A Distributed Multi-Agent System for Collaborative Information Management and Sharing

James R. Chen & Shawn R. Wolfe

NASA Ames Research Center

Mail Stop 269-2

Moffett Field, CA 94035-1000

{jchen, shawn}@ptolemy.arc.nasa.gov

Stephen D. Wragg

QSS Group, Inc., at

NASA Ames Research Center

Mail Stop 269-2

Moffett Field, CA 94035-1000

stephen@ptolemy.arc.nasa.gov

ABSTRACT

In this paper, we present DIAMS, a system of distributed, collaborative agents to help users access, manage, share and exchange information. A DIAMS personal agent helps its owner find information most relevant to current needs. It provides tools and utilities for users to manage their information repositories with dynamic organization and virtual views. Flexible hierarchical display is integrated with indexed query search to support effective information access. Automatic indexing methods are employed to support user queries and communication between agents. Contents of a repository are kept in object-oriented storage to facilitate information sharing. Collaboration between users is aided by easy sharing utilities as well as automated information exchange. Matchmaker agents are designed to establish connections between users with similar interests and expertise. DIAMS agents provide needed services for users to share and learn information from one another on the World Wide Web.

Keywords

Agent, bookmark, collaboration, information management, learning, World Wide Web

1. INTRODUCTION

The Internet revolution has made a wealth of information resources available for direct and easy access on the user's desktop. However, finding appropriate information has become a significant problem for many users. Organized information spaces are easier to search, but finding or authoring these are difficult. Our research focuses on three areas that require significant technological advances: (1) finding information relevant to users' needs; (2) organizing information for facilitating access in various contexts; and (3) collaborative information management and learning.

Current WWW search engines allow users to locate information of interest, but often return vast amount of irrelevant information. On-line centralized catalogs (often called portals) such as Yahoo provide more relevant and well-organized information, but are not always suitable for individual users needs. Personalized catalogs like My-Yahoo can be customized by individual users, but provide limited capacities and cannot support information sharing between users. More recent information discovery and filtering technologies attempt to provide relevant information to users by learning from their previous queries or from other users' queries and feedback [1, 12]. Yet users need an easy way to access information relevant and adapted to their current task and interest at any time.

Once relevant information is found, pointers to it must be locally organized and stored in a manner that allows rapid and effective access for both individuals and workgroups. Current personal information organizing schemes on the WWW are mostly limited to bookmarks (also called hotlists, or favorites). Bookmarks provide an easy way to organize URLs in a hierarchical manner, and to attach personal comments to them. Although clearly superior to unstructured lists, hierarchical folder organization forces users to think in terms of a neatly decomposable structure consisting of disjoint clusters of related URLs. However, a single piece of information is often relevant in multiple ways, and thus is not easily categorized within a single folder. We conjecture that no single static structure will be appropriate in all contexts. With hierarchical schemes, navigational access to information can be tedious and frustrating when information is nested several layers deep. Therefore current bookmarking schemes are monolithic, can be tedious to navigate, and cannot be easily shared with other users. Recent approaches to organize information at the level of collections of documents rely on metadata standards (W3C Resource Description Framework), which require additional authoring effort from Web pages authors, and only support contexts of use anticipated by the author.

There is also a critical need for tools supporting collaboration among distributed users with similar interests, or who are part of the same workgroup. Individual users can author and publish Web pages containing lists of related links. Some of them can be quite sophisticated, organized under single categories, or in tables with multiple categories. However it takes time to author and maintain these lists in a textual format. Sharing a common repository of information is a first step, but doesn't scale up to large distributed

and informal groups. Collaborative tools themselves need to be distributed and dynamic, and support collaborative learning and discovery of information.

In this paper, we present DIAMS, a prototype multi-agent service aimed to help users access and organize online information, and to facilitate sharing and exchange of structured information among users.

2. DIAMS

DIAMS is a system of distributed intelligent agents designed for collaborative information management, sharing and exchange on the World Wide Web (WWW). The system is designed to help a WWW user find needed information from his/her personal collection of URL links, as well as from other remote resources and/or collections of resources. Ideally, the system will find a minimal set of information most relevant to the user's current needs. In a sense, these two desirable capabilities correspond to recall and precision, the two standard measures of effectiveness in traditional information retrieval. Recall is the proportion of relevant materials retrieved, whereas precision is the proportion of retrieved materials that are relevant [16]. With the abundance of information available on the WWW, it has become increasingly more important to have information access tools that can attain good results in both measures. In practice, unfortunately, it is very difficult to achieve both high recall and high precision at the same time in most situations.

To address this problem, DIAMS focuses on accessing and sharing information in a distributed environment of personal or group information repositories, controlled and managed directly by users. DIAMS does not intend to provide WWW users with information stored in huge and complex public repositories or portals, as is done with Yahoo or search engines like Lycos. Instead, DIAMS information agents are designed to provide efficient tools for their users to manage and share high quality, well-organized local information repositories customized for individual needs. DIAMS agents are to provide information services complementary to that of existing resources on the WWW.

DIAMS provides more than services for stand-alone local information repositories; it is designed to facilitate collaborative information management, sharing and learning among distributed repositories of knowledgeable users with similar interests. Information is constantly changing, as are user needs. No single user can always maintain the most updated information links. Portals on the other hand cannot maintain the best information organization to fit all different users at different time. Systems of communicating information agents such as Jasper [3] exchange keywords and URLs, but do not communicate with structured information and knowledge. In order to support collaborative information management and sharing, DIAMS provides utilities to help users learn about and make use of other users' collections. DIAMS also provides for active pushing of useful information to other users to facilitate information exchange.

DIAMS incorporates a multi-agent architecture to help users access, organize, share and learn information on the WWW. Among several different types of information agents employed, personal agents are the ones that work most directly with users to help support the presentation, organization and management of user information collections. A DIAMS personal agent helps its owner manage their information repository with dynamic organization adaptable to current needs. Flexible hierarchical display is integrated with indexed query search to ensure effective information access. Contents of a repository are kept in object-oriented storage to facilitate information sharing. Collaboration between users is aided with matchmakers. Communication between agents is supported with automatic indexing methods in information retrieval. These components and related interface features are presented in the following sections.

3. ACCESS AND MANAGEMENT OF INFORMATION REPOSITORIES

Like most WWW browsers' bookmarking facilities, a DIAMS personal agent maintains a collection of URLs for its owner. An agent supports object-oriented organization of its information contents, and provides dynamic hierarchical display of any part of the structure. Information access and manipulation features are embedded in a friendly graphical interface. This graphical interface is also combined with powerful query search functionalities to ensure quick and effective access.

3.1 Dynamic Organization and Virtual Views

DIAMS supports dynamic hierarchical organization of a collection of information by incorporating multiple indexing categories. A DIAMS category is both a folder for storage of information contents, and an index for search and communication. A category can contain URLs and/or other categories. It can also contain external categories from collections of other personal agents. A category can be at any level of a collection hierarchy. It can be a member of several parental categories, thereby appear in multiple positions within a hierarchy. A category can even be nested within itself if needed. Users can create and edit their collections of categories and URLs. They can change and manage the structure and order of collection contents with drag and drop interaction and menu options. They can also optionally assign weights to categories or URLs with respect to their parental categories, to enhance query performance.

A category does not have to be displayed in the collection hierarchy. The hierarchical display can be narrowed or expanded. Large collections can be made more tractable by hiding portions irrelevant to the current task. Hidden sections can be easily restored. Users can view the whole or part of a collection, or combinations of sections of multiple collections. Users can name and save a particular hierarchical display as a "virtual view". They can later display or modify particular views for different usage. Actual changes to the collection folders and contents will be reflected correspondingly in these saved views.
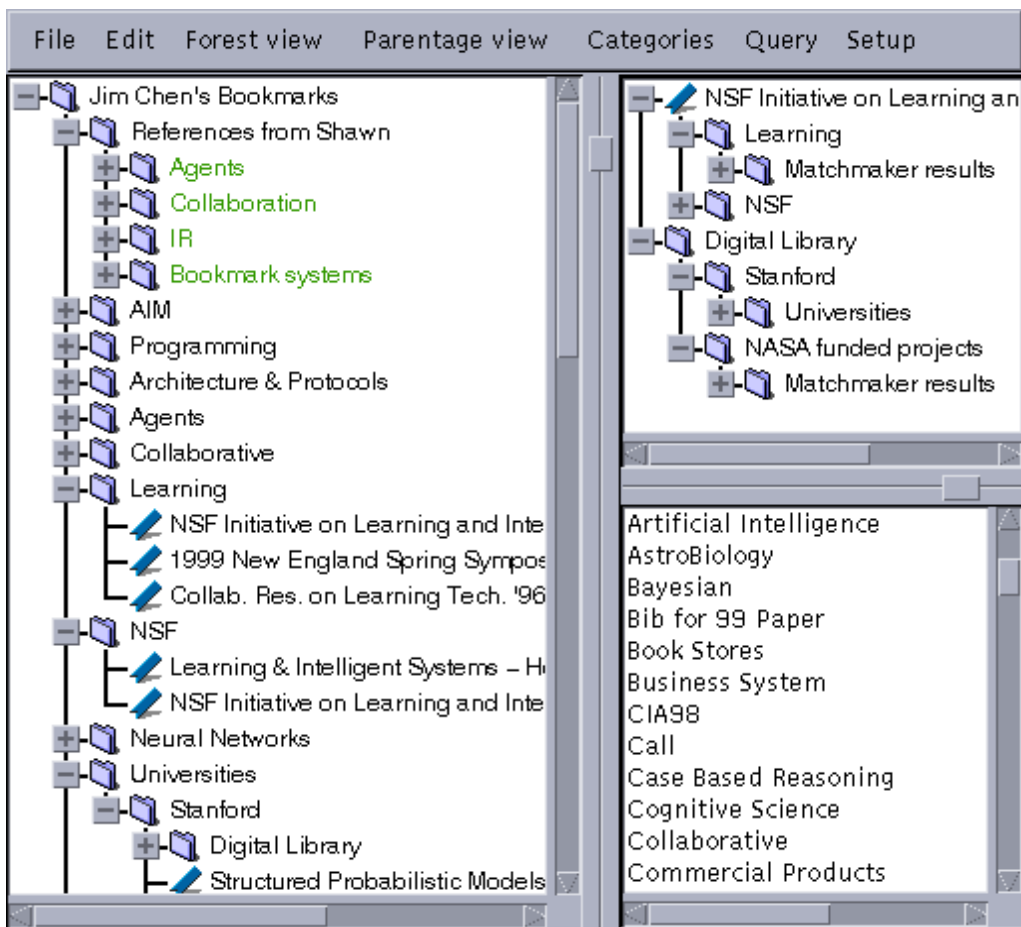
Figure 1: DIAMS User Interface

Figure 1 shows DIAMS main user interface for organizing and browsing personal and other users' collections of URLs. The left pane, Forest View, displays a current view of the contents of the owner's personal collection, which may include selected parts of other users' collections. A different color scheme can be assigned for each user's collection contents.

The upper right pane displays the Parentage View of folders or URLs selected from the left pane. The display order of the parents-children relations in the Parentage View is the reverse of that of the Forest View. Since an URL or a category may have multiple parents, its parentage view can be used to further display forest views of different parent folders. The parentage view also provides the user an easy way to create or follow links to folders owned by other users, yet categorized within this user's own collection.

The lower right pane, the Categories, displays all categories in the owner's local collection. The list can also be extended to display external categories associated with the external collections displayed on the left. This list of categories can be sorted by name or size of contents, or searched by string matching. It provides an alternative way for users to find needed categories for view or query.

3.2 Object-based Information Structure

Information structure of a DIAMS repository is object based. Although the information collection of a personal agent is customized and maintained for its owner, it is likely that a good portion of the collection is composed of links to external information objects of other DIAMS agents. Users can easily select and make connections to any combination of information objects or structures of objects from accessible parts of another repository. The capacity to incorporate existing sub-collections from other repositories promotes rapid construction of useful information collections. Sharing of information objects also helps make updating information easier and minimizes storage. As we will present in later sections, a DIAMS personal agent provides means for its user to locate and access external information repositories. It also provides utilities for the user to make easy suggestions to the owners of external repositories for possible updates if needed.

The information repository of a DIAMS personal agent is customized for its owner. This customization process is made easy by utilizing not only the user's own collection of URLs, but also a distributed collection of information objects from the existing repositories of other knowledgeable users. When used within a company or an institution, a user's DIAMS repository is typically initialized with subsets of standard collections maintained by group agents specialized in some particular information areas on

an intranet. A DIAMS user is also encouraged to incorporate information objects of DIAMS repositories on the WWW from remote experts with similar knowledge and interests.

3.3 Query and Indexed Access

Instead of browsing or searching through views of collections, users can also retrieve needed information directly through queries. Queries are composed of categories and/or index keywords.

Index keywords are text stems extracted from URL pages in a collection. Periodically a DIAMS personal agent runs a background batch process to visit URL pages in the collection, and runs an automatic indexing routine to extract index keywords from these pages. The indexing routine is composed of standard information retrieval procedure. Stopwords are filtered from the text and the remaining words are stemmed. Terms with either too high or too low frequencies of appearances in the collection are excluded from the index list. The conventional information retrieval measure of term importance, TF*IDF (term frequency and inverse document frequency) is used to weigh keywords extracted from documents. The personal agent then maintains a fixed number of keywords of highest weights associated with each category, for indexed query and between-agent communication.

DIAMS supports a number of standard term-weighting TF*IDF formulas [6]. The default within-document TF measure is

$$\frac{freq_{ij}}{maxfreq_j}$$

where

$freq_{ij}$ = the frequency of term i in document j

$maxfreq_j$ = maximum frequency of any term in document j

An alternative entropy measure [10] is used as the system's default IDF measure

$$entropy_i = 1 - \frac{\sum_{k=1}^{N} \frac{freq_{ik}}{totalfreq_i} \log_2 \frac{totalfreq_i}{freq_{ik}}}{\log_2 N}$$

where

N = the number of documents in the collection
$freq_{ik}$ = the frequency of term i in document k

$totalfreq_i$ = the total frequency of term i in the collection

To calculate the measure of term importance, a document unit needs to be defined at different levels of a DIAMS collection hierarchy, either as a category or as a leaf URL. For a URL, the page contents and the contents of pages linked from that page are considered a document. For a category, the combination of its first level contents, including subcategories, is considered a document. Index measurement of a subcategory is adjusted by its size (number of URLs) to represent a single entry within its parent category.

To calculate the inverse document frequency, the collection of documents is by default defined as the entire collection of pages maintained by a personal agent. The current implementation also supports a second algorithmic option, in which the importance measures of keywords within a category are taken with respect to its parent category, i.e., the parent category is taken as the collection in the measure calculation. The employment of this algorithmic option renders a nested hierarchy of index keywords corresponding to the hierarchy of categories. Since a personal agent can only maintain a limited number of index keywords for each category, a nested hierarchy of indices accommodates less redundant information and gives better retrieval accuracy. However, the algorithm requires larger categories with more contents to generate more accurate statistics, hence is more suitable for large and complex group agent collections.

Some URL pages contain binary code or graphical objects instead of word sequences, from which index information cannot be extracted. Although these pages are not legible to DIAMS agents, users can still categorize them. A category containing both text and binary pages will be associated with keywords extracted from the text pages. Users can also enter keywords or notes in DIAMS profiles associated with categories and/or URL pages.

A query composed of categories is translated into a set of weighted index keywords. These keywords are then used to retrieve relevant categories and/or URLs within a collection. A query can be sent to different remote agents to retrieve relevant information. A query can also be issued upon the user's own collection, in which case the personal agent will return not only the categories specified within the query, but also other related categories and URLs within the collection.

4. COLLABORATIVE INFORMATION SHARING WITH MULTI-AGENTS

DIAMS employs a multi-agent architecture to help users access, organize, share and learn information collaboratively on the World Wide Web. Several types of information agents are involved. Among them, personal agents are the ones that work most directly with users for the presentation and management of user information collections. Personal agents work closely with one another for collaborative information sharing and exchange. They also work with other types of information agents in DIAMS, which provide different kinds of services. The functionalities of these different agents and the relations among them are introduced in this section. A typical scheme of collaboration between agents is shown in figure 2.
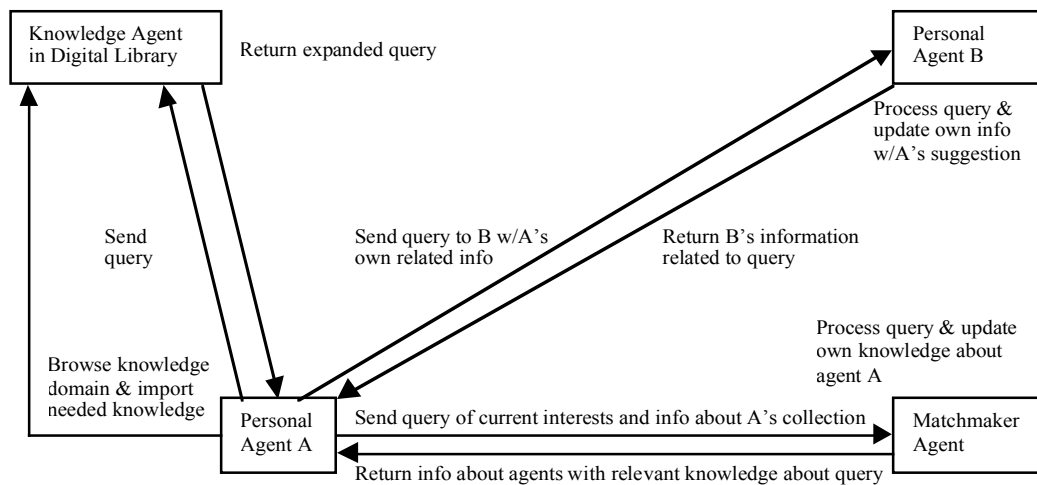
Knowledge Agent in Digital Library

Return expanded query

Personal Agent B

Process query & update own info w/A's suggestion

Send query

Send query to B w/A's own related info

Return B's information related to query

Process query & update own knowledge about agent A

Browse knowledge domain & import needed knowledge

Personal Agent A

Send query of current interests and info about A's collection

Matchmaker Agent

Return info about agents with relevant knowledge about query

Figure 2: Collaboration in DIAMS

A DIAMS user can visit other users' repositories through his/her personal agent. One can also include structured information objects from external repositories in one's own collection. Access to other users' repositories is done through "behind the scene" categories search and translation between agents. External information objects can be displayed in different colors specified by owner of local agent. Read or write protection from self, group or web can be set at the level of categories within a repository.

4.1 Information Exchange between Agents

Collaborative information sharing and learning among users is further supported by automated information exchange. Personal agents exchange information with one another. When a user query is directed to an external agent, the user's agent sends not only the query information, but also sends with the query its own query matches, i.e., its information contents related to that query. The receiver agent, in addition to responding to a query, has several options to handle extra information that comes with the query. The agent can just ignore the incoming information; it can automatically process the new information and place them into appropriate categories within its own collection; it can also keep the information in a temporary space and leave the handling decision to its owner. The user dictates which method will be used in a setup procedure. The temporary space is categorized under both a top level temporary category, and other local categories rendered most appropriate for the information contents. Thus the owner of that agent can handle incoming information of all or any particular categories of interests at any time. Incoming information that has been stored in temporary space over certain time or space limits will be removed automatically.

Since many categories of a personal agent are created by its owner, they are often not known to other agents. As described in a previous section on query and indexed access, index keywords are used to facilitate communication between agents. Each personal agent maintains a set of most useful index keywords extracted from its collection of documents. An agent also keeps track of the relations between the keywords and its categories. The keywords provide language commonality for communication between agents.

4.2 Matchmaker Agents

An important issue regarding collaborative information exchange between users is the possibility to attain knowledge about other users and the ability to find and access the most appropriate ones. A DIAMS matchmaker agent is designed to facilitate collaboration. A matchmaker maintains information about personal agents. Its internal configuration and interface functionalities are very similar to that of a personal agent. However, instead of maintaining a structured information repository about URLs, a matchmaker keeps an indexed collection of personal agents. When responding to a query, a matchmaker provides the inquirer with links pointing to other agents, which may carry information most relevant to the query.

Similar to the information exchange protocol between personal agents, communication between a matchmaker and a personal agent is also bi-directional. When communicating with a matchmaker, the inquiring personal agent brings along with the query a set of categories and keywords representing its current information collection and main interests. A matchmaker agent thereby both learns and provides information about the repositories of visiting personal agents.

A user can inquire a matchmaker with a query to look for relevant repositories. The user can then select the repositories of interests from the result list and issue the query to the remote agents. Users can also query through a matchmaker for direct return of categories and URLs. Users can include useful external categories in their own collections. They can also keep track of personal agents of most interests for future access. A DIAMS interface example with a query pane and direct query results through a matchmaker is shown in figure 3.
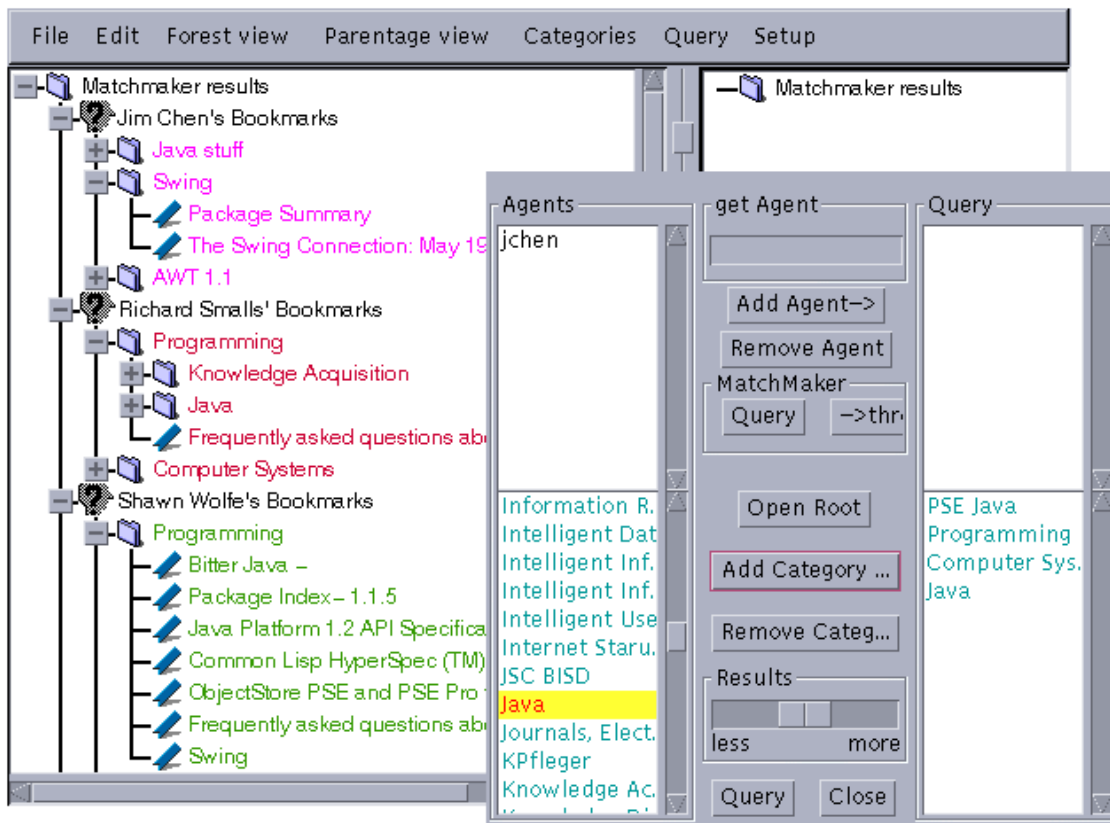
Figure 3: The DIAMS Query Interface with Matchmaker

4.3 Other Current and Future Agents

DIAMS provides several other kinds of agents to facilitate communication and collaboration. Among them, a group agent is the most generally useful. A group agent is very similar to a personal agent, but maintains a common information repository shared and managed jointly by a group of users. A group repository often requires more customization and needs to handle concurrent multi-user access and updates. DIAMS also supports various utility agents to import and export WWW browser bookmarks, to access dictionaries and thesauri, and to translate between DIAMS categories and URL files.

The knowledge agent is an important part of DIAMS currently under development. The use of knowledge base for conceptual indexing and information organization has well been explored [14,15,18]. DIAMS takes a collaborative, distributed knowledge indexing approach built upon the same agent architecture employed by its indexing system. A personal agent maintains its own miniature knowledge base to help organize and manage its collection. A larger knowledge base with more complex semantic relations can be stored in a knowledge agent in DIAMS. Knowledge agents can carry expertise in various special domains. They are initially customized by domain experts, but can also learn new information from their visitors. Personal agents can import knowledge structures and/or commonly used categories from these knowledge agents. Knowledge agents are also used to expand queries to include more elaborate descriptions for better communication between agents.

4.4 Collaborative Learning

DIAMS is designed to encourage collaboration among users. A personal agent helps its owner customize the personal information repository for his/her own needs to ensure the most effective access and management. Personal agents also facilitate collaboration between users through easy browsing, sharing and learning of structured information objects, which carry essential knowledge from their creators and editors. Users can learn from other users with similar interests or expertise. Users are also supported with active information pushing from personal agents to promote collaborative learning of new information and exchange of knowledge. DIAMS agents provide needed functionalities for their users to share and learn information from one another.

5. SUMMARY

We have presented DIAMS, a system of distributed agents that provides services for users to access, manage, share and learn information collaboratively on the World Wide Web. The system is designed to help web users find most needed information from local and/or remote repositories. It incorporates a multi-agent architecture to facilitate information sharing and collaboration. A DIAMS personal agent provides tools and utilities for users to manage their information repository with dynamic organization and virtual views. Object-based structure is used in information repositories to promote easy information sharing and exchange. Flexible hierarchical display is integrated with indexed query search to help ensure effective information access. Automatic

indexing methods are employed to support translation between user queries and communication between agents. Collaboration between users is both aided by easy sharing of information between users, and facilitated by automated information exchange. Connections between users with similar interests and expertise can be established with the help of matchmaker agents. The system also incorporates other utility agents providing needed services. DIAMS is designed to encourage collaboration among users. DIAMS agents provide needed services for users to share and learn information from one another on the World Wide Web.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Balanovic M., An Adaptive Web Page Recommendation Service. *Autonomous Agents* (1997). Marina Del Rey, CA,

[2] Cohen, W. W. A web-based information system that reasons with structured collections of text. In *Proceedings of the 2nd International Conference on Autonomous Agents* (1998), 400-407

[3] Davies, J., Weeks, R and Revett M., Jasper: Communicating Information Agents for WWW. BT Laboratories, Ipswich IP5 3RE UK (1996) (http://www.labs.bt.com/projects/knowledge/jaspaper.htm)

[4] DeRoure, D. C., Hall, W., Reich, S., Pikrakis, A., Hill, G. J., and Stairmand, M., An open framework for collaborative distributed information management. In *Proceedings of WWW7* (1998).

[5] Foner, L.N., Yenta: A multi-Agent, Referral-Based Matchmaking System. *Autonomous Agents*, (1997) Marina Del Rey, CA.

[6] Harman D., Ranking Algorithms, in *Frakes, W.B. and Baeza-Yates R., Information Retrieval, Data Structures and Algorithms,* Prentice Hall (1992)

[7] Kautz, H., Selman, B. and Shah, M., ReferralWeb: Combining Social Networks and Collaborative Filtering. *Communications of the ACM*, 30(3), (1997).

[8] Keller, R.M., Wolfe, S.R., Chen, J.R., Rabinowitz, J.L., and Mathé, N., A Bookmarking Service for Organizing and Sharing URLs. *WWW Conference*, Santa Clara, CA, (1997).

[9] Koller, D. and Sahami, M., Hierarchically classifying documents using very few words. In *proceedings of the 14th international conference of Machine Learning* (1997).

[10] Lockbaum, K.E. and Streeter, L.A., Comparing and Combining the Effectiveness of Latent Semantic Indexing and the Ordinary Vector Space Model for Information Retrieval. *Information Processing and Management,* 25(6), (1989) 665-676.

[11] Mathé, N. and Chen, J.R., User-Centered Indexing for Adaptive Information Access. *International Journal of User Modeling and User Adapted Interaction, Special Issue on Adaptive Hypertext and Hypermedia*, 6(2-3), (1998) 225-261.

[12] Moukas, A., Amalthea: Information Discovery and Filtering using a Multiagent Evolving Ecosystem. *International Journal of Applied Artificial Intelligence*, (1997).

[13] Paepcke, A., Digital Libraries: Searching is not enough, In D-Lib Magazine, (May, 1996), (http://www.dlib.org/dlib/may96/05contents.html)

[14] Pratt, W., Hearst, M, and Fagan, L., A Knowledge-Based Approach to Organizing Retrieved Documents. In *Proceedings of the 16th National Conference on Artificial Intelligence.* (1999).

[15] Sahami, M., Yusufali, S., and Baldonado, M. Q. W., SONIA. a Service for Organizing Networked Information Autonomously. In *Proceedings of the 3rd ACM Conference on Digital Libraries,* (1998) pp.200.

[16] Salton, G., Automatic Text Processing, Addison-Wesley, Reading, MA, 1988

[17] Wolfe, S.R., Wragg S.D. and Chen, J.R., Managing Personal and Group Collections of Information, In *Proceedings of the 4th ACM Conference on Digital Libraries*, (1999) pp. 256.

[18] Woods, W.A., Conceptual Indexing: A Better Way to Organize Knowledge. Tech Report SMLI TR-97-61, (1997) Sun Microsystems Lab. http://www.sunlabs.com/techrep/1997/abstract-61.html