The Intelligent Reference Information System: An Expert System to Select Networked CD-ROM Databases and Other Reference Resources

Charles W. Bailey, Jr. and Kathleen Gunning

Preprint: 5/3/1990

Introduction

In 1989, the University Libraries of the University of Houston were awarded a $99,852 U.S. Department of Education Research and Demonstration Grant to develop and study a prototype Intelligent Reference Information System (IRIS) over a two-year period. This grant will pay for 51% of the estimated costs of the IRIS Project. IRIS will synergistically blend the unique strengths of two promising computer technologies that emerged in the 1980s: CD-ROM databases and expert systems. The system will (1) identify and describe appropriate electronic and printed resources to meet users' reference information needs; (2) link users to appropriate networked CD-ROM databases; and (3) provide users with location information about appropriate stand-alone CD-ROM databases and printed reference sources.

CD-ROM Technology

The power of CD-ROM (Compact Disk Read-Only Memory) technology has allowed academic libraries to provide affordable, fixed-cost database services for library users. Running on a microcomputer workstation, a CD-ROM database can provide access to more than 540 MB of data on a 12 centimeter optical disk. A growing number of databases are becoming available in CD-ROM format, and these databases provide users with an increasing diversity of types of information. Currently, databases are available that contain bibliographic citations, full-text (e.g., articles and reference works), numeric data, and graphic information (e.g, maps).

The success and popularity of stand-alone CD-ROM databases has resulted in a number of challenges for libraries. By their nature, stand-alone CD-ROM databases limit use to one person at a time, preventing users from accessing popular databases when needed due to contention for these products. In response to high user demand for CD-ROM databases, libraries are scheduling the use of these databases to reduce the level of user frustration; however, scheduling entails administrative overhead.

Stand-alone CD-ROM databases also require that libraries either dedicate specific microcomputers to particular CD-ROM databases or else employ general-use microcomputers to run CD-ROM disks that are checked-out from a service desk. Either method has disadvantages. If workstations are dedicated, users must queue up for a particular microcomputer to access the needed CD-ROM database, forcing libraries to consider multiple CD-ROM workstations for popular products. If workstations are general purpose, libraries must provide labor-intensive circulation services for CD-ROM disks, users must continually insert and remove disks from CD-ROM players, and all CD-ROM searching software must be stored and maintained on multiple microcomputers.

Another difficulty with stand-alone CD-ROM workstations is that they typically have no built-in method of maintaining use statistics for CD-ROM databases, requiring that libraries create software for this purpose.

A growing number of CD-ROM network systems are commercially available. They offer many potential benefits. By providing multiple users with simultaneous access to desired CD-ROM databases, networked CD-ROM systems have the potential to (1) lessen user contention for databases, (2) reduce the number of copies of CD-ROM databases required to meet user demand, (3) diminish the likelihood of damage to or loss of CD-ROM disks, (4)

decrease the level of effort needed to install and maintain CD-ROM searching software, and (5) improve the monitoring of CD-ROM utilization through use of network management software. However, the performance characteristics of CD-ROM networks are not well understood, and further work needs to be done to assess their usefulness as a means of providing multiuser access to electronic information.

Expert System Technology

Like CD-ROM, expert system technology provides a powerful new tool for improving library users' access to information. Library expert systems embody and provide to the user the special expertise of library staff in using the library's collections, systems, and services. Expert systems have three primary components: a knowledge base that contains facts, rules, and other representations of human knowledge related to a particular topic; an inference engine, which solves problems by manipulating information in the knowledge base in a way that mimics human reasoning; and a user interface, which permits system interaction with the user and explains system findings.

Expert system technology has the potential to provide library users with sophisticated guidance about appropriate information resources to meet their reference information needs. Reference expert systems have several potential benefits, including: (1) advice can be provided at times that reference staff are unavailable, (2) reference staff can be freed to answer higher-level questions, (3) the specialized knowledge of particular reference staff members can be preserved, and (4) users who are uncomfortable with asking for personal assistance can be given another way of meeting their information needs.

In stand-alone mode, a reference sources expert system can identify appropriate printed and electronic reference sources, describe these sources, and provide the user with location information. By linking a reference sources expert system to selected CD-ROM reference databases that provide different types of information (e.g., citation, full-text, and numeric information) and broad coverage of major disciplines, the user can instantly access a representative selection of reference material in electronic form.

Expert systems have significant potential; however, applications need to be carefully selected based on the current strengths and weaknesses of expert system technology.[1] Moreover, a considerable amount of research and development work needs to be done before expert systems will be commonly used in libraries. As Aluri and Riggs state in the conclusion of a survey of library expert system applications:

> The field of expert systems is still in its infancy. Our use and assessment of expert systems in libraries are few and primitive. During the next 10 years we can expect much more work with the refinement of expert systems in libraries. There will be a major effort made in linking expert systems with external programs/systems. . . . With the noticeable decline in the cost of computer technology and the increased literacy in the use of this technology, now is an opportune time for the application of expert systems in libraries. [2]

IRIS Project Planning

The University Libraries' proposal to develop an Intelligent Reference Information System was a logical extension of two innovative projects in the Libraries: the Intelligent Reference Systems Project and the Electronic Publications Center Project. These projects provided a firm foundation upon which to build the proposed Intelligent Reference Information System and they demonstrated the Libraries' commitment to CD-ROM and expert system technologies. The University Libraries' experience with a successful internal IBM Token-Ring Local Area Network, which has been in operation since 1986, also helped lay the foundation for this proposal.

The Intelligent Reference Systems Project

The Intelligent Reference Systems Project was established in January 1988.[3] The Project's initial goal was to produce a working prototype of an expert system to assist library users in identifying appropriate indexes and abstracts to meet their information needs. The Project

was also charged with recommending whether a production version of the system should be developed.

The University Libraries had previously developed a reference system called the Information Machine, which was written in QuickBASIC.[4] The Information Machine provides users with menu-driven access to explanations of library research strategies, materials, systems, facilities (complete with library maps), services, open hours, policies, and other orientation information. Designed to be easily modified for use in other libraries, the system is currently being marketed by the AMIGOS Bibliographic Council. The Information Machine had been very successful; however, the Libraries staff felt that the power of expert system technology was required to provide more complex assistance to users.

The Project team developed a functional specification prototype and three expert system prototypes by August 31, 1988. The functional specification prototype, created with Dan Bricklin's Demo Program II, was an operational mock-up of the system, demonstrating how it should function. The expert system prototypes were working expert systems that were rough drafts of the desired expert system. These expert system prototypes were created using VP-Expert, the Intelligence/Compiler, and Turbo Prolog. The VP-Expert prototype was created to quickly assess whether the desired system could be built using expert system technology. It demonstrated that this was possible. Since it was not the Project's intention to use VP-Expert to develop the final prototype, the Project obtained the Intelligence/Compiler, which could support a frame-based knowledge representation scheme. This expert system shell was very powerful, but it had limitations in light of the Project's design objectives. (A new version of the Intelligence/Compiler, which we have not evaluated, has subsequently been released.) It was decided to develop two prototypes, one using the Intelligence/Compiler and the other using Turbo Prolog, which offered better graphics capabilities. After assessing the capabilities of these two expert system tools, the Project endorsed continued development of the system using Turbo Prolog.

The University Libraries' Director accepted the Project's recommendation for developing a production version of the system and established the Intelligent Reference Systems Committee to accomplish this goal. A pilot version of the system was programmed, and it was tested by library users from May to June 1989.[5] After the successful test, the system's interface was modified. Currently, the system's knowledge base is being significantly expanded to include all major indexes and abstracts in the Libraries' collection.

The Electronic Publications Center

The Electronic Publication Center (EPC) was established in September 1988 to provide users with access to CD-ROM databases and remote end-user search systems. The Electronic Publication Center incorporated two earlier services: (1) the CD-ROM service, founded in October of 1987, which provided users with access to two copies of Dialog's OnDisc ERIC and one copy of Compact Disclosure; and (2) the Quick Search service, established in January 1987, which provided users with access to remote databases on the BRS After Dark and Knowledge Index systems.

The following CD-ROM products were initially available in the EPC:(1) ABI/Inform Ondisc, (2) Compact Disclosure, (3) MLA International Bibliography, (4) GPO CAT/PAC (Monthly Catalog), (5) OnDisc ERIC (2 copies), (6) PAIS on CD-ROM, (7) PsychLIT, and (8) sociofile.

Three workstations were dedicated to particular CD-ROM products: two to ERIC and one to Compact Disclosure. The other four workstations were used to support the remaining CD-ROM products. These CD-ROMs were checked out to the user from either the Electronic Publication Center desk or the Reserve desk, depending on the day and time of day. Tom Wilson, Computerized Information Retrieval Service Coordinator, wrote a computer program that guides users through the initial steps of starting up the CD-ROM product and asks them brief survey questions when they have finished using the product.

A numeric database, SUPERMAP, is currently utilized by professional searchers in the Computerized Information Retrieval Service to meet library users' needs for census data. The University Libraries are developing the necessary user documentation to make this complex database available directly to library users in the future.

In 1989, the EPC was expanded. Three additional workstations were added, and the Science Citation Index Compact Disc Edition and Compact MED-BASE CD-ROM

databases were made available. The PsychLIT database replaced Compact Disclosure on one of the dedicated machines. Compact Disclosure and the two new CD-ROM databases joined the databases that were checked out from a service desk.

The establishment and subsequent growth of the Electronic Publication Center resulted in the need to significantly increase staff support for end-user searching. The Electronic Publication Center desk is currently staffed over 50 hours per week by a combination of student assistants, catalogers, Information Services librarians, and paraprofessional staff.

IRIS Project Overview

To develop the Intelligent Reference Information System, the University Libraries will network ten user workstations with a commercially available CD-ROM network system. Eight workstations will be for public use, one workstation will be at the reference desk for ready-reference searching, and one workstation will be in the Information Services Department for both LAN administration and online searching by Information Services librarians. Four workstations will be available for searching stand-alone CD-ROM databases.

From November 1989 to February 1990, the IRIS Project Management Group and the Libraries' Computer Networking Committee investigated potential IRIS network configurations. This process was complicated by changes in technology since the grant proposal was written and a dramatic expansion of the state purchasing contract so that many more types of microcomputers were included. These factors were examined in-depth, and changes to the IRIS hardware configuration were made based on them. Changes in the state purchasing contract were especially challenging. The complex network we are building is composed of a number of IBM-compatible units that must work together in harmony. Altering a single component can force the reassessment of the suitability of many other components. Since IRIS was a multi-vendor system, there was no primary vendor who could configure the entire system and certify that its components would work with each other.

The IRIS system will utilize an IBM Token-Ring Local Area Network with a IBM-compatible 80386-based file server, which will run Novell Advanced Netware 2.15. At this time, the IRIS Project has selected a CD-ROM server system; however, this hardware is still being acquired and, until the purchasing process is complete, this system cannot be publicly identified. The CD-ROM server system will accommodate up to twenty CD-ROM disks. Network workstations will be IBM-compatible 80286- and 80386SX-based microcomputers.

During the same three-month period, the IRIS Project Management Group and the Libraries' Information Services Department evaluated a number of commercially available CD-ROM databases. We were looking for a selection of databases that: (1) could be networked, (2) provided adequate coverage of the academic disciplines taught at the University of Houston, and (3) included a mix of citation and non-citation databases. After considerable internal discussion and lengthy discussions with database vendors, the IRIS Project identified the following CD-ROM databases to be used on the network:

ABI/INFORM Ondisc

Art Index

Biological and Agricultural Index

Business Dateline

Compact Disclosure

Compendex Plus

Computer Library

ElectroMap World Atlas

General Science Index

Grolier Electronic Encyclopedia

Humanities Index

Microsoft Bookshelf

OnDisc ERIC

Periodical Abstracts Ondisc

PsychLIT

Social Sciences Index

sociofile

Statistical Masterfile

SuperMap

The Index Expert system will be modified to function on the CD-ROM network as an intelligent database selection system that will link users to appropriate CD-ROM databases. The expert system will be extended to include all major types of reference materials, such as atlases, dictionaries, encyclopedias, and statistical sources. It is anticipated that the modified expert system will be primarily written in Turbo Prolog. Turbo C may be used for aspects of the system (e.g., network operations) where a procedural language is a better development tool than a logic programming language.

The IRIS Project has three primary committees: (1) The Project Management Group, which reports to the Director of the Libraries, plans and implements the IRIS system as a whole; (2) the Knowledge Engineering Group, which reports to the Project Management Group, designs and programs the expert system; and (3) the Electronic Publications Instruction Group, which also reports to the Project Management Group, coordinates IRIS instructional support and user documentation. (The Intelligent Reference Systems Committee has been replaced by the Knowledge Engineering Group.) The first two committees normally meet on a weekly basis (the last committee has just been formed). It is anticipated that an additional work group will be added to the Project to deal with evaluation issues.

Current Project members are:

Project Director

Robin N. Downes, Director of the University Libraries.

Project Management Group

Charles W. Bailey, Jr., Project Manager and Assistant Director for Systems (Chair).

Kathleen Gunning, Assistant Director for Public Services and Collection Development.

Donna Hitchings, Acting Head of Information Services.

Judy Myers, Assistant to the Director.

Thomas Wilson, Computerized Information Retrieval Services Coordinator.

Knowledge Engineering Group

Judy Myers, (Chair).

Charles Bailey.

Jeff Fadell, Information Services Librarian.

Thomas Wilson.

Electronic Publications Instruction Group

Cherie Colbert, Coordinator of Library Instruction (Chair).

Ivan Calimano, Information Services Librarian.

Carolyn Meanley, Information Services Librarian.

Derral Parkin, Head of Branch Libraries.

Other committees in the Libraries are called upon as needed to support the Project. For example, the Computer Networking Committee was involved in evaluating LAN options and the Collection Management Committee was consulted about issues related to CD-ROM database acquisition.

Initial Project activities are focused on selecting, purchasing, and assembling into a workable system the components of a CD-ROM network system. A considerable amount of energy was spent working with multiple vendors to obtain licenses for using their products in a network environment. Getting these products to run successfully in a multi-vendor local area network environment should prove to be equally challenging.

Once the CD-ROM network system is being utilized by library users, the Project will begin development of four versions of the expert system. The Project will use the existing Index Expert system as the foundation for the new system. The first version of the new system, the functional specification prototype, will be an operational mock-up of the system. The second version of the system, the expert system prototype, will be a working expert system intended as a rough draft of the system. The third version of the system, the pilot expert system, will be refinement of the expert system prototype, suitable for testing by end-users. The last version of the system, the production expert system, will be a completed version of the system intended for actual use. Prototyping is a common expert system development strategy that offers a flexible, incremental approach to the development process.

The Project will gather and analyze a variety of data about the use and effectiveness of the Intelligent Reference Information System.

The Project will acquire or develop a logging system to capture information about user activities, gather usage data, and analyze this data. Ideally, the Project would like to have the capability to sample or record all data at the transaction level (e.g., search key entered); however, at minimum, the Project intends to record data at the database-use level for CD-ROM databases and at the transaction level for the expert system.

The Project will evaluate the pilot expert system to assess the effectiveness of this test expert system with actual users. A mixture of techniques will be used: observation of user behavior, analysis of transaction logs, and a user survey.

The Project will conduct two user surveys to determine the effectiveness of different aspects of the Intelligent Reference Information System. The first survey will be conducted eight months after the initial networked CD-ROM system is installed. This survey will focus on user perceptions of the benefits and problems associated with (1) the selection of and access to heterogeneous CD-ROM databases (i.e., databases covering different types of information and also having different user interfaces); (2) the networked delivery of CD-ROM information, and (3) the use of electronic information itself. The second survey will be conducted one month after the production expert system is installed. This survey will examine user perceptions of the benefits and problems of using this expert system.

Conclusion

The primary focus of library automation efforts has shifted from computerizing back-room library functions to providing patrons with access to an increasingly diverse group of public-access computer systems.[6] As these heterogenous systems proliferate, issues related to system integration become more critical if we are to provide users with a easy-to-use and comprehensible information environment.[7] By creating an experimental expert system to

assist users in selecting appropriate networked CD-ROM databases, the Intelligent Reference Information System Project hopes to address a small, but significant, part of the larger problem of how to create future integrated public-access computer systems.

The marriage of networked CD-ROM databases and expert system technologies provides a relatively low-cost laboratory for exploring the potentials of the powerful knowledge-based electronic information systems that libraries will provide to users in the future. Libraries and vendors who will create such systems based on minicomputer or mainframe computer technology will be engaged in a very complex, expensive, and time-consuming effort. By creating the prototype IRIS system, the University Libraries can conduct investigations that will hopefully benefit future large-scale system development efforts aimed at providing knowledge-based access to electronic information.

The Intelligent Reference Information System will also provide a variety of more immediate benefits by advancing the state of the art of CD-ROM network systems and expert systems themselves. It is hoped that as the IRIS project explores these new technologies and reports its findings that this information will provide useful guidance to the library community about benefits and pitfalls of these technologies.

References

1. Charles W. Bailey, Jr., "Intelligent Library Systems: Artificial Intelligence Technology and Library Automation Systems," *Advances in Library Automation and Networking* 4 (1990), forthcoming.

2. Rao Aluri and Donald E. Riggs, "Application of Expert Systems to Libraries," *Advances in Library Automation and Networking* 2: 1-43 (1988).

3. Jeff Fadell and Judy E. Myers, "The Information Machine: A Microcomputer-Based Reference Service," *Reference Librarian* no. 23 (1989): 75-112.

4. Charles W. Bailey, Jr., "Building Knowledge-Based Systems for Public Use: The Intelligent Reference Systems Project at the University of Houston Libraries," in *Convergence: The Proceedings of the Second LITA National Conference,* edited by Michael Gorman (Chicago: American Library Association, 1990), forthcoming.

5. Charles W. Bailey, Jr., Jeff Fadell, Judy E. Myers, and Thomas C. Wilson, "The Index Expert System: A Knowledge-Based System to Assist Users in Index Selection," *Reference Services Review* 17, no. 4 (1989): 19-28.

6. Charles W, Bailey, Jr., "Public-Access Computer Systems: The Next Generation of Library Automation Systems," *Information Technology and Libraries* 8 (June 1989): 178-185.

7. Charles W. Bailey, Jr., "Integrated Public-Access Computer Systems: The Heart of the Electronic University," *Advances in Library Automation and Networking* 3 (1989): 1-33.

Published Version Citation

Charles W. Bailey, Jr. and Kathleen Gunning, "The Intelligent Reference Information System," *CD-ROM Librarian* 5 (September 1990): 10-19.

Bailey, Charles W., Jr., and Kathleen Gunning. "The Intelligent Reference Information System." *CD-ROM Librarian* 5 (September 1990): 10-19.