

On Adopting Software Agents for Distributed Digital Libraries

Ch Aswani Kumar & S Srinivas

Abstract

Digital Libraries (DLs) are an access to the collections of electronic documents over network. They serve the scientific community with large amount of electronic data and thereby increase the levels of education and literacy in the society. Software agents have been proved to be the best for the handling of distributed information. This paper surveys about the digital libraries and software agents and analyses the adoption of agents for various issues of distributed DLs.

1. INTRODUCTION

In the present digital world, more and more electronic information is made available to the users by interconnected environments like internet and corporate intranets. Also the number of the users accessing this electronic information is rapidly increasing. With the advent of high speed communication networks and wide area distributed systems users are able to access the electronic information in multimedia two dimensional and three dimensional graphics, audio, and video formats. In particular, the world wide web provides an exponentially growing amount and range of information through which people can browse. Particularly this electronic information serves the research community through online research journals and in other forms. It enables the universities to start various academic programs through e-learning. The number of students enrolling for e-learning at higher education level is increasing rapidly as shown in figure1. But this electronic information provided in the internet suffers from serious drawbacks. There is too much information i.e. information

overload. Finding the appropriate, required information is tedious. The web has no standards for metadata. Also there is a lack of revenue or billing applications.

In many areas like libraries, cultural heritage, health care, entertainment, protection of the environment, etc. there is an increasing awareness towards the building of very large heterogeneous digital information repositories which can be called as digital libraries (DLs), which are interconnected and accessible through global information infrastructure. Software agents play an important role in distributed information management. The role of software agents in various tasks of digital libraries is a current research area. Researchers working in this area have identified agent based digital libraries as the key technology for solving most of the important problems associated with the internet.

2. DIGITAL LIBRARIES

Information scenario keeps changing with the development of national and global information infrastructure. The concept of DL¹

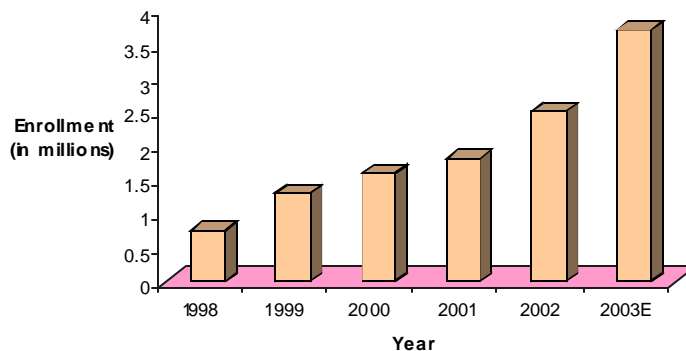


Figure: Growth of E-Learning

arose from the analogy with a place-based repository library containing an organised collection of print-on-paper and other physical artifacts combined with systems and services to facilitate physical, intellectual, and long-term access. DL is a distributed technology environment which reduces the creation, dissemination, manipulation, storage, integration, and reuse of information by individuals and groups.² In USA, in 1994 the National Science Foundation (NSF), the Department of Defense Advanced Research Projects Agency (DARPA) and the National Aeronautics and Space Administration (NASA) launched a Digital Library Initiative (DLI) jointly. Some of the currently existing digital libraries are.

- ◆ USA: Stanford Digital Library Technologies
<http://www-diglib.stanford.edu/WP/WWW/>
- ◆ China: Peking University Digital Library
<http://162.105.138.23/tapian/tp.htm>
- ◆ India: Archives of Indian Labour.
<http://www.indialabourarchives.org/>
- ◆ New Zealand: New Zealand Digital Library
<http://nzdl.org>
- ◆ International: UNESCO Project
http://www.unesco.org/webworld/build_in_fo/gct/bestpractices/anthologies.shtml

The components of the DLs³ include: digital objects, metadata, repositories and harvesting, rights management, indexing, resource discovery, searching and retrieving, linking, interfaces and interaction, architectures, and interconnections. DLs are developed in highly distributed environments.

Library objects are served to remote client processes, which will send information requests and receive results by employing varied communication protocols. Client processes will make diverse interfaces available to users. Repositories on the server side rely on advanced database management systems (DBMSs) for object storage, indexing, and selective retrieval.

2.1 Services of Digital Libraries

There are so many terms related to the DL which include multimedia database, information mining, information warehouse, information retrieval, on-line information repositories, electronic library, and wide area information services. A DL service⁴ is an assemblage of digital computing, storage, and communications machinery together with the software needed to reproduce, emulate, and extend the services provided by conventional libraries based on paper and other material means of collecting, storing, cataloguing, finding, and disseminating information. Public, private, professional, school, commercial, and other kinds of libraries emphasise different services, different types of information, and different service styles.

A DL must accomplish all essential services of traditional libraries and also exploit digital storage, searching, and communication. In addition to delivering documents, DL services will involve all stages of the creative process, from reformatting files, to summarizing the daily news, editing collaborative works, and even helping to plan

new ventures. Services will be customized to fit specific tasks, and personalized to fit individual organisations and users.

The initial emphasis was on the retrospective conversion of print-on-paper objects to digital objects. The vision of digital libraries is, 'Integration and use of computing, communications, and digital content on a global scale, combined with the increasing possibility of cost effective digitization and convergence of formerly separate media types to create the conditions for new infrastructure/environments to support humans as individuals and organisations in distributed knowledge-based activities'. The DLs support individuals or organisations in a broad range of distributed knowledge based activities from electronic commerce to scientific collaboration. They support teaching and learning, especially in the context of distance or lifelong learning. Now the digital libraries are there at universities, publishers, government agencies and public libraries.

2.2 Digital Library Vs. Traditional Library

DLs can handle multimedia data, which can present information more effectively than print media and can be accessed easily and understood even by those who are illiterate. The following points illustrate the potential differences^{2,4} between traditional and digital libraries:

- ❑ Traditional libraries are based upon centralised control and relatively few access locations; digital libraries can be distributed and ubiquitous
- ❑ Traditional libraries support one-way, loosely coupled (slow) interaction; digital libraries support two-way communication with tight, fast interaction
- ❑ Traditional libraries are based upon a model of one-way search: a consumer looking for an object; digital libraries support symmetric search: consumer looking for an object and the producer of the object looking for a consumer.
- ❑ In traditional libraries structured text queries (and some browsing) are used to aid intellectual access; in digital libraries

complex interactions of query, navigation/browsing, and social filtering can be used.

- ❑ Only a librarian may add to the collection of a conventional library, because of the discipline essential to create a quality catalogue. In a digital library, cataloguing discipline and search restrictions to authorized data can be automatically enforced.

3. SOFTWARE AGENTS AND THEIR APPLICATIONS

The concept of agent became important in artificial intelligence (AI) and its sub field distributed artificial intelligence (DAI). AI is the imitation of human intelligence by mechanical means. This paradigm gained momentum in early 90's. There is no single definition for agent. Researchers in agent development gave a number of definitions. Agent can be described⁵ as 'a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.' Ideally, an agent that functions continuously in an environment over a long period of time would be able to learn from its experience. Table 1, depicts the important characteristics of agents. Agents may be usefully classified according to the subset of the properties they enjoy.

Agent systems are usually employed in applications^{6,7} where complex, purpose-directed behaviour is needed in changing environment. There are three important characteristics, which are often cited for adopting agent technology. They are: (a) Distribution of data, control, expertise or resources, (b) Natural metaphor, and (c) Legacy systems.

The number of agent based applications being developed and deployed in real world settings is rapidly increasing. When the domain involves a number of distinct problem solving entities or data sources which are physically or logically distributed in terms of their data, control, expertise or resources and which need to interact with one another or be combined in order to solve their problems or

one common problem, then agents can often provide an effective solution. The general application areas⁸ of agents include personal use, network management, data access, mobility management, e-commerce, computer user interface, application development, and military applications.

4. SOFTWARE AGENTS IN DIGITAL LIBRARY SERVICES

A DL system offers the users a coherent view of an organised, selected and managed body of information. Traditional software development techniques and models cannot support the complexity of these kind of tasks. Software agents provide useful abstraction to system designers to understand the functionality of complex distributed systems like DLs.^{7,9}

A DL can be realised as a collection of agents. The characteristics of the agents, allow in a flexible way to develop a system providing higher search functions for DLs. Such functionality enables the system to help the user in their activities. Agents¹⁰ can translate, communicate and publish information as well as, guide the search process by taking the users' query. Agents can also negotiate access to and the exchange of information with other agents. Personal agents can be used in the digital libraries for the user interface.

Through the personal agents, users can customize their interfaces. The user could be hidden from the DL¹¹ and *vice versa*. Autonomous agents are needed for the user

interface. Depending on the interest of the user, they will search the DLs and return the information which the user is looking for. Some autonomous agents, called 'Knowbots', instead of relying on the rules, or other specifications, develop their own specifications after watching the user interests. Some other autonomous agents find out the dead links in the digital libraries and try to fix them or inform the administrator. Search agents as the name interprets, provide the functionality of basic search engine.

The working mechanism of a search agent is simple.¹² When a user issues a specific query, then a search agent can use a thesaurus to look for related results. Also the search agents may broaden the search if sufficient matches are not found. The key functioning of the Library Service Agents¹³ is to manage the access to the contents of the library, billing, and providing security to the contents. They provide the DL with an interface to the owner's content. The meta agent is a central communication point and acts as an information broker to facilitate communication between other agents. It optimises this communication by routing the messages to the agents to which it is immediately relevant.

In distributed problem solving or DAI, a problem can be split into sub problems and is concerned how a problem can be divided among a number of modules or agents to achieve a solution. Intelligent agents¹⁴ can affect the solutions for sub problems using

Table 1: Characteristics of Agents

Characteristics	Meaning
Social Ability	Able to interact with other agents and humans in order to achieve their tasks
Reactivity	Perceiving the environment and respond in a timely fashion to changes that occur in it
Mobility	Ability to move around the electronic network
Veracity	Assumption that agent, knowingly will not give the false information
Pro-activity	Exhibiting the opportunistic, goal directed behaviour
Autonomy	Exercises control over its own actions

their unique expertise. Further these sub problems can be merged into a complete solution. In DLs, intelligent agents can play a crucial role in information retrieval. When the user poses query, this query will be split into sub queries. An intelligent agent that knows about the sub query will take on the task of fulfilling that sub query. Again the results of all sub queries are recombined to obtain the solution to the entire query. The DL function¹⁵ is expensive in terms of resource consumption when compared to the search engines on the internet. A DL may need to store substantial amounts of user profiling information; in addition, it may need to manage large numbers of regularly scheduled alerting query services. Efficient resource management within a DL is thus particularly important.

Indeed, as the internet and WWW become increasingly commercialised, the need for effective profit oriented agents which can act on behalf of their owners and seek payment for services rendered is likely to increase dramatically. However, for such sophisticated agents to materialise, key work remains to be done in defining and deploying techniques for brokering of information products and also for linking the requests of demand consumer agents to the ability of supply producer agents to provide the required information items.

CONCLUSIONS

Software agents have proved to be the best for handling distributed information issues like information retrieval, integrity and navigation assistance. DLs are large information repositories built on distributed environment. In this review paper, an attempt has been made to analyse the adoption of software agents for various DL services.

REFERENCES

1. Borgman, C. Social aspects of digital libraries. NSF Workshop Report, UCLA, Los Angeles, 16-17 Feb. 1996.
2. Borgman, C.L. What are digital libraries? Competing visions. *Information Processing and Management*, 1999, **35**, 227-43.

3. Sornil, O.A. distributed inverted index for a large-scale, dynamic digital library. Virginia Tech Computer Science, Blacksburg, Ph. D. Dissertation Draft, 2000.
4. Kilker, J. & Gay, G. The social construction of a digital library: A case study examining implications for evaluation. *Information Technology and Libraries*, 1998, **17**, 60-70.
5. Genesereth, Katchpel. Software agents. *Communications of ACM*, 1994, **37**(7), 48-53.
6. Maes, P. Agents that reduce work and information overload. *Communications of ACM*, 1994, **37**(7), 31-40.
7. Jennings, N.R. An agent based approach for building complex software systems. *Communications of ACM*, 2001, **44**(4).
8. Graesser, Art. Is it an agent, or just a program: A taxonomy for autonomous agents. Proceedings of the 3rd International Workshop on Agent Theories, Architectures, and Languages, Springer-Verlag, 1996.
9. Isaias, P. Agents in digital libraries and electronic publication: An architecture proposal. Proceedings of WebNet 99-World Conference on the WWW and Internet, held during 24-30 October 1999. 1542-43.
10. Jennings, N.R. & Wooldridge, M. Applying agent technology. *Journal of Applied Artificial Intelligence*. Special Issue on Intelligent Agents and Multi-Agent Systems, 1995.
11. A. Paepcke, C.-C. K. Chang, H. Garcia-Molina, & Winograd, T. Interoperability for digital libraries worldwide. *Communications of the ACM*, 1998, **41**, 33-43.
12. Jennings, N.R. A roadmap of agent research and development. *Autonomous Agents and Multi-agent Systems*, 1998, **1**, 275-306.
13. Sánchez, J.A. & Leggett, J.L. Agent services for users of digital libraries.

- Journal of Network and Computer Applications*, 1997, **20**, 45-58.
- 14 Fox. E. How to make intelligent digital libraries. Methodologies for Intelligent Systems: Proceedings of the 8th International Symposium (ISMIS'94), edited by N.C. Charlotte, Springer-Verlag, New York, October 1994, 27-38.
15. Sánchez, J.A.; Leggett, J.J. & Schnase, J.L. AGS: Introducing agents as services provided by digital libraries. *In* Proceedings of the Second ACM International Conference on Digital Libraries (DL'97). ACM, Philadelphia, 1997, 75-82.

Contributors: **Ch Aswani Kumar**, Lecturer, Dept of Information Technology, Vellore Institute of Technology, Vellore, Tamil Nadu - 632 014. e-mail: aswani@vit.ac.in
Sh S Srinivas, Asst Professor, Dept of Mathematics, Vellore Institute of Technology, Deemed University, Vellore, Tamil Nadu - 632 014.