Managing Personal and Group Collections of Information

Shawn R. Wolfe*, Stephen D. Wragg*, and James R. Chen
NASA Ames Research Center
Moffett Field, CA 94035-1000
{shawn, wragg, jchen}@ptolemy.arc.nasa.gov

The internet revolution has dramatically increased the amount of information available to users. Various tools such as search engines have been developed to help users find the information they need from this vast repository. Users often also need tools to help manipulate the growing amount of useful information they have discovered. Current tools available for this purpose are typically local components of web browsers designed to manage URL bookmarks. They provide limited functionalities to handle high information complexities.

To tackle this problem, we have created DIAMS, an agent-based tool to help users or groups manage their information collections and share their collections with others. The main features of DIAMS are described here.

Flexible Organization and Indexed Access

Hierarchies are the most common structure for organizing large collections. A given hierarchy, however, cannot always provide the most suitable display for a particular need. The complexity associated with large hierarchies can also hinder quick access and manipulation. DIAMS addresses the problem of information complexity by integrating flexible organization with indexed access. It supports dynamic hierarchies embedded in a directed graph of categories. Users can customize flexible organization of categories and manipulate virtual views. DIAMS leverages automatic indexing with the user's own categorization to characterize the elements in the collection. It employs the traditional information retrieval measure of term frequency and inverse document frequency, to weigh common keywords extracted from categories and documents. Users can search or query a collection with categories or keywords. The system returns organized information substructures related to the query.

Information Sharing and Agents

Information sharing between users are built upon a distributed agent architecture. Personal or group agents govern the communication, information exchange and privacy protection among users.

Users can query other collections through their agents. While a user's personal agent keeps its own set of categories, it can communicate with other agents through connections between categories and extracted keywords. A querying agent also provides its own knowledge related to the query, to facilitate bi-directional information exchange.

Collection owners can set read/write policies on any portion of their collections. When provided the proper access, a user can create pointers to portions of another collection, or even modify the contents of that collection. This introduces another powerful way of gaining knowledge (by mining the collection of an expert in the field) and allows for the development of a distributed collection.

DIAMS also employs matchmaker agents to assist users in finding other agents that are knowledgeable in a given domain. In response to a given query, matchmakers return a list of relevant information sources. In exchange, the matchmaker requests a characterization of the querying agent, and in this way learns which agents may contain what information.

Conclusion

The vast amount of information available today has lead to the overwhelming chore of keeping track of it all. DIAMS is a prototype system developed to assist users in the access, organization, sharing and continuous discovery of their most valuable information assets.