/aboutky.htm</a>	D
8.3 Civil War	Kentucky and the Civil War: KET EdWeb Site index page	<a href="http://www.ket.org/Education/civilwar/">http://www.ket.org/Education/civilwar/</a>	B
8.4 Reconstruction	HQ Reference: NMA History--the Reconstruction Era	<a href="http://www.nmaonline.com/hq/reference/hist_recon.html">http://www.nmaonline.com/hq/reference/hist_recon.html</a>	D
8.5 20th Century	Hazard Kentucky & Perry County: A Photographic History	<a href="http://hazardkentucky.com/">http://hazardkentucky.com/</a>	C
<b>Mail Order</b>	<b>Page Title</b>	<b>Page Address</b>	<b>Grade</b>
1. Catalog	The Catalog Site! Browse our award-winning site for the perfect catalogs to fit your shopping needs.	<a href="http://www.catalogsite.com/">http://www.catalogsite.com/</a>	A
2. On Line	New Wave Mail Order - Your One Stop On-line Source for Gaming!	<a href="http://www.newwave.org/">http://www.newwave.org/</a>	B
3. On the Web	National Mail Order Association Web Links	<a href="http://www.nmoa.org/links_public.htm">http://www.nmoa.org/links_public.htm</a>	B
4. e-commerce	How To Combine E-Commerce With Mail Order For Greater Profits	<a href="http://www.dealconsulting.com/technology/mail.html">http://www.dealconsulting.com/technology/mail.html</a>	C

Figure 3. Taxonomy Repository Entries for Kentucky and Mail Order

Outdoor Equipment in Appalachia	Keywords
1. States	
1.1 Kentucky	
1.1.1 Forest	Kentucky Forest
1.1.2 Appalachian Mountains	Kentucky "Appalachian Mountains"
1.1.3 River	Kentucky River
1.2 Tennessee	
1.2.1 Forest	Tennessee Forest
1.2.2 Appalachian Mountains	Tennessee "Appalachian Mountains"
1.2.3 River	Tennessee River
2. Weather	"Appalachian Mountains" Weather
3. Outdoor Equipment	
3.1 Manufacturers	"Outdoor Equipment" Manufacturer
3.2 Types	
3.2.1 Back Packs	"Outdoor Equipment" "Back Packs"
3.2.2 Sleeping Bags	"Outdoor Equipment" "Sleeping Bags"
3.2.3 Tents	"Outdoor Equipment" Tents
3.2.4 Tools	
3.2.4.1 Axe	"Outdoor Equipment" Axe
3.2.4.2 Saw	"Outdoor Equipment" Saw
3.2.4.3 Knife	"Outdoor Equipment" Knife
3.2.4.4 Lantern	"Outdoor Equipment" Lantern
3.2.4.5 Stove	"Outdoor Equipment" Stove
3.3 Canoe	Canoe
3.4 Skis	Skis
3.5 Climbing	"Appalachian Mountains" Climbing
4. Other Outdoor Equipment Retail Outlets	
4.1 Kentucky	"Outdoor Equipment" Kentucky
4.2 Tennessee	"Outdoor Equipment" Tennessee
4.3 Mail Order	
4.3.1 Catalog	"Outdoor Equipment" Catalog
4.3.2 On Line	"Outdoor Equipment" On-Line
4.3.3 On the Web	"Outdoor Equipment" Web
4.3.4 e-commerce	"Outdoor Equipment" e-commerce

Figure 4. Completed Taxonomy and Keywords

Jack's Taxonomy Branch	Web Address	Jack's Grade
Outdoor Equipment - Types - Sleeping Bags Outdoor Equipment - Types - Tents Outdoor Equipment - Types - Tools - Saw Outdoor Equipment - Types - Tools - Stove	<a href="http://www.outdoorreview.com/techtalk/">http://www.outdoorreview.com/techtalk/</a>	D
Other Outdoor Equipment Retail Outlets - Mail Order - Catalog Other Outdoor Equipment Retail Outlets - Mail Order - On Line Other Outdoor Equipment Retail Outlets - Mail Order - On the Web Quality Repository	<a href="http://www.northernmountain.com/catindex.html">http://www.northernmountain.com/catindex.html</a>	A
States - Kentucky - Forest States - Kentucky - River Quality Repository	<a href="http://www.gorp.com/gorp/location/ky/ky.htm">http://www.gorp.com/gorp/location/ky/ky.htm</a>	B
States - Kentucky - Appalachian Mountains States - Tennessee - Appalachian Mountains Outdoor Equipment - Climbing	<a href="http://peakbagger.com/range/appalac.htm">http://peakbagger.com/range/appalac.htm</a>	C
Outdoor Equipment - Types - Tools - Knife Outdoor Equipment - Types - Tools - Lantern Quality Repository	<a href="http://www.northernmountain.com/gearinfo/packcheck.html">http://www.northernmountain.com/gearinfo/packcheck.html</a>	C
Other Outdoor Equipment Retail Outlets - Kentucky Other Outdoor Equipment Retail Outlets - Tennessee Quality Repository	<a href="http://www.northernmountain.com/links.html">http://www.northernmountain.com/links.html</a>	B
States - Tennessee - Forest States - Tennessee - River	<a href="http://users.multipro.com/cumberlandtrail/index.html">http://users.multipro.com/cumberlandtrail/index.html</a>	A
Weather Quality Repository	<a href="http://www.breathitt.k12.ky.us/">http://www.breathitt.k12.ky.us/</a> (GPA: 1.76) <a href="http://uswrp.mmm.ucar.edu/uswrp/PDT/four/App.html">http://uswrp.mmm.ucar.edu/uswrp/PDT/four/App.html</a>	D B
Outdoor Equipment - Manufacturers Outdoor Equipment - Types - Back Packs	<a href="http://www.macscouter.com/General/Outfitters.html">http://www.macscouter.com/General/Outfitters.html</a>	C
Outdoor Equipment - Canoe	<a href="http://www.canoe.ca/">http://www.canoe.ca/</a>	D
Outdoor Equipment - Types - Tools - Axe	<a href="http://www.jmw.net/virtual_mall/mntzone.html">http://www.jmw.net/virtual_mall/mntzone.html</a>	A
Outdoor Equipment - Skis	<a href="http://www.mtn.co.uk/gearguides/nordicski_skis.html">http://www.mtn.co.uk/gearguides/nordicski_skis.html</a>	A
Other Outdoor Equipment Retail Outlets - Mail Order - e-commerce	<a href="http://www.hoxie.org/guy.htm">http://www.hoxie.org/guy.htm</a>	C

Figure 5. OSA Web Search Results