Managing Personal and Group Collections of Information

Shawn R. Wolfe
Caelum Research

Stephen D. Wragg
Caelum Research

James R. Chen
NASA

NASA Ames Research Center, Mail Stop 269-2
Moffett Field, CA 94035-1000
Tel: 01-650-604-4760
{shawn,stephen,jchen}@ptolemy.arc.nasa.gov

ABSTRACT

In this paper, we describe DIAMS, an agent-based information management tool.

Keywords

Intelligent agents, information access, collaborative system.

1. INTRODUCTION

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, but these alone cannot satisfy all of their needs [7].

Users often need tools to help them 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 address 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.

2. 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 [3]. It supports dynamic hierarchies embedded in a directed graph of categories (see Figure 1). Users can customize flexible organization of categories and manipulate virtual views.

Figure 1: DIAMS User Interface

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 [10] 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.

3. INFORMATION SHARING AND AGENTS

Information sharing between users is built upon a distributed agent architecture [1]. 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 (as described in section 2).

A querying agent also provides its own knowledge related to the query (in the form of related links) 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.

4. IMPLEMENTATION

DIAMS is written as a Java applet and is designed to work within the context of an internet browser. The GUI is written using Java Foundation Classes (Swing), and Java Development Kit. CORBA [7] is used for the agent communication, and all the agents are also written in Java. A small amount of Javascript is used to facilitate communication between the DIAMS GUI applet and the browser.

5. CURRENT WORK

We are adding to the functionality of DIAMS in several ways.

5.1 Structured Keyword Extraction

For the purpose of keyword extraction, DIAMS views each collection as a corpus and each URL as a document in the corpus. The results have been acceptable, but not ideal. The hierarchical structure of collections is not considered in our current algorithm, and research indicates that this is a valuable source of information [3] [5]. By treating the categories as part of a hierarchy rather than a flat collection, we believe that we can improve the utility of the automatically generated keywords, and as a result, improve the quality of search results.

5.2 Knowledge-Based Indexing

At present, categorization is the only direct relationship between elements in DIAMS. Categorization may be sufficient for very small collections, but clearly a wider array of relationships are necessary to represent the complexities inherent in a larger collection. To better convey the structure of a collection and to enable more powerful searching capabilities, we plan to incorporate new relationships into DIAMS and to provide for rudimentary reasoning capabilities.

5.3 Integration with Other Services

Currently, DIAMS treats all external resources as URLs. In many cases, though, greater gains may be achieved by interfacing directly with a structured online collection. We have developed an interface to Postdoc [9], a NASA general-purpose online repository, but our implementation is specific to Postdoc.

Interoperability has been identified as an issue for digital libraries. One solution to this problem used CORBA objects to translate from one protocol to another, thus limiting the number of translation stages needed [8]. We will use a similar approach, and in addition, we plan on adding the ability to interpret and serve Extensible Markup Language (XML) documents [2].

6. CONCLUSION

The vast amount of information available today has led 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.

7. ACKNOWLEDGMENTS

We wish to thank Dr. Nathalie Mathé, a prior prinicipal intestigator, for her contributions to DIAMS.

8. REFERENCES

[1] Chen, J. R., Mathé, N., and Wolfe, S. R. Collaborative information agents on the world wide web. In Proceedings of DL '98 (Pittsburgh, PA, June 1998).

[2] Extensible Markup Language, http://www.w3.org.

[3] Koller, D. and Sahami, M. Hierarchically classifying documents using very few words. In Proceedings of Laguna, Sardinia, June 1997).

[4] Mathé, N., Chen, J. R. and Wolfe, S. R. Organizing and sharing information on the world-wide web using a multiagent system. In Proceedings of ED-MEDIA '98 (Freiburg, Germany, June 1998).

[5] Mladenic, D, an Grobelnik, M. Feature selection for classification based on text hierarchy. In Proceedings of the workshop "Learning from Text and the Web", CONALD-98 (Pittsburgh, PA, June 1998).

[6] Object Management Group, CORBA/IIOP 2.2 Specification, http://www.omg.org/corba/corbaiiop.html.

[7] Paepcke, A. Digital libraries: searching is not enough. In D-Lib Magazine, May 1996, http://www.dlib.org/dlib/may96/05contents.html.

[8] Paepcke, A., Cousins, S. B., Garcia-Molina, H., Hassan, S. W., Ketchpel, S. P., Röscheisen, M., and Winograd, T. Towards interoperability in digital libraries: overview and selected highlights of the Stanford digital library project. In IEEE Computer, May 1996.

[9] Postdoc, http://ace.arc.nasa.gov.

[10] Salton, G. Automatic Text Processing, Addison-Wesley, Reading, MA, 1998, 279-281.