1-1-2000

An Analysis of Using Expert Systems and Intelligent Agents for the Virtual Library Project at the Naval Surface Warfare Center-Carderock Division

J. Liebowitz
University of Maryland - Baltimore County

Monica Adya
Marquette University, monica.adya@marquette.edu

Published version. "An Analysis of Using Expert Systems and Intelligent Agents for the Virtual Library Project at the Naval Surface Warfare Center-Carderock Division" in *World Libraries on the Information Superhighway: Preparing for the Challenges of the New Millennium*. Eds. John Carlo Bertot and Patricia Diamond Fletcher. IGI Global, 2000: 169-188. DOI.
Monica Adya was affiliated with the University of Maryland - Baltimore County at the time of publication.

World Libraries on the Information Superhighway: Preparing for the Challenges of the New Millennium

Patricia Diamond Fletcher
University of Maryland Baltimore County

John Carlo Bertot
University at Albany, State University of New York

IDEA GROUP PUBLISHING
Hershey USA • London UK

Senior Editor: Mehdi Khosrowpour
Managing Editor: Jan Travers
Copy Editor: Maria M. Boyer
Typesetter: Tamara Gillis
Cover Design: Connie Peltz
Printed at: BookCrafters

Published in the United States of America by
Idea Group Publishing
1331 E. Chocolate Avenue
Hershey PA 17033-1117
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: jtravers@idea-group.com
Website: http://www.idea-group.com

and in the United Kingdom by
Idea Group Publishing
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 171-240 0856
Fax: 171-379 0609
http://www.eurospan.co.uk

Library of Congress Cataloging-in-Publication Data

Fletcher, Patricia Diamond, 1948-
World libraries on the information superhighway : preparing for the challenges of the new millennium / Patricia Diamond Fletcher, John Carlo Bertot.
p. cm.
Includes bibliographical references (p.) and index.
ISBN 1-878289-66-7 (paper)
1. Internet (Computer network) 2. Library information networks. 3. Libraries--Special collections--Computer network resources. I. Bertot, John Carlo. II. Title.

Z674.75.I58 F54 2000
021.6'5--dc21 99-048437

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter XI

An Analysis of Using Expert Systems and Intelligent Agents for the Virtual Library Project at the Naval Surface Warfare Center — Carderock Division

Jay Liebowitz and Monica Adya
University of Maryland Baltimore County

The Virtual Library Project[1] at the Naval Surface Warfare Center/Carderock Division (NSWC/CD) is being developed to facilitate the incorporation and use of library documents via the Internet. These documents typically relate to the design and manufacture of ships for the U.S. Navy Fleet. As such, the libraries will store documents that contain not only text but also images, graphs and design configurations. Because of the dynamic nature of digital documents, particularly those related to design, rapid and effective cataloging of these documents becomes challenging. We conducted a research study to analyze the use of expert systems and intelligent agents to support the function of cataloging digital documents. This chapter provides an overview of past research in the use of expert systems and intelligent agents for cataloging digital documents and discusses our recommendations based on NSWC/CD's requirements.

The explosion in the use of the Internet for information exchange and retrieval has significantly increased the production and distribution of digital information. This has increased our reliance on rapid

information availability and exchange for decision making. The benefits of such technology motivated trends are several. Information is available globally within seconds. Multiple users can view the same information in the same format all over the world. More importantly, this data exchange is not limited to text but extends to audio, video and pictorial data exchange.

As with any situation, the above benefits are tempered by the challenges that accompany such technological advances. The volume of digital data available is increasing rapidly. Consequently, the Internet has become a library without a catalog. Although several search and retrieval functions are increasingly becoming available, their effectiveness is dependent largely on the effectiveness with which digital documents are cataloged.

Digital documents are being generated and propagated at tremendous rates. In fact, in our daily experience, we find that our exchange of paper-based documents has been reduced to the minimal. Most of the documents are exchanged via e-mail or through the Web and are stored electronically. The challenge then is ensuring that each document being generated and exchanged digitally be effectively and rapidly cataloged and archived. How can the cataloging function be automated such that a document is cataloged as soon as it 's received electronically? In this study we explore the use of Artificial Intelligence (AI) techniques for cataloging. In particular, we examine the use of Expert Systems at the Naval Surface Warfare Center/Carderock Division (NSWC/CD) which is actively exploring the storage and exchange of their books, drawings, manuscripts, documents and photographs digitally so that they can be transmitted at high speed to computer screens anywhere in their original format

THE CATALOGING FUNCTION

According to Anderson (1990), the Head of Cataloging Services at Virginia Commonwealth University, cataloging is, arguably, among the most complex of library processes. Typically, as with NSWC/CD, a cataloger provides both the descriptive and the subject cataloging for each item handled and must also be able to use appropriate classification "schedules" and subject heading thesauri. Familiarity with the principles of uniform headings and the application of name authority control procedures are also standard aspects of the cataloger's responsibilities (Anderson, 1990).

Because the cataloging function is very labor-intensive and costly, many organizations are looking at alternative ways to either aid or automate this function. Some libraries, like at NSWC/CD, are looking at providing a system that helps the contributors catalog their own documents. According to Ruschoof (1995), Head of the Cataloging Department at Georgetown University, the various approaches to cataloging fall into four categories:

- streamlining cataloging through curtailing procedures and simplifying records,
- sharing cataloging through cooperative enterprise,
- obtaining cataloging through contracting schemes, and
- automating cataloging through artificial intelligence or "expert systems."

Chander, et al. (1997) built an expert system to aid cataloging and searching electronic documents on digital libraries. While understanding the domain, they found that to catalog and describe the document, reference librarians relied on their knowledge of classification schemes, terms, indices, structures and resources available in the domain of the user's needs.

Cataloging digital documents is different from cataloging traditional documents in several ways. According to Gaynor (1994), Head of Original Cataloging and in charge of the Cataloging Electronic Texts Project at the University of Virginia Library, the issue with digital documents is that of what constitutes an edition or version or state of an electronic text. Digital documents are more dynamic than traditional documents particularly when they are in print. Once catalogued, traditional documents become a static entity in the catalogs. This does not hold true of digital documents which can change frequently. Another pressing concern for the cataloging community in particular is the relationship, both short and long-term, of the TEI (Text Encoding Initiative) header to the MARC record. Gaynor (1994) states that in the short-term, the cataloging community needs to reevaluate its tools for cataloging electronic texts.

In spite of these concerns, the need for document cataloging is unquestioned. The trend though is to use non-traditional techniques for such cataloging particularly those that rely on Artificial Intelligence (AI). The use of such automated techniques can support the

dynamic nature of the task more effectively and cost-efficiently than traditional cataloging. An example of the use of AI techniques for cataloging is from the University of Exeter. The University is exploring the use of expert systems for cataloging digital resources.

Efforts at OCLC have been directed at automated descriptive cataloging. According to Weibel (1987), the head of the Automated Title Page Cataloging project at OCLC, automated descriptive cataloging entails three broad functional activities: (1) capturing a page image in machine-readable form; (2) identifying functionally related text strings from the page image; and (3) assigning these text strings to appropriate bibliographic fields, e.g., title, statement of responsibility, publisher.

Assigning subject fields entails the following: selecting from a document certain words, phrases or sentences that may suitably represent its information content; if necessary, transforming these sets of words into a standard terminology and casting them into standard form; if necessary, translating the text words or standard terms into a code; and choosing some of these words, terms or codes as access points for the information search. The inherent problem in automated indexing is that the meaning of a word depends on the context in which it is used (University of Texas Report, 1998).

Automatic summarizing is another technique that may help in cataloging. Automatic summarizing is the process by which a computer creates a condensed version of a text. The new version should well represent the original in meaning and scope.

In order to see what has been done in terms of developing expert systems for the cataloging function (Michaelson 1991, Palmer 1990), the next section highlights projects that have been accomplished in this area.

EXPERT SYSTEMS FOR CATALOGING

In reviewing the literature, a number of expert systems have been developed for cataloging documents. However, most of these expert systems are in the prototype or conceptual stages and have not been developed and implemented as *fully operational systems*. This section will highlight some of the leading projects in this area.

Chander et al. (1997), at Concordia University in Montreal, developed an expert system to aid in cataloging and searching electronic documents on digital libraries. For cataloging and searching, they

used a metadata description called a "semantic header" to describe the document. The expert system is designed to mimic the behavior of a reference librarian. This system was built using CLIPS, a rule-based expert system shell.

SKICAT, developed by Usama Fayyad at the Jet Propulsion Laboratory and CalTech Astronomy, includes tools for the analysis and exploration of a large catalog database—namely, a comprehensive northern sky survey catalog that will ultimately contain on the order of one billion entries.

At Northwestern University, several expert system projects were developed for cataloging. CLARR is an expert system to assist in MARC field validation. DELICAT (Data Enhancement of Library Catalogues) is an expert system capable of automatically detecting errors in library catalogues and drawing these to the attention of library staff. IESCA (The Interactive Electronic Serials Cataloging Aide) is a tutorial which assists libraries in cataloging electronic documents. It guides the user through the process and includes most applicable rules and standards from cataloging organizations.

OCLC has developed the Automated Title Page Cataloging System. This prototype is a rule-based system, written in PROLOG, which interprets typographic files and builds an approximation of an AACR2 first level description. In preliminary tests, 75% of the fields in a sample were identified correctly; half of the title pages were captured correctly in their entirety.

Another OCLC-sponsored project by Svenonius focused on the problem of automatically deriving name access points. She found that 93% of all personal names and 80% of all corporate names assigned by the Library of Congress (LC) and the National Library of Medicine (NLM) could be successfully assigned by an expert system.

A newer research project at OCLC is called Scorpion[2]. Scorpion explores the indexing and cataloging of electronic resources. The primary focus of Scorpion is the building of tools for automatic subject recognition based on well-known schemes like the Dewey Decimal System. According to the project leaders on Scorpion, Scorpion cannot replace human cataloging. There are many aspects of human cataloging that are difficult if not impossible to automate, according to the Scorpion research team.

At Exeter University in England, an expert system was developed for cataloging. The system used AACR2 and the MARC manual as

sources of rules for the knowledge base. In a keynote address, Tillett (1995) discussed the following conclusions from the University of Exter's efforts:

- rules for determining access points could be reinstated as production rules,
- rules governing bibliographic description could be entered through the use of templates,
- local rules could be accommodated,
- menus and screen forms could be designed which would act as a user-friendly interface between a cataloger and an expert system, and
- explanation facilities, in terms of the systems' attempt to decide which rules from AACR2 (Anglo-American Cataloging Rules, 2nd ed.) apply, could easily be incorporated.

At UCLA, MAPPER was an expert system designed for the descriptive cataloging of maps. This doctoral project used AACR2 cataloging rules for maps and knowledge contributed by expert catalogers. A similar project called ESSCAPE (Expert Systems for Simple Choice of Access Points for Entries) was developed at Linkoping University in Sweden. ESSCAPE was designed for the cataloging of books versus maps.

The University of Michigan Digital Library (UMDL) project is an agent-based approach for rendering library services in a digital networked environment.

It is worth noting that even though expert systems are being prototyped, built and designed for cataloging, their full potential in the cataloging function is questionable. Even if the development of such a system is possible, the complexity and expense of encoding current processes will be a deterrent. Furthermore, the expert system should be customizable so that it may observe local practices.

NSWC/CD'S REQUIREMENTS FOR AN EXPERT SYSTEM FOR CATALOGING

Based on discussions with the Virtual Library project leaders at NSWC/CD, the following requirements for an expert system for cataloging digital records emerged as follows:

- the Virtual Library project will be operated over the Web with a Web interface;
- documents will ultimately be added to the Library by individual users over the Web (thus, a cataloging expert system should be able to catalog these documents automatically or at least aid the user in the cataloging process);
- Center reports are accessed the most so the expert system should focus initially on cataloging Center reports;
- for digital document check-in, the following fields are **required** to be cataloged, according to NSWC/CD officials:
  - corporate source
  - title
  - personal author
  - publication date
  - series numbers
  - major descriptors (thesaurus terms)
  - identifiers (keywords that are not in the thesaurus)
  - barcode
- the following fields would be **desirable**, according to NSWC/CD officials:
  - pagination
  - subject descriptors (thesaurus terms)
  - circulation restriction
- cataloging should follow acceptable standards and guidelines (e.g., American National Standard's Scientific and Technical Reports—Elements, Organization and Design—adopted by the Department of Defense).

IS A CATALOGING EXPERT SYSTEM FEASIBLE FOR NSWC/CD'S NEEDS?

It appears, on paper, that expert systems technology should be suited for descriptive cataloging (not subject cataloging) due to well-established rules. However, in reviewing the work that has been done over the past 15 years in applying expert systems to cataloging, very few systems, if any, have been implemented beyond the prototyping or conceptual stage. As Lancaster and Smith (1998) point out, most of the studies have provided conceptual models of expert systems rather than actual implementations. They cite several studies that have revealed that tasks that are very easy for human catalogers are difficult

to delegate successfully to automated routines. For instance, the distinction between title and subtitle may not be as evident to the machine as to a human cataloger (Lancaster and Smith, 1998).

None of the experiments on the automation of descriptive cataloging has produced a significant prototype system, much less a truly operational one (Lancaster and Smith, 1998). The costs involved in developing a complete and comprehensive cataloging system are prohibitive. The National Library of Medicine abandoned a $130,000 project for identifying the correct form of an author's name since it was determined that putting an operating tool out would be too expensive an undertaking. Similarly, the QUALCAT project, which identified duplicate records and selected the one that was of superior quality, was declared infeasible due to the limitations of MARC records (Lancaster and Smith, 1998).

Although, Clarke and Cronin (1983) suggested 16 years ago that an appropriate application of expert systems might be to catalog electronic publications as they are generated online, in light of the findings and conclusions discussed above, it does not seem promising to develop an expert system for the cataloging function. Poulter (1994) also confirms our findings.

Based on the above analysis, we reiterate the infeasibility of developing an operational expert system for cataloging. We base our conclusions on the following reasons:

- In an earlier publication, Liebowitz (1998) identified several criteria for expert system problem selection. They are:
  a. the task involves mostly symbolic processing,
  b. test cases are available,
  c. problem task is well-bounded,
  d. written materials exist explaining the task,
  e. task requires only cognitive skills,
  f. experts agree on the solutions to the problem,
  g. at least one expert exists and the expert is cooperative,
  h. the expert is articulate,
  i. the expert's knowledge is based on experience, facts and judgment,
  j. a need exists for developing an expert system for the problem,
  k. the task will be provided with the necessary financial and moral support,

l. top management supports the project,
m. the domain-area personnel have realistic expectations regarding the use of an expert system,
n. users would welcome the expert system, and
o. the knowledge used in the expert system is not politically sensitive or controversial.

It appears that an expert system for cataloging digital documents could meet many of these criteria. However, the key deterrent is that many leading authorities have tried to develop expert systems in the cataloging area over the years, and have only produced mainly prototypes versus real, fully operational systems.

- It might be possible to develop a *limited* version of an expert system for **descriptive** cataloging of digital documents, but it would be difficult to develop a system for automatic **subject** cataloging.
- The expert systems that have been developed so far in cataloging have had very limited success.
- There are more viable AI technologies versus expert systems that could have a greater likelihood of success for handling library functions.

In the following section, we propose a preferred solution as the next step for rendering library services in a digital networked environment as part of the Virtual Library project at NSWC/CD.

INTELLIGENT AGENTS FOR SEARCHING MULTIPLE SITE DATABASES

One of the recent trends in Artificial Intelligence has been the use of intelligent agents for searching and indexing functions, particularly over the Internet for Web-based documents. Intelligent agents are software that assist people and act on their behalf. They can automate repetitive tasks, remember things, intelligently summarize complex data and learn from and make recommendations to the users (Russell and Norvig, 1995). Agents are:

- autonomous, that is they have control over their own actions;
- goal-driven, in that they have a purpose and act in accordance with those actions;

- reactive, in that they sense changes in their environments and respond in a timely fashion to these changes;
- adaptive, that is they change their behavior based on previous experience.

Intelligent agents can be classified in a number of ways based on the functions they perform (Gilbert):

- *Interface Agents:* These are used to decrease the complexity of the increasingly sophisticated and overloaded information systems. They may add speech and language understanding to otherwise dumb terminals.
- *System Agents:* These run as integrated parts of operating systems or network protocol devices and help manage the complex distributed computing environments.
- *Advisory Agents*: They are used in complex help or diagnostic systems.
- *Filtering Agents*: These agents reduce information overload by removing data that does not match the user profile. Agentware and InfoMagnet provide more general kind of filtering capabilities.
- *Retrieval Agents*: These agents search and retrieve information. ATI, Bullseye, Go-Get-It and Surfbot are some of the tools that provide these capabilities.
- *Navigation Agents*: Navigation agents are used to navigate through external and internal networks, remembering short-cuts, preload caching information and automatically bookmarking interesting sites. IBM's Web Browser Intelligence is an example.
- *Monitoring Agents*: These provide users with information when particular events occur, such as information being updated, moved or erased.
- *Recommender Agents*: They are collaborative agents that need many profiles to be available before an accurate recommendation can be made.
- *Profiling Agents*: These agents are used to build dynamic sites with information and recommendations tailored to match each visitor's individual tastes and need.

For digital library initiatives, most of the research and interest lies in the areas of filtering, retrieval, navigation, monitoring, recommender

and profiling agents. In the section below, we discuss briefly some research in the design and use of intelligent agents in these functions.

Research on Intelligent Agents

In the context of digital libraries, several studies have indicated their findings with the use of intelligent agents on online databases and on the Internet. Harvest was developed to index topic-specific collections rather than to locate and index all HTML documents it can find (Information Interchange Report). It allows the users to control the enumeration in several ways by including stop lists, depth limits and count limits. Searchbots are a simple example of a cooperative multi-agent approach to information retrieval. To satisfy one's query, multiple agents perform search at heterogeneous remote sites via the Web. Some of the search methods may include using existing database search engines. Domain experts determine what sites to search and the path to the best solution. The best solution is the one with the lowest search cost (CIG).

Newt is an example of an information filtering system utilizing a society of agents that inhabit the user's computer (Beerud, 1994). Each agent is a user profile that searches for documents that matches the profile and recommends these documents to the user. The user can provide feedback to the agent for the documents recommended. User feedback changes the fitness of the profiles. If the user provides positive or negative feedback for a document, the fitness of the profile that retrieved the document is either increased or decreased. Second, user feedback modifies the profile. Therefore, each agent learns and adapts during its lifetime to the changing needs of the user.

Rapoza (1996) describes the use of filtering agents for feeding corporate intranets. Such agents push news items or other information to network users within selected categories such as sports, financial news and others. Richardson (1996) provides a discussion of an agent that indexes a local database maintained on a Web server. This tool though has limited text capabilities.

Certain commercial products provide several facilities for document management. AppleSearch is an agent system that searches and retrieves text from computers linked together by AppleShare, Apple's file sharing application, or by System 7's personal file sharing capabilities. Up to 50 users can operate on a network as AppleSearch clients. AppleSearch uses agents to examine text and to read and index

documents that exist in a variety of formats (Valauskas, 1994).

The University of Michigan, as an undertaking in NSF's Digital Libraries Initiative, has created a digital library for science in the schools, grades six through nine. The project was a vehicle for investigating the use of ontologies and intelligent agents, as well as for creating economic incentives for use[3]. They now have an operating model that includes age-appropriate materials, as well as tools for working with the information.

A comprehensive review of agent use can be found in Lancaster and Smith (1998) and Liebowitz and Beckman (1998).

Issues with Agent Technologies

Research on the use of intelligent agents for information retrieval and management is in its infancy. As with most new technologies, there are some concerns with using intelligent agents over the Internet:

- they can generate a substantial amount of load on servers;
- sites can block out access by agents (Eichman, 1995);
- agents roaming the server can threaten the server with theft of resources, assets and reputation (Ordille and Travis, 1996).

Most of the initiatives are in preliminary stages although commercial products using agent capabilities are emerging rapidly. Despite its recent emergence, agent-based technologies are gaining popularity in document retrieval and management, as the next section reveals.

Ongoing Research on Intelligent Agents in Digital Libraries

STARTS—Stanford Protocol Proposal for Internet Search and Retrieval—is a project initiated at Stanford University to develop an agent-based protocol that text search engines should follow to facilitate searching and indexing multiple collections of text documents. Research and development of intelligent agent technology, particularly in the areas of digital libraries, document and workflow management, and computer games, is underway at Mitsubishi Electric Digital Library Group in London. The largest current undertaking is the NSF-sponsored Digital Libraries Initiative (DLI). Under the DLI, six universities, including the University of Michigan and Carnegie Mellon University are exploring various new information technologies to support various functions within digital document management. All

six universities involved in the DLI are utilizing intelligent agent technologies in one form or another.

RECOMMENDATIONS FOR THE NSWC/CD REQUIREMENTS

Instead of using expert systems technology, a preferred approach to aiding the Virtual Library Project at NSWC/CD may be the use of intelligent agents for searching multiple site databases and acting as a virtual TIC user agent. Intelligent agents are already being applied in a number of digital library initiatives. For example, the University of Michigan Digital Library project designs and builds a flexible, scalable infrastructure for rendering library services in a digital networked environment. In order to meet this goal, intelligent agents are utilized for representing the library as a collection of interacting agents, each specialized to perform a particular task and all of them acting in an artificial economy.

Other projects like WebSEEk have been developed at Columbia University as a content-based image and video catalog and search tool for the World Wide Web. It has already catalogued over 650,000 images and 10,000 videos from the Web. Netscape Communications Corporation also has an agent-based product titled Catalog Server where catalog agents gather resource information from a variety of primary sources (including Web servers, end users, archivists and legacy systems). The agents communicate indexing information to one or more Catalog Servers via an open standard. For other projects, the reader may want to consult the National Science Foundation's Digital Library Initiative[4] and Project Aristotle[5].

The goal of digital libraries, according to the University of Michigan Digital Library, will be to provide mechanisms by which a digital library can continually reconfigure itself as users, contents and services come and go. Challenges for digital libraries include: digitizing contents, particularly a variety of forms such as documents and images; computerizing services to support this digitization; and networking users.

The NSWC/CD Virtual Library initiative could require several of the following components: document acquisition, document cataloging and management, document retrieval and other services (e.g., library security, bibliographic information, copyright protection). Each of these areas will be discussed next.

Document Acquisition

Since Internet access and input is a projected goal for the NSWC libraries, one of the challenges will be the identification of remote documents in addition to the local documents. Presuming the local acquisitions will be supported by traditional library functions, support for extracting indexes from suitable sites on the Internet and cataloging and managing these resources will be critical. For this task, the research team proposes the use of a multi-agent architecture.

Document acquisition via the Internet can take on two aspects. The first would be input from users of the Internet. These would be documents created by users as well as those found over the Internet via traditional search mechanisms. In the latter case, to help users search over heterogeneous information services that support nonuniform query languages, users will be allowed to compose simple boolean queries in a front-end natural language interface that may be interfaced with an intelligent agent that will then take on the task of searching the Internet.

For automatic search of the Internet, we propose the use of an intelligent agent that would take initiative in searching the Internet for relevant resources. It may be beneficial to promote communication between this agent and the one discussed in the above paragraph to enhance this agent's search capabilities.

Once retrieved, some mechanism for rating and recording the contents of digital documents will be required for facilitating the cataloging functions. This may be best supported by some parsing mechanism.

Document Cataloging and Management

The cataloging and management of documents obtained through various sources discussed above can be supported by traditional keyword indexing, knowledge-based systems and an agent-based architecture. For documents that are input by users, titles and keywords can be used for supporting cataloging and management functions. This will typically require traditional indexing capabilities. A knowledge base for known indexes or analogous terms may also be maintained.

For documents searched via the Internet, some parsing mechanism will be required to determine keywords from frequently occurring phrases in the document. An intelligent agent may be able to

support this function most effectively by filtering out nonsignificant characters and strings. SGML tags from these documents can provide further insights into the nature of the documents.

Document Retrieval

The digitization of libraries will be able to enhance the search and document facilities provided by traditional library functions. These can be supported by simple query functions that will allow, among others:

- author search for authors who have produced a document, have cited certain other authors, have a name in the body of the text, etc.;
- search on keyfields, titles, bibliographic information and their combinations;
- modifying searches based on search results;
- Boolean searches;
- full-text SGML retrieval;
- rating and site information.

Other Services

Library Security

Traditional library functions have been open with security issues, particularly with the intention of promoting usability. In the NSWC/CD libraries, security becomes a concern since the libraries will contain both public and private documents relating to naval warfare. In particular, security architectures are required for mobile computing to allow the frequent migration of computers in and out of security enclaves and wide-area collaboration and to create dynamic sessions that stretch across organizational boundaries (as done in the University of Illinois Digital Library Initiative).

In response to the security issues, we propose the development of a system access and encryption mechanism. This mechanism will undertake the process of verifying the authority and validity of new security services and policies and enforcing the restriction of system resource access by these services and policies. It will provide basic public-key encryption and authentication. Together, these components will restrict access to the system and will determine if a user should be allowed access to a particular operation. Simple Java applets can be used to provide the support for this part of the initiative.

However, to promote flexibility and extensibility of this system, a mobile agent may be used.

Bibliographic Information and Copyright Protection

A tool for maintaining bibliographic information should be able to support various formats and provide a uniform searchable database of bibliographic information. Simple SQL query mechanisms that permit searches on authors, journals, documents/reports and keywords would be able to address the needs for the NSWC/CD digital libraries.

The generation and distribution of illegal copies of digital documents is an issue. Some mechanism will be required to track the unauthorized distribution of illegal versions of documents.

Next Steps for the NSWC/CD Virtual Library Project

The proposed initial research for the Virtual Library project at NSWC/CD will involve developing two intelligent agents. The first agent will search multiple database sites to retrieve relevant documents per the user's keyword requests. The second agent will develop a user's profile dynamically to better tailor the user's search.

Specifically, an intelligent agent is needed to query multiple site databases (for example, the Carderock TIC, the Dahlgren Library and the Indian Head Library) simultaneously from a proxied Web browser interface at either site and deliver the results back to the end-user's browser. It must know a definable query language (SQL) and be able to communicate using the ANSI standard Z39.50 protocol (Web-based protocol for querying databases). The agent must be able to authenticate itself in order to access databases. The agent must also be able to accept and return binary, as well as ASCII text data.

The second intelligent agent is needed to act as a proxy "virtual agent" for TIC online users. Each user will be able to define their own information profile which will be used to set an intelligent agent's criteria to periodically search the Web and deliver search results back to the end-user, by e-mail or some other user-defined way. The agent must also build a user's profile on the fly. It must be able to discern duplicate information and filter out redundant results. The agent must have a definable scope for querying (i.e., the level of links to follow, etc.). The agent must be able to avoid sites/files where bots are not authorized (i.e., recognize robots.txt files). The research steps

proposed for this effort are:

- *Web Development*: Develop the appropriate Web sites that will support user querying, access, and search modifications. In particular, emphasis must be given to the mode of presentation of query results. The Web site must be able to support display in various forms such as html and pdf among others. Appropriate support must also be provided for display of images and sound.
- *Testbed Development*: Identify and develop a testbed for testing the effectiveness of the proposed solution. We propose that the research team in collaboration with appropriate authorities at NSWC identify:

a) the volume of documents stored at various databases
b) the types of documents and their quantities stored at each location
c) the volume and urgency of requests from users for each of the locations, and
d) the rate at which these requests are satisfied from each location.

These measures will allow the research team to sample appropriate documents from each location to form a part of the testbed. This will also facilitate development of an objective measure of success of the proposed solution versus the traditional approach. If successful on the testbed, the proposed solution can eventually be phased in over the entire database.

- *Tool Exploration*: Research and demo available commercial agent tools (e.g. Verity's Search '97/Agent Server and Open Sesame/ Browne and Company's Learn Sesame), as well as public domain agent tools.
- *Tool Integration*: Develop the design for integrating the two proposed agents within the Virtual Library project.
- *Agent Testing*: Encode and test these intelligent agents on the testbed described above.
- *System Refinement-1*: Refine the agents based upon results from the initial tests.
- *User Evaluation*: Once the system has been refined by the research team, we propose the release of the system to a select set of users. These may be obtained from the usual sample of users as well a group from the library team. It is then recommended that the research team obtain feedback from the test user community. We propose an online feedback mechanism that may be offered after

each search or after every few searches performed by the end user. In addition, a more detailed survey can be requested from the user at the end of the test period.

- *SYSTEM REFINEMENT-2*: Refine the Web sites and the agents based on results from user evaluations.

Since several of these activities such as testbed development, tool testing and integration can be conducted in parallel, we anticipate that this would be a six to nine-month effort whose deliverable would include: the software code for the two intelligent agents and a brief user's guide on how to use them.

FUTURE RESEARCH

Several issues will need to be examined in the future of this and other intelligent agent projects in this domain. To accommodate the increasing volume of information retrieval and cataloging over the next few years, developers must examine the need for high speed access to document databases. With the advances in telecommunications technology, such high speed access is increasingly becoming possible. Furthermore, state-of-the-art storage devices will be required to support documents, images and possibly sound.

One of the important benefits of digital libraries is the long-term maintenance of digital documents. Careful thought must go into acquiring the appropriate infrastructure to ensure this. Furthermore, with the increasing use of Internet for accessing digital documents, the incorporation of security modules both on the Web site and the catalog and document databases will require significant energies.

CONCLUSIONS

In the past decade, computer technology has become increasingly powerful and easy to use. With the advent of the World Wide Web, the production and exchange of digital information has increased exponentially. Consequently, the development of digital libraries is gaining increasing significance. The large volume of documents in digital libraries, the myriad forms of these documents and the dynamic nature of digital documents call for automated mechanisms that expedite the function of library cataloging while maintaining the efficacy of human catalogers. Expert systems have been proposed as a possible solution for this cataloging need. However, 15 years of

evidence indicates that expert systems are not suited for the task of digital cataloging, primarily because of the cost constraints. In this chapter we have proposed the use of intelligent agents for library cataloging. Although the technology is still in its infancy, it is potentially viable for digital libraries. The increasing use of intelligent agent technology in ongoing library projects confirms and corroborates our conclusions.

ENDNOTES

[1] This research was made possible by sponsorship from Dave Glenn, Director of the Virtual Library Project, U.S. Naval Surface Warfare Center-Carderock Division, West Bethesda, Maryland.

[2] http://purl.oclc.org/scorpion

[3] Discussion of this is available at http://mydl.soe.umich.edu.

[4] Details of the Digital Library Initiative are available at www.nsf.gov.

[5] Details of the Aristotle project are available at (www.public.iastate.edu/~CYBERSTACKS/Aristotle.htm)

REFERENCES

Anderson, B. (1990). Expert Systems for Cataloging: Will They Accomplish Tomorrow the Cataloging of Today?. *Cataloging & Classification Quarterly,* 11 (2).

Beerud, S. (1994). A Learning Approach to Personalized Information Filtering. *Learning and Common Sense Section T. R. 94-01*, MIT Media Laboratory. [On-line]. Available: http://agents.www.media. mit.edu/groups/agents/papers.html.

CIG. (n.d.). [On-line]. Available: http://dis.cs.umass.edu/research/searchbots.html

Chander, P., Shinghal, R., Desai, B., & Radhakrishnan, T. (1997). An Expert System to Aid Cataloging and Searching Electronic Documents on Digital Libraries. *Expert Systems With Applications Journal, 12*, (4).

Clarke, A. & Cronin, B. (1983). Expert Systems and Library/Information Work. *Journal of Librarianship, 15*.

Eichman, D. (1995). Ethical Web Agents. *Computer Networks and ISDN Systems, 28*.

Gaynor, E. (1994). Cataloging Electronic Texts: The University of Virginia Library Experience. *Library Resources and Technical Services, 38*, (4).

Gilbert, D. (n.d.). *Intelligent Agents: The Right Information at the Right Time*. [On-line]. Available: http://www.networking.ibm.com/iag/iaghome.html.

Information Interchange Report. (n.d.). *Intelligent agents and information*

retrieval. [Online]. Available: http://www.techapps.co.uk/iiartagt.html .

Lancaster, F.W. & Smith, L. C. (1998). "Intelligent Technologies in Library and Information Service Applications. *Report for the Special Libraries Association*, University of Illinois at Urbana-Champaign, Graduate School of Library and Information Science.

Liebowitz, J. (Ed.). (1998). *The Handbook on Applied Expert Systems*, CRC Press, Boca Raton, FL.

Liebowitz, J. & Beckman, T. (1998). *Knowledge Organizations: What Every Manager Should Know*, CRC/St. Lucie Press, Boca Raton, FL.

Michelson, A. (1991). Expert Systems Technology and Its Implications for Archives. *National Archives Technical Information Paper No. 9*, March 1991.

Ordille, J.A. & Travis, L. (1996). When Agents Roam, Who Can You Trust? *Proceedings of the First Annual Conference on Emerging Technologies and Applications in Communications*. Los Alamitos, CA, IEEE Computer Society Press. 188 - 191.

Palmer, M. (1990). *Expert Systems and Related Topics: Selected Bibliography and Guide to Information Sources*, Idea Group Publishing, Hershey, PA.

Poulter, A. (1994). LIS Professionals as Knowledge Engineers. *Annual Review of Information Science and Technology, 29*, 1994.

Rapoza, J. (1996). Service Filters News for Intranets *PC Week 13* (21)

Richardson, J.V. (1996). Add an Engine, *Internet World, 7* (5).

Ruschoff, C. (1995). Cataloging's Prospects: Responding to Austerity with Innovation. *The Journal of Academic Librarianship, January.*

Russell, S. & Norvig, P. (1995) *Artificial Intelligence: A Modern Approach*, Prentice Hall, NJ.

Tillett, B. (1995). Keynote: Cataloging Rules and Conceptual Models. *National Cataloguing Conference*, New South Wales, Australia, October 19, 1995.

University of Texas Report. (1998) "AI/Expert Systems in the Library—Cataloging," University of Texas—Graduate School of Library and Information Science, http://fiat.gslis.utexas.edu/~palmquis/courses/project/

Valauskas, E.J. (1994). AppleSearch: How smart is Apple's intelligent agent? *Online. 18*, (4), 52-64.

Weibel, S. (1987). Automated Title Page Cataloging. *The Annual Review of OCLC Research*, July 1986-June 1987.