Online Public Access Catalog Retrieval Structures and Techniques: With Reference to Recent Developments in the United States and Great Britain

Sharon E. Clark
*University of Illinois at Urbana-Champaign*

William H. Mischo
*University of Illinois at Urbana-Champaign*
ONLINE PUBLIC ACCESS CATALOG RETRIEVAL STRUCTURES AND TECHNIQUES:
WITH REFERENCE TO RECENT DEVELOPMENTS IN
THE UNITED STATES AND GREAT BRITAIN

Sharon E. Clark and William H. Mischo
University of Illinois at Urbana-Champaign
Urbana, Illinois, USA

INTRODUCTION

The last few years have witnessed the introduction of several important
information technologies, including powerful microcomputer workstations,
optical disk and advanced magnetic disk storage media, computer graphics
technologies, the availability of nationwide and campus wide computer
networks, and the use of sophisticated search engines (such as BRS/SEARCH) in
local online catalogs. These technologies have greatly influenced all aspects
of library information services, including the provision of enhanced subject
access and access to a wider variety of bibliographic resources for library
users.

The United States and Great Britain are both taking leadership roles in
developing effective solutions to improved access and retrieval methods for
information. After a review of some of the main issues in user access to
information needs, a few important developments are highlighted to illustrate
the measures that are being taken to improve access.

ONLINE CATALOG STUDIES

Online catalogs represent, for the library users, the most visible and heavily
used end user computer interface for information retrieval (IR). Since the
introduction and rapid proliferation of online catalogs in the early 1980's,
numerous user surveys and use studies have been conducted. Principal among
these have been the federated studies conducted under the aegis of the Council
on Library Resources using questionnaires, transaction log analysis, and focus
group interviews to determine patterns of catalog use. Among the other
findings the various studies have shown that:[1]

1) Most catalog users want materials on a topic.

2) Subject searching is the predominant mode of searching;
in some libraries it accounts for more than one-half of
all searches.

3) Catalog users report the most problems with subject
searching.

4) One third to one half of all searches result in no items
retrieved.

5) Subject searches using keywords with a non-Boolean strategy,
or with search arguments which provide a partial match with
controlled vocabulary terms, often produce a large number of
citations.

6) User entered subject search terms match the Library of
Congress Subject Headings (LCSH) controlled vocabulary only
25% of the time or less.

7) Systems with keyword searching appear to receive more
subject searching.
Catalog users place the highest priority for improvements on various subject search enhancements.

Users approach online catalogs expecting to find enhanced access to a broader field of materials, including citations to periodicals at the article level, than covered by the traditional card catalog.

While the online catalog use studies indicate an overwhelming acceptance by users, including high user satisfaction, the studies have also been interpreted as a mandate for enhanced subject access.

The provision of enhanced online access to periodical information has received a great deal of attention lately. As users become accustomed to the presence of online catalogs and CD-ROM databases, they are indifferent to the form—book or periodical—in which the information they desire is packaged. The card catalogs of the late nineteenth century provided access to periodical articles via 3 x 5 cards prepared by vendors or prepared in-house. There is a renewed interest in shaping the modern catalog into an "analytical" catalog capable of fulfilling the same function but not limited to linear access.

This paper will report on projects that:

1) Employ workstation and interface software along with locally mounted databases to expand access to materials not typically found in library catalogs, such as periodical article citations; and

2) Use of augmented bibliographic records and software and user-feedback techniques to increase access to online catalog materials.

ACCESS TO THE PERIODICAL LITERATURE

Libraries are presently investigating three means of providing enhanced online access to periodical information. These are:

a) mounting commercially produced bibliographic databases in local online systems, using either the same software employed in the online catalog or employing a separate search engine;

b) providing a gateway searching service, utilizing remote search software and databases such as CARL UNCOVER or the DRA central periodical index databases, or the commercial vendor systems; and

c) making available fixed cost searching with optical disk (CD-ROM) databases, either on stand-alone workstations or in a networked environment.

While many academic libraries are implementing all three options, the latter two methods of providing access to periodical databases have been essentially independent of the local online catalog and are typically offered as separate services. There are economic, service, and access advantages and disadvantages associated with each approach. Clearly, the most comprehensive method of offering patron access to the periodical literature is to provide a complementary approach, employing a combination of the three.

Libraries have provided online access to commercial database vendors since the late 1960's, but until recently direct interaction with these systems has been
almost the exclusive provenance of trained intermediaries. However, with the introduction of CD-ROM and locally mounted periodical index databases, direct end user searching has become commonplace. In response, libraries have begun offering self service database searching facilities and training for end users.

Studies of these end user search services offered in library settings show:

1) Users are very enthusiastic about performing searches in easy to use, inexpensive systems, but are not routinely performing effective searches.

2) Users have serious problems with Boolean logic and search strategy formation.

3) Users resist formal training sessions and printed instructions, preferring computer aided instruction (CAI) and direct one to one instruction from library staff and peers.

4) The services are very staff intensive and require additional library (reference) personnel to assist in database selection and search strategy formulation and modification.

5) Users have difficulty with the simpler interfaces that employ command mode searching.

6) Searches performed with intermediaries still yield better results than those without intermediaries.

7) For searches to be truly effective they are going to have to move beyond the second generation Boolean searching which is now so commonplace.

Libraries are investigating several techniques to improve end-user access to information resources, including the development of information workstations.

LIBRARY WORKSTATION

Some of the early applications of these technologies center around the development of campus information systems comprised of scholar workstations (located both in libraries and on individual's desks) linked via a campus network to local and remote computing resources. These microcomputer workstations provide access to heterogeneous information resources and, to function effectively, must rely on customized interface software to guide users into and between the various resources available. It has become clear that the role of the user interface in information retrieval systems is of paramount importance, and the design of intelligent interfaces to facilitate retrieval has become an area of increased interest.

THE UIUC MICROCOMPUTER WORKSTATION

Microcomputer workstations play an important role in the University of Illinois at Urbana-Champaign (UIUC) Library's online systems. They are used both as public terminals and staff workstations and employ interface and gateway software: to facilitate access to the campus and statewide online catalog; to access locally mounted periodical index databases; to provide a gateway function to external online catalogs and other databases; and to provide access to bibliographic, referral and directory information databases.
stored on microcomputer hard disk and optical disk files. The effective
utilization of microcomputers is made easier in the Illinois setting because
the UIUC Online Catalog employs microcomputers as public terminals.

The Library Information Workstation is an IBM-based microcomputer software and
hardware platform used as an online catalog public access terminal in the UIUC
Library system. The UIUC Library is seeking to establish a unique
demonstration site for the design and testing of state-of-the-art public
access workstations. These workstations will provide access to a wide variety
of local, statewide, and national information resources.

The UIUC Library Information Workstation has been in development for several
years. The workstation (now available on 95 PS/2) terminals features expert
system interface and gateway software, and graphic interface techniques that
provide library users with access to: the ILLINET Online (IO) statewide online
catalog, a local BRS/SEARCH implementation with associated periodical index
databases, the CARL UNCOVER system, and customized databases stored on the
individual microcomputer workstations. Figures 2-7 follow a subject search
from the selection of the database through linking the journal citation to
the UIUC Library's call number and location.

The potential now exists to develop a state-of-the-art, full-featured, Library
Information Workstation. The UIUC staff will focus on the design and develop­
ment of the microcomputer workstation software that allows library patrons and
staff to effectively access and link multiple information resources from one
common interface and gateway software, advanced multimedia and graphic tech­
niques (including voice synthesis and pen/hand lettering input), and a
database management module for files stored on the microcomputer. This
workstation development has been and will continue to be implemented and
tested in operational environments with library users in the UIUC library
system.

The software features and functions of the Library Information Workstation
will continue to evolve over the next several years. Many workstation
functions already have been at least partially implemented. These include:

1) Access to the statewide online catalog (IO) and the holdings
of over 800 libraries in Illinois through customized
microcomputer and mainframe interfaces;

2) Access to bibliographic databases (including full-text) and
non-bibliographic databases through the campus network,
UIUCNET (TCP/IP) and over the Internet, the Consortium for
Cooperation's CICNET, etc. Specifically provided will be
access to select Association for Research Libraries online
catalogs which are available via the Internet. Network
database access and interface software are two emerging
information technologies for academic libraries.

3) Access to a multi-campus local BRS/SEARCH implementation
with associated periodical indexes (Current Contents,
MEDLINE, six Wilson files, AND ERIC) with customized
interfaces, which are transparent to the user, for each
database. Plans call for mounting full-text files (campus
information and news) and extending access to the statewide
environment.

4) Access to remote database vendors (BRS, DIALOG, STN, etc.)
via a gateway function, either dial-up or through a network
such as the IBM IIN.
5) Access to databases stored on CD-ROM (including full-text) and their search engines, run from the interface using shell software.

6) Access to the OCLC EPIC search system and other national bibliographic databases, such as the Library of Congress catalog.

7) Access to customized databases stored on the microcomputer workstation's disk drive(s). These databases are site specific and can be easily updated. Databases may include hot topics and customized bibliographies, and pathfinders; bulletin board files; serial and new book lists; class reserve lists; calendars; current contents databases; and customized files downloaded from the online catalog, BRS searches, or remote vendor searches. These ASCII databases can be searched with a fast sequential searching database management system developed by the UIUC library or another system such as a package purchased with a database.

8) The capability of invoking specific software packages from the interface and returning to the interface main menu level.

9) Graphic files such as building maps, mixed graphics, sound, and text (hypermedia files). The Audio Visual Connection (AVC) software from MCA architecture has been used to integrate AV, interactive graphics, digitized images, video images, real-time multimedia.

10) Point-of-contact instruction and the use of customized expert system and multimedia software (emphasizing voice synthesis) to provide search navigation and access resources such as reference collection databases and pathfinders and refer patrons to both bibliographic and human resources that can best meet their information needs.

11) The use of optical scanning technologies to digitize text and contents pages, reserve readings, graphical images, etc. and the delivery of these bit-mapped text files over campus and national networks.

12) The capability of linking these separate modules for enhanced user searching, such as linking elements of local microcomputer-stored databases with online catalog information (holdings, availability), linking retrieved references in a local BRS search to the online catalog, or linking online catalog data to information in the OCLC EPIC national database. One important feature already implemented is a link between a journal title file (with call numbers and library locations) and the Current Contents database. With this link, the library call number and location(s) are displayed, printed, or downloaded with each Current Contents citation.

13) An electronic mail function to provide library users with a reference referral and a suggestion option.

14) Customized user interfaces, window-based menus and displays, and mouse support.
OKAPI AND ITS IMPLEMENTATION

Funded by the British Library Research and Development Department, OKAPI, provides a cutting-edge system in addressing the search design in the online catalog.[4,5] The system provides automatic search sequencing mechanisms using Boolean operators on different combinations of search terms which allows for retrieval of records containing all terms in a search (exact match) and also those containing some of the search terms (partial match). Using a term weighting algorithm, records are then displayed in rank order with the best match first. Using a relevance feedback model, the user is asked to indicate which reference in a retrieved set is most useful. The system responds by automatically reformulating the search and looks for references on the basis of the user's relevance judgments. That is, when the user enters a search in natural language, the system asks the user if the response is what was intended. If the answer is yes, the system automatically looks at title terms, subtitles, subject descriptors in LSCH or PRECIS, series and Dewey class numbers. It then weights, re-ranks and re-displays the results with the most relevant item first. Future plans include interface development, use and display of the thesauri directly to the user and making classification automatic. There is also an interest in improving concurrence since until now only single terms have been used. The system is also used for INSPEC and LISA.

While OKAPI has been a research and development project for the past eight years, the principles developed by Stephen Walker and his colleagues have begun to be incorporated in other catalogs, e. g. LIBERTAS.

The Centre for Bibliographic Management (University of Bath) and Stephen Walker (City University of London) are collaborating on a British Library R & D Department funded project to examine the design and use of a database of subject-enhanced catalog records.[6] The project, lasting approximately two years, began at the end of 1989. In 1991 a system for live use will be tested at the University of Bath Library. The project involves using existing machine-readable data to enhance the catalog records, and mounting them on a system which can exploit the additional information more fully than most existing OPAC's.

The search system will be a version of OKAPI. The University of Bath catalog records for the most recent three years will be mounted and selectively enhanced with abstracts, content pages, and other subject-descriptive materials derived from Book Data files. It is estimated that there will be matching Book Data records for between 25% and 40% of the Bath records. The project when completed will move the OKAPI research one step further to a practical and functioning search system.

Online access to databases is being provided by the U.K. Office for Library Networking, located at the University of Bath. The Combined Higher Education Software Team (CHEST) made an agreement with the Institute for Scientific Information (ISI) database producer to mount databases and make them accessible to participating libraries on Britain's Joint Academic Network (JANET). The initial databases mounted include Science Citation Index, Social Science Citation Index, Arts and Humanities Index, Index to Scientific and Technical Proceedings, and Research Front Data. Some 50 higher education sites will participate at a cost of 5,000 pounds sterling plus VAT per site for four years from April 1991. The service called BIDS (Bath ISI Data Service) will begin with a hundred ports and increase to 300 ports by the end of 1991. Mounted on STATUS, there will be a menu driven interface. An Interface Working Party to elicit input from librarians on the interface design has been established. Full service with free access for staff and students at participating sites is due to be available this summer.[7]
The Scottish Academie Libraries Bibliographic Information Network (SALBIN) at the present includes the eight Scottish university libraries plus the National Library of Scotland and just recently the Edinburgh and Glasgow public libraries [8, 9]. The SALBIN group was formed in response to the universities awareness that budget constraints of the previous decade had seriously eroded library services to users both through collection building and library staff. In 1985, the Scottish librarians were asked to consider resource-sharing as a vehicle towards improving library support for teaching and research in Scotland.

Following a feasibility study funded by the BLR & DD in 1987, University Grants Committee funding was awarded to Edinburgh University to implement the first stage of SALBIN, the development of a user friendly communications package that would provide access to all of the SALBIN library OPAC's. The result is one of the most interesting gateway systems via networks to online catalogs. The SALBIN interface is designed to give library users easy, menu-driven access to the online catalogs of the other libraries in the group, and also via sub-menus, to those of other UK university and polytechnic libraries accessible over JANET. The software runs on an IBM PC or clone preferable with a hard disk drive. It offers the user catalog access taking him/her directly into the chosen library’s OPAC, and the option to log the session to an attached printer or download to a diskette. Further features include an alternative SALBIN EXIT and facilities for extra Help screens for remote users.

The interface is distributed in two versions. The User's Version is designed to be as friendly and intelligible as possible by masking log on and communications reports. This version will be the standard public access version. Built in defaults and time outs are provided as well as a safeguard to prevent users from deleting the system. Provision of a direct PAD connection on the main menu in the Edinburgh User's version provides for campus facilities, such as E-mail, without exiting from the program. The Professional Version allows for local tailoring and maintenance, such as controlling and changing menus and Help Screens. Changes in campus network and JANET protocols, and increasing number of accessible OPAC's have led to the decision to give participating libraries the guidance required to amend "Connect" calls also rather than attempting to issue new releases.

Following an initial Welcome/Logon screen, the user can then choose general information help screens or proceed directly to the main menu. Upon choosing an option, the system makes the connection to it, flashing a message to keep the user informed, as the time taken to respond will vary with the state of the networks involved as well as the speed of the microcomputer being used. The user may print or download the session onto a diskette.

A Help screen provides guidance for printing or logging individual screens rather than entire sessions. If neither option is chosen, the system defaults to search mode. The user may then to search the chosen OPAC, experiencing the advantages and limitations of a particular system, using it's bulletin Help facilities, since there is facility in SALBIN for site-specific Help screens which can be incorporated within each library's version. Use of a particular OPAC's log-off directions will trigger an instant return to the SALBIN menu. If, as in so many systems, these directions are not visible at the point from which one wishes to exit, then F10 can be used as an emergency exit.

CONCLUSIONS

The budget constraints facing libraries in the U.K. and the U.S.A. resulting in less staff available to help users use OPAC's and perform online searching
are providing the impetus for improving user access and retrieval. The necessity to make users independent consumers of library resources, i.e. self-service is putting the user needs in the forefront. One reason an interface like SALBIN is so important lies in its' potential for capturing real user's responses to OPAC's. At the University of Edinburgh, when library users were introduced to searching other OPAC's via SALBIN, they actually made suggestions for how to improve Edinburgh's OPAC based on features they had seen in other systems. Libraries and vendors always have user groups but how often, if ever, have real patrons been tapped as a resource for OPAC development?

Libraries have the opportunity to capture user enthusiasm and use it as a power-base to insist on providing improved access and retrieval systems. The perfect system will not happen overnight. However, it will not happen at all without making librarians and users part of the development process. This is an exciting, but complex, time for users and developers of online information systems. We have a rich variety of information resources and information technologies to offer to library patrons. There is no doubt that increased systems potential and user expectations are challenging the structure of today's OPAC's. An online catalog doing no more than replicating the paper systems must no longer be tolerated. Given this challenge it is heartening to realize that effective answers are being sought.

REFERENCES

4. HANCOCK-BEAULIEU, MICHELINE. Interview at Centre for Interactive Systems Research, Department of Information Science, City University, London, October 15, 1990.
5. WALKER, STEPHEN. Interview at Centre for Interactive Systems Research, Department of Information Science, City University, London, October 17, 1990.
6. DEMPSEY, LORCAN AND STEPHEN Prowse. Interview at Centre for Bibliographic Management, University of Bath, October 12, 1990.
7. BRYANT, PHILLIP. Interview at Centre for Bibliographic Management and U. K. Office for Library Networking, University of Bath, October 12, 1990.