Volume 8
Number 1

ISSN 0730-9295

Information Technology and Libraries

March 1989

CONTENTS

ITLBDC 8(1) 1–96 (1989); Inf Technol Libr

Ulrich's *Plus*™ Another Plus!

Reshaping The Way You Work With Periodicals Forever

ULRICH'S *PLUS*™ combines the remarkable technology of CD-ROM and ULRICH'S INTERNATIONAL PERIODICALS DIRECTORY DATABASE to create the most powerful periodicals acquisition and reference tool ever.

ULRICH'S *PLUS* has all of the comprehensive data now included in ULRICH'S INTERNATIONAL PERIODICAL DIRECTORY with IRREGULAR SERIALS AND ANNUALS along with the ULRICH'S update – merged together to revolutionize the way you find and order periodicals.

LOCATE THE DATA YOU NEED IN SECONDS

ULRICH'S *PLUS* is so incredibly fast and efficient, at just the touch of a few computer keys, you can search for data in 22 different categories, alone or in combination – including...

- Title
- Editor
- Keyword
- Year First Published
- ISSN
- Frequency
- Online Availability
- Country Code
- Subject
- Publisher
- Price
- Circulation
- Dewey Decimal
- Status Code
- Media Code
- U.S. State/Zip
- Abstracting and Indexing Services
- Special Features (including book/film/record/dance reviews, advertising, charts, illustrations, maps, statistics)
- Special Index (including International Congress proceedings, UN publications, European Communities publications)

...and much more.

ULRICH'S *PLUS* is so thorough, it's possible to search with just a fragment of a title, publisher's name or keyword – so complete you'll be able to find detailed contact information for every publisher in the database.

ULRICH'S *PLUS*
THE MOST COMPREHENSIVE PERIODICAL DATABASE EVER

Now the same comprehensive periodical information you've always depended on is yours on a single compact disc:

- More than 143,000 records from 61,000 publishers in 184 countries
- Brief descriptions of all periodicals
- LC classification numbers
- CODEN designations for thousands of entries
- Identification of publications available online or on CD-ROM.

PLUS QUARTERLY CUMULATIVE UPDATES

Quarterly updates assure that you'll receive complete and up-to-date information on new titles and revisions to old titles throughout the year. **ULRICH'S *PLUS*** also retains all ceased titles so you'll always know the status of any publication.

AND THE CD-ROM INTERFACE IS INCREDIBLY EASY TO USE

The same interface that is used for BOOKS IN PRINT *PLUS*™, BOOKS IN PRINT WITH BOOK REVIEWS *PLUS*™, and BOOKS OUT•OF•PRINT *PLUS*™, as well as other CD-ROM products, is also used for **ULRICH'S *PLUS***. There are no new commands to learn, no new searching strategies to become familiar with, and no problems switching between CD-ROM discs.

Sensibly priced at $395 for a one-year subscription or $1,126 for a cost-effective three-year subscription – YOU CAN'T AFFORD NOT TO HAVE **ULRICH'S *PLUS***.

ORDER YOUR SUBSCRIPTION NOW!

Call **TOLL-FREE 800-233-3288**
(In NY, AK, HI call **COLLECT 212-337-6989**)

Or **SEND YOUR ORDER TO:**
Bowker Electronic Publishing
245 West 17th Street
New York, New York 10011

ULRICH'S *PLUS*™
Incredibly powerful, easy to use, right for you!

Bowker Electronic Publishing,
a Division of R. R. Bowker, The Information Reference Company

ISSN 0730-9295

Information Technology and Libraries

Volume 8, Number 1: March 1989

CONTENTS

ITLBDC 8(1) 1–96 (1989); Inf Technol Libr

Information Technology and Libraries

Information Technology and Libraries is the official publication of the Library and Information Technology Association, a division of the American Library Association, 50 E. Huron St., Chicago, IL 60611; *Executive Director:* Linda J. Knutson. The journal is issued quarterly in March, June, September, and December.

Information Technology and Libraries publishes material related to all aspects of library and information technology. Some specific topics of interest are: Automated Bibliographic Control, AV Techniques, Communications Technology, Cable Systems, Computerized Information Processing, Data Management, Facsimile Applications, File Organization, Legal and Regulatory Matters, Library Networks, Storage and Retrieval Systems, Systems Analysis, and Video Technologies. *ITAL* welcomes unsolicited manuscripts. Submissions should follow the guidelines stated under "Instructions to Authors" on page 83 of the March 1989 issue.

Manuscripts of articles, communications, and news items should be addressed to the editor-designate: Thomas W. Leonhardt, *Information Technology and Libraries*, University of the Pacific, Stockton, CA 95211. Copies of books for review should be addressed to: Karin A. Trainer, *ITAL Book Reviews*, Yale University Library, 120 High St., P.O. Box 1603A, Yale Station, New Haven, CT 06520. Copies of software for review should be addressed to: David T. Buxton, *ITAL Software Reviews*, A349 Main Library, University of Arizona, Tucson, AZ 85721. Advertising arrangements should be made with Kurt Murphy, Hayden Library, Arizona State University, Tempe, AZ 85287.

Information Technology and Libraries is a perquisite of membership in the Library and Information Technology Association. Subscription price, $17.50, is included in membership dues. Nonmembers may subscribe for $35 per year in U.S., Canada, Mexico, Spain, and other PUAS countries; $40 in other foreign countries. Single copies, $9. Second-class postage paid at Chicago, Illinois, and at additional mailing offices. *Postmaster:* Send address changes to *Information Technology and Libraries*, 50 E. Huron St., Chicago, IL 60611.

Circulation and Production: Central Production Unit (Eileen Mahoney, Director of Production and Distribution Services; Chere Elliott, CPU Supervisor; Bruce Frausto, Debra M. Hall, Colleen King, and Debra J. Teague), American Library Association, 50 E. Huron St., Chicago, IL 60611.

Abstracted in *Computer & Information Systems, Computing Reviews, Information Science Abstracts, Library & Information Science Abstracts, Referativnyi Zhurnal, Nauchnaya i Tekhnicheskaya Informatsiya, Otdyelnyi Vypusk*, and *Science Abstracts Publications*. Indexed in *CompuMath Citation Index, Computer Contents, Computer Literature Index, Current Contents/Health Services Administration, Current Contents/Social & Behavioral Sciences, Current Index to Journals in Education, Education, Library Literature, Magazine Index, NewSearch*, and *Social Sciences Citation Index*. Microfilm copies available to subscribers from University Microfilms, Ann Arbor, Michigan.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984. ♾

The "New" NOTIS/LUIS Indexes

Velma Veneziano

In its NOTIS/LUIS system, Northwestern University Library has recently implemented a new set of indexes that provide author, title, and subject access to a database of nearly 1.5 million bibliographic and authority records in the NOTIS/MARC format. This paper provides information on the index record structure and content, outlines the design philosophy and process, and describes the functions that are supported and how this is done.

For several years the Information Systems Development Office at Northwestern University Library, in close cooperation with library staff, has been engaged in a major project to redesign and improve the bibliographic/authority indexes that provide author, title, and subject access to MARC records.

The early indexes served the library well, enabling it in 1981 to close the author/title card catalog and to close the subject catalog in 1982. The aim of the project was to enhance rather than replace these early indexes. The redesigned indexes would follow their model, capitalizing on the strengths of the early indexes but remedying some of their deficiencies and increasing their power and functionality.

THE EARLY INDEXES

Each index record derived from a bibliographic record consisted of a heading derived from one field and precoordinated with a heading from another field. To this pair of headings was added some additional supplementary information (date or place of publication, location/call number, etc.). The net effect was that each index record constituted a "brief entry" catalog record, identifying, more or less distinctively, a "bibliographic unit." Users could browse through an alphabetically ordered list of these brief entry records to locate the entries of interest to them. From this list they could then call up full bibliographic and holdings records complete with circulation or acquisitions status. This technique had the advantage of minimizing the number of full bibliographic, holdings, and item records that had to be accessed and displayed.

The early indexes were stored in large blocks with both front-end compression of common characters and back-end compression of trailing blanks. This compactness minimized the number of input-output operations and resulted in high system efficiency and good response time. The latter was an important factor in user satisfaction.

Searching was performed using a simple FIND command that could be omitted in public mode, followed by a one- or two-character index identifier [author, title, subject (LCSH), subject (MeSH), authority], followed by a search term. Left-anchored string search with implicit right truncation was used to find a group of index records, which were then displayed, multiple index records per screen.

GOALS OF THE INDEX REDESIGN PROJECT

Despite the simplicity, efficiency, and effectiveness of the early indexes, it became

Velma Veneziano is Library Systems Analyst, Northwestern University Library, Evanston, Illinois.

clear by 1983 that it was time for a new generation. The index redesign project—its process, problems, and outcome—is the subject of this paper.

Improved access, obviously the main goal, meant not only simply improving recall by increasing the number of access points but also enhancing precision and relevancy. Improved access called for a syndetic structure enabling the user to be referred from a variant heading to an established heading and from one established heading to another related, established heading. It meant that all the works under a particular heading would be collocated. It meant dynamic updating.

THE PROCESS

The task was formidable and time-consuming. Functional specifications were first prepared and reviewed by a task force of librarians representing a large number of library departments. After several recyclings through the review process, the specifications were finalized in March 1983 to the point where detailed design and programming could begin. At about that time, staff of the Information Systems Development Office were diverted to a major redesign of the circulation module of NOTIS; as a result, actual programming had to be postponed until late 1985.

The first test versions of the new indexes became available to staff in mid-1986. These test versions were reviewed by the members of a second task force, who identified a number of changes that they felt were needed and in late 1987 formally approved the indexes from a content standpoint. The new indexes were incorporated into LUIS (the public access mode of NOTIS) in September 1988.

Although several refinements remain to be added, all the major goals of the index redesign project have been achieved.

DESIGN CHALLENGES

Systems and programming staff, as well as consulting librarians, came to the design process committed to the idea that the new indexes should be an elaboration of the old ones, not a radical departure from them. They were convinced that an index record structure in which multiple headings were precoordinated had advantages over records in which the terms were single words or phrases requiring coordination at search time. Such a structure makes it possible to:

- relate a heading in one record to all records in which that heading occurs, regardless of the source record type or the function of the heading in the source record, and
- relate any particular heading to other headings in the same record.

Commitment to Multifunctionality

A major concern of the software designers was that the new indexes be simple, flexible, and easy to maintain and manipulate. The structure had to be sufficiently generalized so that indexes could be adapted to changes occurring in the MARC format. The rules for deriving index records from source records had to be concise, with an absolute minimum of special cases.

Keeping the records simple and flexible proved a real challenge due to the fact that a single index file had to meet the needs of a wide variety of users, with diverse needs and different viewpoints as to what was essential and nonessential. Although the needs of the public were paramount, the needs of staff, which differed from department to department and from function to function within departments, were equally important. Staff doing book selection had different needs than those using the system for ordering books and serials. Staff responsible for processing invoices had needs that differed from those checking-in items. The needs of staff checking-in serial issues differed from those checking-in new books. Staff creating item records and printing spine labels had special needs, as did staff in the bindery section. Circulation staff's needs differed from those of reference department staff. Catalogers' needs differed from those of staff doing pre- and postorder searching.

Needs of several "special" libraries using the system also had to be met. At Northwestern, NOTIS is used not only by the main library on the Evanston campus but also by a library that specializes in materials for the field of transportation. The transportation library uses NOTIS for the "indexing" of journal articles as well as for the cataloging of books and serials. Since NOTIS includes a course reserve module, the needs of reserve room staff had to be taken into consideration. The university archives are maintained by the library, and the staff of that department had special needs. The music library had unique problems. The law and medical libraries on the Chicago campus had needs that were different from those of libraries on the Evanston campus. If the indexes were to serve the needs of all users, those differences had to be reconciled; the design process involved many compromises and trade-offs.

To Keyword or Not to Keyword

Very early in the design process, it was decided that the "new" indexes would not be keyword indexes. Although they had strong advocates, there was agreement that keyword indexes were not satisfactory as the *sole* means of access to a catalog. Even if Boolean searches of keyword indexes could be proved superior from a retrieval standpoint to browsable, brief entry indexes, there was concern over what they would do to response time, given the large size of the database (more than 1.5 million bibliographic and authorities records) under high transaction loads on a small computer. There was also concern that casual users of the catalog would not know how to search efficiently and effectively using Boolean logic and would miss materials because they did not understand the inherent shallowness of cataloging data as compared to indexing and abstracting data.

DESIGN DETAILS

Effect of NOTIS/MARC Format on Index Redesign

A factor that influenced design of the new indexes was the structure and content of the NOTIS/MARC bibliographic and authority records. With the exception of the way directory information is carried, the format of the NOTIS online records differs only in minor respects from the USMARC communications format. Headings are carried as text strings in the bibliographic records, not as pointers to authority records. Authority records do not carry pointers to bibliographic records. An important function of the new indexes would be to associate all occurrences of any particular heading, providing a mechanism for linking headings without the use of explicit pointers.

Given the fact that each NOTIS/MARC record is a "unit" record, containing one or more headings which exist in physical isolation from the same headings in other NOTIS/MARC records, the indexes, to be maximally useful, had to provide more than just passive links between the same heading in different records. In order to actively assist in the maintenance of referential integrity, each index record had to carry control information that could be used to detect invalid or missing relationships (established headings that conflicted with reference headings; referenced headings which, if used as a search term, would result in a failed search; etc.).

A further factor influencing the redesign was the structural complexity of the MARC authority and bibliographic formats. Records in these formats are difficult to manipulate, and editing them to a user-friendly format is processing intensive. For this reason it was important that an index record format be developed which would minimize the frequency with which the full MARC records had to be accessed, manipulated, and displayed. Each index record needed to be, in effect, a "brief" bibliographic record.

Indexes of this type require more storage than simple control number linkages carried in the source bibliographic and authority records. At one time the cost of storage to duplicate headings in two places (in the bibliographic/authority records and in the brief entry indexes) would have been cause for concern; however, with decreased storage costs this concern has diminished, especially since data compression techniques can be used to reduce storage requirements by a third or more.

The decision to have a single "dictionary" index file, instead of separate author, title, and subject index files, was also influenced by the NOTIS/MARC format. A heading in a MARC authority record, whether an established heading or a reference heading, can be coded as appropriate for use as either a main/added entry or a subject or both. This led naturally to the conclusion that a single index record from an authority record could likewise be coded to function as a reference in more than one type of search. This also allowed index records from bibliographic records to be coded so that they could be retrieved by more than one type of search. Experience had demonstrated that conference names are not necessarily viewed by the user as "authors"; they need to be retrieved also by a title search. The concept of a uniform title heading used as a main entry is, moreover, hard to explain to anyone not intimately involved in cataloging. Because uniform titles headings are used as either main or added entries, some staff may want to search them as names, but the public and other staff must be able to retrieve them in a title search.

Computing and Programming Resources Affecting Design

If a single index file was to serve all types of users, major hardware and software barriers had to be overcome. The designers had to be concerned not only about system efficiency but also about staff efficiency. These dual concerns led to a decision that update of the index file should be performed dynamically and that the index file should be a dictionary file. Dynamic update reduced the need for frequent index regeneration, and the dictionary approach reduced the number of index records which had to be generated. Dynamic update enabled staff to see immediately the results of modifying a record, and the dictionary file enabled them to save time in searching. The fact that all index records are interfiled in a single sequence, regardless of use, also simplified the design of programs that monitor the database to detect invalid relationships and interrecord inconsistencies.

File and Record Size Considerations Affecting Design

In addition to concern over disk space for storage of the index file, another environmental factor that strongly influenced index design was disk space available for sorting. Although online update of the indexes reduced the frequency of index generation, an occasional regeneration was unavoidable. Since the amount of disk space available to the library was limited, record and file size were important considerations.

Considerable research went into determining the optimum index record size. The size had to be large enough so as not to jeopardize collocation nor seriously interfere with the ability to distinguish one work from another. At the same time it had to be small enough so as not to require more sort work space than was available and so that the time required for sorting would not be excessive. The maximum size of an index record finally settled on was 216 characters. With 1.5 million bibliographic and authority records and more than 5 million index records, regenerating, sorting, and reloading the indexes presently takes between five and six hours.

Record Structure

Because so many different record types are intermixed in a single index file, it was necessary to develop a very generalized record structure, one which was independent of the types of data carried by the structure. At the highest level of structural abstraction, an index record is divisible into two mandatory *fields*: a variable length *entry field* (with a maximum of 190 characters) and a 26-character, fixed-length *control data field*.

The entry field is divided into five *segments*. Each of the first four segments carries one or more *headings*. A heading may be either a name or a title or a subject term. A segment, in many cases, is derived from more than one MARC subfield in a variable MARC field. Segments are divisible into *subsegments*, with each subsegment carrying a logical heading or term in a heading.

Segments 1, 2, and 3 are grouped into a data aggregate termed the *full heading*. Segment 1 is referred to as the *simple heading segment*; segment 2, as the *subheading segment*; and segment 3, the *subdivision segment*.

Segment 4 contains a heading that, if from a bibliographic record, is referred to as the *qualifying heading* or, if from an authority record, a *referenced heading*.

Segment 5 contains supplementary (nonheading) data and is used only in index records derived from bibliographic records.

The maximum size of the full heading is

120 characters. A referenced heading segment cannot exceed 70 characters; a qualifying heading cannot exceed 58 characters; a supplementary data segment cannot exceed 12 characters.

Developing a coding structure that would maximize the use of the 26-character control data field proved to be a real challenge. This field is divided into a number of fixed-length subfields, plus a number of flag bytes, each of which carries eight 1-bit flags. Because of the variety of data types, a tree-structured format was necessary to enable a particular code or flag to take on different meanings depending on data elements that precede it in the field.

Deciding which fields to index (access fields) and how many index entries should be created per access field was another area where many trade-offs had to be made between what was "nice" and what was "necessary." The following access fields from bibliographic records are presently indexed: (1) all 1xx main entry fields; (2) 700, 710, 711, 730, and 740 added-entry fields; (3) 400, 410, 411, and 440 series statement fields; (4) 800, 810, 811, and 830 series added-entry fields; (5) 212, 214, 240, 242, 243, 245, 246, and 247 title fields; and (6) 600, 610, 611, 630, 650, and 651 fields. All 1xx, 4xx, and 5xx fields in authority records are indexed. As additional processing power and storage become available, and as the need is demonstrated, other fields can be added.

Sequence Problems

Lack of disk space to sort and store very long index records made it impossible to carry two forms of a heading in the record—one normalized[1] for sorting and one for display. Designers and librarians alike had no choice: index entries had to display in essentially the same way as they were stored, in uppercase without conventional inter- and intraelement punctuation.

Even by drastic normalization, it was not possible to preserve all the filing conventions used in the traditional card catalog. With one exception (birth dates for personal names), numeric digits embedded in access fields file character-by-character (i.e., as "words"), in IBM collating sequence (with numeric characters after alphabetic characters). Theoretically, it would have been possible to use very sophisticated techniques which would produce sequences that more closely conformed to conventional catalog filing rules; however, such techniques were ruled out as impractical in an online update environment.

An exception was made for names with birth dates. A major effort was directed to preserving the principle that a name without a birth date files directly ahead of that same name with a birth date. The following examples illustrate how the original version of the indexes, which used the standard IBM collating sequence, violated this principle:

JONES JOHN
JONES JOHN J
JONES JOHN 1908 1978

A programming solution was finally arrived at which manipulated the bit configuration of numbers derived from birth date subfields to alter the collating sequence to achieve the following sequence:

JONES JOHN
JONES JOHN 1908 1978
JONES JOHN J

This means, of course, that when the index record is displayed, the bits have to be remanipulated.

Many of the librarians who acted as design advisers would have preferred that titles in a series be listed under the series entry in volume number order. Given the extreme variability in the $v subfield in series statement and added entry fields, this was not practical. The compromise which was worked out was to subfile from the series heading to the title of the volume and then to the volume number. Given the ease of browsing through a list of titles under a series, this presents no problem unless the list is very long. To minimize the problem of long files for the user who knows the number but not the title, search qualification (described later in this paper) may be used.

Accommodating Variations in Relationships in the Format

Developing a generalized "entity-type

insensitive" structure that would be hospitable to the wide variations in content from one record to the next required identifying all possible types of relationships that could occur between entities, such as

- relationships between two headings in a single record,
- relationships between different portions of a single heading in a single record, and
- relationships between a heading from one record and the same heading in another record.

Once these relationship types were identified, conventions had to be developed to build information about them into the indexes in such a way as to achieve a reasonable order without adversely affecting retrieval or processing efficiency.

A number of different types of relationships were identified. Some relationships were determined to be *supportive*, as in the case of the relationship of an established heading in an index record derived from an authority record to the same heading in another index record derived from a bibliographic record. Another type of relationship identified was an *associative* relationship. An associative relationship is defined as one in which two headings, although they come from the same record, are either different entity types (e.g., one is a name and one is a title), or, if the same entity type, each has a different relationship to the bibliographic item described by the record (e.g, one is a composer and one is a performer). Other elements have a *subordinate* relationship to the data which they follow (e.g., the relationship of a subject subdivision to the subject terms which precede it in a heading). Other elements have a *supplementary* relationship to what precedes them in the entry (for example, the relation that a date or a place of publication or a volume number has to a work). In an index record in which a reference heading is paired with a referenced heading, the referenced heading has a *corresponding* relationship with the reference heading.

Some of the above-identified relationships occur between records (interrecord relationships); some relationships are between two segments in a single index record.

The Content of the Index Record

Given the limitations on record size and given the fact that so many data types had to be accommodated, it was a real challenge to decide exactly which data elements should be defined, how they should be "named," what rules should govern them, and how to locate them in the record.

Complicating matters further was the fact that only in rare cases does it happen that all segments that potentially *can* appear in an individual heading *will* appear. Dealing with missing segments was as much a problem as dealing with entries that require more segments than the record structure would allow.

Ideally, qualifying headings and referenced headings should have been structured exactly as for full headings (i.e., be composed of up to three segments: a simple heading, a subheading, and a subdivision segment) with a maximum size of 120 characters. Record size considerations ruled this out.

Ideally, also, more than one segment should have been allowed for supplementary data, enabling, for example, a work to be identified by both place of publication and date of publication, and perhaps even by edition. Again record size considerations made it necessary to limit the number of supplementary segments to one per index record.

It was decided that the supplementary data in entries derived from serial records would be place of publication; entries derived from series statements and series added entries would be supplemented with volume numbers; and all other entries would carry date of publication.

Only rarely does the maximum number of segments occur, as in the following example from a name/title subject heading with subject subdivisions:

SHAKESPEARE WILLIAM 1564 1616
.KING HENRY IV—BIBLIOGRAPHY
.HENRY THE FOURTH PART ONE
<1977>

In the above example the first heading, which occupies two display lines, is derived from a 600 subject heading field consisting of personal name plus a uniform title plus a subject subdivision. The third line carries the piece title, which is followed by the date of publication.

Most index records have fewer than five segments, with an average of three. Sometimes, as in the case of bibliographic record without a 1xx field, the index entry derived from the 245 field consists solely of a title heading followed by a date or place of publication, as in the following example:

JOURNAL OF THE AMERICAN ACADEMY OF PSYCHOANALYSIS <NEW YORK>

If another index record with the same title were qualified by a heading from a 1xx field, it should file *after* the record without the qualifying heading.

Further contributing to the difficulty of finding a structure that would accommodate all data types and all combinations of relationships was the fact that, within a single index entry, two separate elements, not necessarily adjacent, will each have the same relationship to the elements that precede them. In the Shakespeare example above, the uniform title and the name have an associative relationship to each other, but the qualifying heading segment (derived from the 245 field) also has an associative relationship to the subject heading that it follows.

A technique had to be developed to ensure that, when two records had the same heading but one was from an authority record and the other from a bibliographic record, the authority record filed first. Further, if multiple records—all from authority records—contained the same heading, records from 4xx tracings filed after those from 1xx fields but before those from 5xx tracings. Further, if there are multiple records from 5xx tracings, "earlier" headings must file together, and "later" headings must file together. For example, a record containing the established heading AMERICAN GAS ASSOCIATION from the 1xx field of an authority record must sort ahead of a record with the heading AMERICAN GAS ASSOCIATION derived from a 5xx tracing derived from a different authority record. Both these records must sort ahead of a record with the heading AMERICAN GAS ASSOCIATION derived from a field in bibliographic record.

After much experimenting, the technique decided upon involved defining two separator characters: a *segment terminator character* and a *segment lead character*. Each segment after the first starts with a lead character that indicates the relationship of the segment (associative, subordinate, qualifying, supplementary, or corresponding) to the segments that precede it. Each of the three segments that constitute the full heading must end with a terminator character. The fourth segment (either a qualifying or referenced heading) must end with a terminator character unless it is the last segment in the entry. The fifth segment, since it is always the last segment, does not require a terminator. The terminator and the lead character combine to control sequence and to insure that headings that follow other headings subfile according to their relationship to the heading that they follow.

There were only a limited number of characters suitable for use as separators. The characters picked for lead characters were as follows:

- A period (.) indicates an associative relationship.
- A left angle bracket (<) indicates a supplementary relationship.
- A hyphen (-) indicates a subordinate relationship.
- A numeric code (1, 2, etc.) indicates a corresponding relationship. By using different numeric digits, correspondence subtypes (established heading, *see also* reference heading, earlier heading, later heading, *see* reference heading) can be defined.

Because all authority-derived records with a particular heading must file ahead of all bibliographic-derived records with the same heading, a null character was picked as a terminator for the *last* segment of the full heading derived from an authority record, which might be a simple heading, a subheading, or a subdivision segment. In all other cases, for both bibliographic- and authority-derived records, a segment that is not the last segment in the entry is terminated with a space.

Name/title headings posed a particular

filing problem, since it was desired that a title in a name/title heading file with the same title derived from a 245 field. In order to file correctly, both title segments (regardless of the access field from which they were derived) had to have the same lead character, a period (.). The result, from a filing standpoint, is illustrated in the example below:

TWAIN MARK
  .LETTERS FROM THE EARTH <1962>
  .LIFE ON THE MISSISSIPPI
    .MARK TWAINS MISSISSIPPI <1974>
  .MAN THAT CORRUPTED HADLEYBURG <1985>

The titles "LETTERS FROM THE EARTH" and "MAN THAT CORRUPTED HADLEYBURG" are piece titles that have been added as qualifying headings to the name heading "TWAIN MARK." The title "LIFE ON THE MISSISSIPPI" occurs as a subheading segment derived from the title portion of the name/title heading "TWAIN MARK/LIFE ON THE MISSISSIPPI." This segment is followed by a qualifying heading carrying the piece title "MARK TWAINS MISSISSIPPI."

The convention of using a space and a period ahead of each qualifying heading or subheading segment, while it solved one sorting problem, created another. In a subject search, works about an author's work interfile with works about the author, as in the following example:

DICKENS CHARLES 1812 1870
  .AS THEY SAW HIM <1970>
  .BLEAK HOUSE
    .CHARLES DICKENS BLEAK HOUSE <1974>
  .BLEAK HOUSE—CONGRESSES
    .HOME SWEET HOME OR BLEAK HOUSE <1985>
  .CARLYLE AND DICKENS <1972>
  .CHARLES DICKENS <1983>

The two entries for "BLEAK HOUSE" are works about Dickens' work *Bleak House*. The other entries are about Dickens himself.

References as a Requirement

Provision of references was considered mandatory by all involved in the design process; however, no one was sure how such a feature should be implemented. After examining the alternatives, it was decided that the most practical way was, for simple references, to incorporate the full reference (both the reference heading and the referenced heading) directly into the index record. Simple *see* and *see also* references could be produced by pairing a 4xx or 5xx authority record tracing field with the 1xx heading in an authority record. This technique would not work for complex references. For those it was decided to include in the index entry only the established or reference heading from the 1xx field. This heading is followed by an "empty" referenced heading segment, which consists solely of a lead character that identifies the reference instruction phrase. In the example:

ZIMBABWE
  *For information on this heading type 1.

the referenced heading segment consists solely of the code 0, which on display is used to generate the appropriate reference instruction phrase.

When the user types the line number, the system generates a reference display from the established heading or reference heading, using the 1xx field and the appropriate complex note field(s) (260, 360, 663, 664, 665, 666, or 680). In the Zimbabwe example above, the authority record contains a 665 (information/history note) field, which is displayed as follows:

ZIMBABWE
  Southern Rhodesia, Northern Rhodesia and Nyasaland were united in 1953 to form. . .

Role of the Indexes In Authority Control

Catalog department staff attached high priority to having an efficient and economical means of achieving authority control via the new indexes. They were particularly anxious to be able to visually identify split files, to catch headings that conflicted with *see* references, and to catch references that were blind.

To facilitate visual as well as machine detection of conflicts and blind references in the union catalog, Northwestern added three new "locally defined" heading use codes to the MARC heading use codes in the authority record (bytes 14, 15, and 16 of the 008 field). A code *c* in any of the three fields in an established heading record means that the heading in the 1xx field has been used in a bibliographic record in the indicated way (as a main/added entry, subject, or series). In an index record generated from a tracing field, unless the reference is overridden by a tracing use restriction code in the $w subfield, a *c* "subject" heading use code means that since the *referenced* heading (the heading derived from the 1xx field) has been used as a subject, the reference should consequently be displayed in response to a search.

This use code, modified by the tracing use restriction code where appropriate, is included in the staff display of an authority-derived index record, enabling the searcher to spot blind references without searching on the referenced heading. If the established heading has *not* been used but is appropriate for use, the cataloger can also determine this from a display of the index record.

In certain staff views of the index, each entry is coded as to whether or not it displays in public mode. This enables catalogers and other staff to include in the NOTIS database "resource" records (for internal use only). This can be used for both bibliographic and authority records. The index display also indicates whether the source record contains full cataloging or if the bibliographic data is provisional. The processing unit responsible for the maintenance of the record can also be identified.

Because headings from both authority and bibliographic records are displayed "in context," staff time is saved since it is often unnecessary for the full MARC authority record to be displayed. Also, because there is no delay between the time a record is stored and the indexes are updated, the cataloger or searcher can immediately see the results of adding, deleting, or changing a heading in an authority or bibliographic record.

Although the emphasis is on preventing integrity problems in the database by spotting them at the point of record creation/update, some conflicts, blind references, and interrecord inconsistencies inevitably escape detection by staff involved in searching, cataloging, or data entry. Programs have been written which periodically scan the index file and report these problems. Many of the "flags" carried in the fixed control data field were defined specifically to allow interrecord inconsistencies to be identified. For example, the system requires that the heading derived from the 1xx field in a reference record coded as "traced" match a heading derived from a 4xx field in a record for an established heading. Or, if a 5xx tracing is coded "make a reference," that heading must match a 1xx heading in another record that is coded as "used."

Also available are programs that use two copies of the index file, created at difference points in time, to identify headings that are "new" to the database or have been "dropped" from it. At Northwestern, such new/dropped headings lists are limited to subject headings; however, the same programs can be used to produce lists of new or dropped name or series headings.

It should be emphasized that the indexes do not operate to "control" headings. Northwestern catalogers made a conscious decision that they did not want to be forced to create an authority record for every used heading. It is possible, however, by a batch scan of the index file, to produce lists of headings used in bibliographic records which are *not* supported by authority records. These lists, available on demand, may be limited to a particular category of heading (e.g, subjects, series, conference names, and corporate names), or they may include all categories of headings.

A question that had to be answered when these programs were being written was What constitutes support for a heading? Northwestern catalogers decided that in the case of multiunit corporate and conference names and with subject headings with subject subdivisions a heading in a bibliographic record could be considered "supported" by an authority record even if it contained subfields that were not in the heading of the authority record. Using a topical subject heading as an example, the heading ARCHITECTURE in an authority record would support the heading ARCHITECTURE—AMERICA in a bibliographic record but would not support the heading ARCHITECTURE AMERICAN in a bibliographic record.

At present, unless there is an authority record that includes subject subdivisions, the system does not make any judgment as to the correctness of subject subdivisions in records for which no exactly matching authority record exists. In theory, if and when the Library of Congress distributes authority records for subject subdivisions, it is possible that each subdivision could be checked to see that the subdivision is a valid one (which does not mean it has been correctly applied).

A similar policy applies to corporate bodies with multiple subunits. An index record that contains the established heading INTERNATIONAL BUSINESS MACHINES CORPORATION from a 1xx field in an authority record would support the heading INTERNATIONAL BUSINESS MACHINES CORPORATION DATA PROCESSING DIVISION in a bibliographic record. Of course if there were also an authority record for the heading INTERNATIONAL BUSINESS MACHINES CORPORATION DATA PROCESSING DIVISION, the heading in the bibliographic record would be considered supported by both authority records. This policy enables a decision to make or not make an authority record to be based on whether or not a reference is needed.

One option, which was rejected as unworkable from a work-flow standpoint, was to have the system reject—at the point of record creation/update—an unsupported or conflicting heading. Under consideration as a future enhancement, however, is a VERIFY command that can be used from an already stored record and that will identify any unsupported or conflicting headings.

Collocation Requirements

Because staff wanted to be able to detect split files easily, another requirement for the new indexes was that all index records with a particular heading file together. In order to achieve such collocation, the number of characters allotted to the heading in the new index entries had to be much greater than in the earlier indexes. Again, heading length variability was the problem. Multiunit government body names and multisectioned name/title series headings can sometimes be quite long, with uniqueness often far down in the heading. Investigations into the prevalence of headings that require more than 120 characters to achieve uniqueness turned up only a handful (mostly name/title series issued in multiple sections)—too few to justify increasing the maximum length of the full heading segments beyond 120 characters.

Because of the importance of collocation, it was necessary to include in the first heading almost all the subfields from the source field. In the case of personal names, this includes birth dates and qualifiers; in the case of uniform title headings, it includes numbers, parts, versions, etc.

THE ROLE OF CONTROL DATA

Control data elements selected for inclusion in the index records had to be chosen very carefully in order to minimize the size of the field and maximize the usefulness of data in the field.

This field is divided into fixed-length subfields, some of which contain data codes and some of which carry 1-bit flags. The field has a generic tree structure, with the first 2 bytes of the field (the function code and the heading type code) defining the rest of the field.

All records contain the NOTIS record number and the code of the processing unit responsible for the record. Each record, whether from an authority or bibliographic record, contains a subject system code. In addition a number of "length" subfields are carried in each record.

If derived from an authority record, the kind of record code can be deduced. If derived from a bibliographic record, the bibliographic level code, the type of record code, and the form of reproduction code are carried.

Each record contains 32 as yet unused bit flags that eventually will be used to identify the location-based catalogs to which an entry is appropriate, allowing a user to limit a search to materials in a particular location.

Other flags are used to indicate such conditions as:

- Should the record be displayed in public mode?
- If derived from an authority record, does the source record contain a complex reference note (260, 360, 663, 664, 665, 666, or 680)?
- If from a bibliographic record, is the bibliographic data provisional or full?
- Is the full heading truncated (i.e., did the source field generate more than 120 characters)?
- Is the full heading in the record a "composed" heading (i.e., derived from more than one access field)?
- If the record is derived from a 1xx field in an authority record, is it appropriate for use as a main/added entry? As a subject? As a series? If appropriate for one or more of these uses, has it been used in that way?
- If the record is derived from a tracing field in an authority record, has the referenced heading been used, or is it appropriate for use in one of the three ways?

THE INDEXES IN USE

Designing the Display

Although, for collocation and work identification, the longer the headings the better, this need was counterbalanced by the need to display as many index records as possible on a single screen. Some form of indented list display format was essential, one which would allow for collapsing of segments held in common with adjacent records. Developing a comprehensible but compact display format was quite a challenge. As illustrated in the following example, present logic calls for a line break on the occurrence of a subheading, a referenced heading, or qualifying heading. A subdivision segment that does not follow a subheading also triggers a line break.

With an authority-derived index record, a reference instruction phrase is inserted ahead of the first referenced heading requiring that phrase, as in the following examples:

QUAKERS
*Search also under:
1 SOCIETY OF FRIENDS
2 .STEWARDSHIP OF WEALTH <1985>
—AUSTRALIA—BIOGRAPHY
4 .BACKHOUSE AND WALKER <1981>
—ENGLAND—BIOGRAPHY
5 .APOCALYPSE OF THE WORD <1986>
6 .JAMES NAYLOR 1618 1660 <1982>
7 .MARGARET FELL <1984>
8 .RECORDS AND RECOLLECTIONS OF JAMES JENKINS <1984>

To display the eight records listed above requires eleven lines. Highlighting and color can also be used to distinguish between different types of index entries and different elements within an entry.

Limiting Searches

One advantage of a dictionary index is that it opens the way to a variety of index subsets. These subsets can be customized to fit the needs of various types of users (public, catalogers, acquisitions, or reference staff, etc.).

Index subsets are controlled by a search term qualifier termed "view type," and within the view type by "search type" (author, title, subject, series, name, etc.). Within search type, a "search subtype" code also operates to refine a search, based on subject heading list/thesauri.

The view type is system-controlled in public views, but in staff mode the user can switch between view types. View types serve as filters to enable certain broad categories of index entries to be selected and others bypassed. For example, index records for established headings from authority records which do not contain any complex note fields display in some but not all staff views, whereas they do not display at all in public views. Similarly, reference entries from authority records where the established heading (the referenced heading) has not been used are suppressed or displayed based on view type.

In subject searches, a search "subtype"

code enables a search to be restricted by subject heading list. A user in the medical library, for example, may restrict a subject search to MeSH subject headings.

Staff, particularly catalogers, often want a search that does not distinguish between names, titles, and subject headings. By using a search type or search subtype of *x*, a dictionary search can be performed which ignores the "use" and "type" of the heading. This enables a cataloger, for example, to obtain in a single list of index records works by and about an author or to retrieve a group of index records in which LCSH and MeSH subject headings are intermixed.

One index subset defined as essential for authority control provides a "headings view" in which only one record per heading is retrieved. If a heading from an authority record exists, that record is preferred over the same heading from a bibliographic record; if there is no authority record for a particular heading, then the heading from the first bibliographic record is displayed, becoming, in effect, a "surrogate" authority record. Again, the "level of establishment" can be determined from the index display.

Limiting Searches by Masking

One concern was that, as the size of the index file grew, the simple "left-anchored" string searches which had always characterized NOTIS might result in many searches retrieving very large numbers of records. This was a problem with prolific authors, particularly corporate authors. It also was a problem with many subject headings.

To solve this problem, a technique was developed which allows the user to include a masking symbol in his search term which operates to "mask out" character strings occurring in the middle of an entry. The portion of the term to the left of the masking symbol (the anchoring term) is used to identify a set of candidate index entries. The portion of the term to the right (the floating term) is "floated" through each of the candidate entries to determine if a match can be found. Only if a match occurs on both segments of the search term is the record retrieved. The masking operates against all segments in the entry. In many cases it can greatly reduce the number of characters the user has to include in his search term. For example, the following search term

NORTHWESTERN # DEARBORN

retrieves only entries with the heading

NORTHWESTERN UNIVERSITY EVANSTON IL DEARBORN OBSERVATORY

Limiting by Qualification

A variation on the masking technique is a feature termed "search qualification." This allows certain records in a group of candidate index entries retrieved by the first search term to be bypassed if they do not also match a second search term. For example, the user can limit an author search by title, or limit an author, title, or subject search by date of publication or by a range of dates. Searches on a series heading may be qualified by a volume number. Searchers can also be restricted based on type of record code (language materials, music recordings, maps, films, etc.); form of reproduction code; bibliographic level (monographs, serials, component parts), etc.

The following is an example of a search term that includes both masking and qualification:

t = PROCEEDINGS # CATHOLIC PHILOSOPHICAL & v = 51

This search would retrieve only the following entry:

PROCEEDINGS OF THE AMERICAN CATHOLIC PHILOSOPHICAL ASSOCIATION
.ETHICAL WISDOM EAST AND OR WEST <V 51>

Searches may also be qualified based on the processing unit responsible for the record, providing, in effect, the ability to restrict by broad groups of locations (main library, medical library, law library, transportation library, etc.). Planned for implementation during the coming year is a method whereby searches can be limited by "location-group" within the processing unit. A "location-group" is defined as one or more locations. A location can be defined finely (e.g., a browsing area) or broadly (the stacks of a research library).

Limiting by Catalog

When a user uses qualification or masking to limit a search, the limitation stays in effect only for the duration of the search. Planned for implementation some time during the coming year is an enhancement to provide a method by which the user can, at the start of a session, "set" a particular catalog using a CHOOSE command. The catalog selected will then stay in effect until another catalog is chosen or until the user returns to the introductory screen.

Catalogs that cut across locations can be defined. For example, a catalog could be defined to include films only or serials only. It is planned that catalog selection will enable users not only to select subsets of the index records that apply to a particular database but also to select an entirely different database. For example, users should be able to switch from a database containing conventional cataloging records to one that contains indexing records.

Increasing Recall

Although the ability to limit a search is useful to avoid having to page through many screens of index entries, there are many ways in which recall can be increased without resorting to keyword indexes. One way involves using a single access field to produce multiple index records (explosion). For example, one record can be produced in which all the segments are in the same order as in the access field, and another record can be produced in which the simple heading is derived from the access field but which contains subheading and/or subdivision segments "borrowed from" other access fields in the record (composed heading records).

Another technique that explodes one access field into more than one index record involves creating one record with the segments in the same order as in the access field and another record in which the segments are interchanged (rotated heading records.)

Rotation for Increased Recall

Rotation is used to provide access to titles in name/title headings. In the following example, two index entries are generated from a name/title added entry, one in direct order, searchable by author:

HAWTHORNE NATHANIEL 1804 1864
.BIOGRAPHICAL STORIES FOR CHILDREN
.TRUE STORIES FROM HISTORY AND BIOGRAPHY <1972>

and one in rotated order, searchable by title:

BIOGRAPHICAL STORIES FOR CHILDREN
.HAWTHORNE NATHANIEL
.TRUE STORIES FROM HISTORY AND BIOGRAPHY <1972>

Although rotation is presently limited to name/title headings, this technique has promise for the future, providing access, for example, by subunit in a corporate body, by subject subdivision in a subject heading, or by keyword in a title.

Composed Headings for Increased Recall

In the following example, where the source bibliographic record contains both a main entry field and a uniform title field, two index records are generated. In the standard record, the main entry field is the sole source of the full heading. In the extra index record, the main entry field and the uniform title field are combined to form a name/title heading. In both records, the qualifying heading consists of the piece title.

PLATO
.CRITO
.PLANTONIS CRITO <1903>

PLATO
.PLANTONIS CRITO <1903>

Another type of composed heading which is useful for increasing recall involves creating "name/name" headings. Name/name headings are presently derived only from 700, 710, and 711 added entry fields that do *not* contain a $t or $k subfield and which contain a $4 (relator code) subfield. In such headings the simple heading is derived from the 7xx; a subheading is derived from the 1xx. The following is an example of such a composed heading:

SOLTI GEORG SIR 1912
.BARTOK BELA
.BLUEBEARDS CASTLE <1980> SOUND
.MIRACULOUS MANDARIN <1970> SOUND

An evaluation is presently underway to determine if, in the case of topical subject headings which describe the "form" rather than the actual subject content of the work, an index record should be created with a subject/name heading, constructed from the 650 field and the 1xx field and the 245 field, as in the following example:

PIANO MUSIC
.ORNSTEIN LEO
.A LA CHINOISE

In the example above, the source field should really have been tagged as a 655 (genre/form) field; however the Library of Congress does not presently make this distinction in their subject authority records. Northwestern is considering identifying these headings and changing the tag to 655.

Amplification

Experience with the early indexes indicated that, especially in subject searches, users often bypass accessing the full bibliographic record if the indexes themselves contain location and call number information. Since this is more efficient from a system standpoint, the challenge to the designers was how to incorporate location/call number information in the new index records without greatly increasing record size. The problem was solved by implementing an "amplify" command that can be issued from an index display, which causes location/call number information to be extracted dynamically from the holdings record and displayed following each index entry in an index display, as in the following example:

HALEY ALEX
1 .RACINES <1977>
-> COPY FOR MAIN africana IN PROCESS
2 .ROOTS <1976>
-> SEABURY E185.97 .h168a3 1976
3 .ROOTS <1976>
->MAIN africana 929.20973 H168r
->MAIN core 929.20973 H168r
->DEERING special collections Falley Am H168r

Although a fair amount of system overhead is involved in amplification, it is expected that it will be offset by a reduction in the number of full bibliographic/holdings displays that must be generated. Although the primary beneficiary of this facility is the library patron, catalogers also find it useful in assigning call numbers.

Enriching Access Through Use of References

Probably the ease with which references can now be created is the most significant contribution that the new indexes have made to improved access. No longer is the cataloger deterred from making references by concern over the amount of work involved in maintaining those references, a concern which was always present in the days of the card catalog. Further, because of the "logs" of searches maintained by the system, the cataloger is able to obtain concrete evidence as to what types of references are needed. It is also very easy to see, from the index displays, which references are unneeded, and these can then be coded, in the tracing field, so they will not display in public mode but will still be available for staff use.

INDEX ASSESSMENT TO DATE

The new brief entry index file became available for use by the staff of the Northwestern University Libraries in mid-1987. They were made available in LUIS (the public access mode of NOTIS) in September 1988. Reaction so far has been positive.

By increasing the number of references and access points, by more appropriate references, by rotation, by explosion, by composition, by masking, by qualification, by amplification, and by catalog selection, many of the "recall enhancement" and "search limiting" features of keyword-Boolean search can be provided without the "noise" and "misses" characteristic of keyword searching and with substantial cost savings compared to Boolean searches of keyword indexes. Because the vocabulary is controlled and a syndetic structure maintained, precision is enhanced, and changes in terminology over time can more easily be accommodated.

FUTURE POSSIBILITIES

The full potential of the indexes is far from being realized. There are a number of index record types that could be adapted to the index structure to provide additional access (title keyword, subject subdivision, class numbers, for example).

Because information about the vocabulary is intermixed with citations to works using that vocabulary, an opportunity exists to experiment with dynamic reconciliation between terms from different subject heading lists, thus breaking through the barrier which has traditionally existed between general and special collections. By dynamically "mapping" headings and references derived from multiple subject heading lists, the possibility exists for the creation of a "super" vocabulary that brings closer the day when the online catalog can be truly termed an *expert system*.

The potential also exists for including additional supplementary data in the index, derived from either fixed or variable fields. There are a number of ways in which this could be done without greatly affecting record size, although the file size would be increased considerably. The 245 field could, for example, be exploded to produce multiple index records, each with a different element of supplemental data (edition, publisher, class number, etc.).

One very interesting area for research is to determine if NOTIS-type index records, with their similarities to the relational database model, and searchable using commands such as JOIN could not supplant the full bibliographic records for the majority of users. Most index records from bibliographic records contain, in addition to the full heading, a second heading that itself is always indexed. That second heading in one record, in combination with the bibliographic control number, could be used as a search term to locate the record or records in which it appears as a first heading. The second heading for such a found record could then also be used as a search term. The net effect would be that multiple index records could be assembled into a near complete bibliographic record.

The NOTIS indexing technique also has potential for use when a USMARC authorities format for classification numbers is developed. One index record might be built from an authority record that would contain a class number followed by the textual "caption" associated with that number. Another record could be produced in which the elements are rotated—caption first, followed by the class number. References (including general explanatory references) could connect related numbers. Merging a library's call number indexes with these authority records and using rotation could provide another level of subject access.

Another very intriguing area for investigation is to determine if precoordinated "brief entry" type indexes could be used to achieve a truly national, "distributed" database. It is interesting to speculate if a network of systems could be established in which each node contributed only its indexing records to other nodes in the network. This would enable users at any node to have access to a national union catalog of brief entry and reference records without the expense of duplicating the full MARC records at each node. With a network of interconnecting references, the problem of variations in headings from one library to another could be overcome or at least minimized. For such an approach to work a "standard" index record format would need to be developed.

Northwestern University Library's index redesign project has demonstrated that the data in MARC-formatted bibliographic and authority records can be decomposed into multiple brief entry records, in a format that provides superior access, is simple to manipulate and economical to process, and can be used to monitor the integrity of the database. With this latest enhancement to NOTIS/LUIS, we feel we have made one more giant step toward our goal of a seamless, totally integrated library system.

REFERENCES AND NOTES

1. Normalization, as it works in index entry generation, involves stripping out of an access field all punctuation characters, diacritics, delimiters and subfield codes. Lowercase alphabetic characters are all set to upper. Extra spaces between words are removed. In some cases, spaces left by removing a punctuation character are closed up. Normalization also involves adding certain characters necessary to achieve the desired sort sequence. ■■

METAMARC: An Extension of the MARC Format

Mark Hinnebusch

A hierarchical extension to the ANSI Z39.2 American National Standard for Bibliographic Information Interchange on Magnetic Tape is defined. In the extension multiple records are contained in a single record, which in turn can be combined with other records to make up a single record. How this record structure is accommodated in the existing standard is shown. Then, modifications to the existing standard are proposed, and the hierarchical structure is defined as an implementation of the proposed modified Z39.2 standard. Finally, applications of the hierarchical format are presented. This paper is presented as a starting point for discussion toward a goal of promulgating standards changes.

The American National Standards Institute (ANSI) promulgated the standard structure for data records used for the exchange of bibliographic data known as Z39.2.[1] This standard is the base structure upon which the MARC record standard is built; that is, the MARC record is a particular implementation of the Z39.2 standard. In addition to the structure defined by the Z39.2 standard, MARC further defines the content of the bibliographic record by assigning meaning to data fields, specifying standard field identifiers (tags), fixed portions of fields (indicators), and the content of variable fields.[2]

MARC has been a widely accepted standard for bibliographic data since its development in the late sixties. Now, twenty years later, the experience of the library community has shown the MARC format to be both comprehensive and usable.[3] MARC-based library automation systems have become the rule rather than the exception.[4] Algorithms for handling data in the MARC structure have become widespread and readily adaptable to new systems.

In many instances, it is desirable to connect logically bibliographic records. The MARC format has attempted to deal with this requirement through the use of control numbers in 0xx fields and defined linking fields, such as the 400–490 and 800–840 fields for series notes and tracings, the 505 field for formatted contents notes, and 7xx fields for analytic added entries. In 1980 MARBI Proposal 80-5.1 attempted to deal with the problem by using the implementation-defined portion of the directory entry to indicate a "subrecord" to which each particular field in the record belonged and defining a "relationship code" in a new field 002.[5] In 1981 MARBI Proposal 81-13 provided new bibliographic level values, defined position 19 of the MARC leader as a "linked-record code," defined a subfield 7 for all linking entries and defined field 773, "host item entry."[6] The MARC Format for Holdings and Locations (MFHL) uses the 004 field to identify the parent bibliographic record and the 014

Mark Hinnebusch is Assistant Director for Technical Services, Florida Center for Library Automation, Gainesville, Florida.

to identify network linkages.[7] OCLC has defined the 019 field to link duplicate records that have been deleted.[8] The MULVER committee has recommended to MARBI a linking field solution to the problem of multiple versions of a single work.[9] All of these techniques excepting the subrecord technique of Proposal 80-5.1 depend on the contents of a record to define the link between that record and other records. Moreover, these links are only defined between two records. From a processing standpoint, these techniques are unreliable and difficult to handle.

A hierarchical data structure represents data in a manner analogous to a tree, consisting of a "root" containing data which occurs once and only once and branches which contain multiple occurring data. In general, a root can support any number of branches, and each branch can support any number of branches. There is no theoretical limit on the amount of branching and levels of branching within the structure. The degenerate case is the "flat" structure containing only the root.[10] Two well-known software products supporting hierarchical structures are IBM's Information Management System[11] and Informatics' MARK IV and MARK V languages.[12] A representation of a hierarchical data structure is shown in figure 1.

A hierarchical record structure carries the relationships between its component records (or branches) inherent within the structure of the record. It is capable of handling vertical and horizontal relationships, as well as chronological relationships, which are a special case of a horizontal model. It can define a potentially infinite number of both horizontal and vertical relationships within one hierarchy.

There are many potential uses of this format. It would be useful in any situation where multiple records are to be combined into a single record. The records to be combined could be MARC but need not be. There is theoretically no limit to the number of records that can be combined into a single composite record. Multiple composite records could be combined into a single record, which in turn could be combined with other records, and so on.

MARC was originally designed for information interchange between computer systems.[13] This assumed usage predicated certain design decisions that made sense in the interchange application but have caused problems in the use of MARC records in local systems. The lack of a unique record identifier and a creation date and time in fixed locations in the record (indeed anywhere in the record prior to 1979) has made the matching of records in local systems difficult and sometimes impossible. In an interchange environment, these data were unnecessary and even problematic; MARC-based systems require them for internal processing.

THE METAMARC RECORD FORMAT AS AN IMPLEMENTATION OF THE Z39.2 STANDARD

A hierarchical data structure is a structure in which a record can be logically contained in another record and can itself logically contain a record. Using a tree analogy, the entire tree is a record and each branch is also a record. Traditional data processing has recognized the importance of hierarchical structures for many years, especially in those applications where recursive algorithms are required.[14] The library community, on the other hand, has historically limited itself to the processing of nonhierarchical (i.e., flat) records. The most common record structure in use by libraries is the MARC format, an implementation of the ANSI Z39.2 standard, the American National Standard for Bibliographic Information Interchange on Magnetic Tape. The MARC structure is a flat structure. However, the Z39.2 standard allows for a hierarchical structure, and that possibility promises to solve many important problems in library automation, some of which are dealt with in a later section.

A hierarchical MARC record is a record which meets the definition of a hierarchical record, i.e., it can contain and be contained by other records, and also meets the demands of the MARC record format. Such a hierarchical MARC (henceforth known as METAMARC) record can be defined as an implementation of the ANSI Z39.2 record structure. This implementation allows for the recursive processing of hierarchical

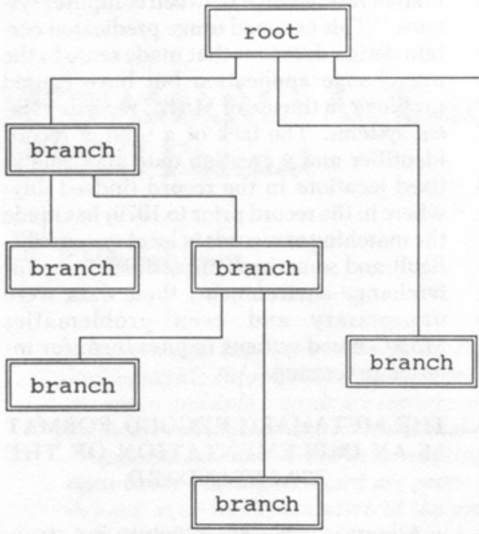


Fig. 1. Hierarchical Data Structure.

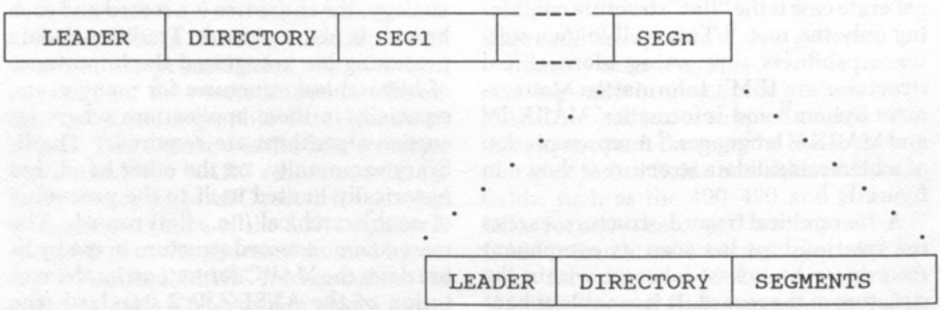


Fig. 2. Hierarchical Record Structure.

records using existent MARC record handling software. The structure of such a hierarchical record is shown in figure 2.

In a METAMARC record, a segment can be a field or a record. A segment can be another METAMARC record, a standard MARC record, or a non-MARC record. By generalizing the definition of a segment to be either a record or a field as defined by the MARC format, the METAMARC record is a generalization of the MARC record, and a MARC record is a special case of the more general METAMARC record.

Since a METAMARC record can be contained within a METAMARC record, the level of nesting is potentially infinite and limited only by the length restrictions imposed for each record at each level. These length restrictions are inherent in the Z39.2 standard, which allows only five characters to define the record length; hence no record can exceed 99,999 characters.[15]

The nesting of records within records is important from a processing standpoint. Nonrecursive software designed to manipulate flat records can be modified to allow for recursive use only if, at each nesting level, the structure of any given record is exactly the same as that of a record at any other level, excepting the lowest level, which may differ in a manner that identifies it as the lowest level. Recursive pro-

Table 1. Leader Contents of Proposed Hierarchical Record for Bibliographic Information Interchange on Magnetic Tape

Position	Contents
0–4	Record Length as a five-digit character field
5	Status as defined in MFBD
6	Type of record If a MARC record, as defined in MFBD If a METAMARC record, blank
7	Bibliographic level If a MARC record, as defined in MFBD If a METAMARC record, blank
8	Hierarchical Record type If a MARC record, blank If a METAMARC record, values are 1—the record consists of a single segment that is a MARC record. 2—the record consists of multiple segments, all of which are MARC records. 3—the record consists of a single segment that is a non-MARC record. 4—the record consists of multiple segments, all of which are non-MARC records. 5—the record consists of multiple segments, some of which are MARC records, some non-MARC; all segments are flat. 6—the record consists of multiple segments, all of which are hierarchical and consist only of MARC records. 7—the record consists of multiple segments, some hierarchical and some flat.
9	Undefined, blank
10	Indicator count = 2
11	Subfield code count = 2
12–16	Base Address of data, a five-digit character field containing the displacement of the byte following the directory. For a MARC record, this is the displacement of the 001 field. For a METAMARC record, this is the displacement of the first segment.
17	Encoding level If a MARC record, as defined in MFBD If a METAMARC record, coded as z, not applicable.
18	Descriptive cataloging form If a MARC record, as defined in MFBD. If a METAMARC record, coded as z, not applicable.
19	Linked record code as defined in MFBD.
20–23	Directory entry map as defined in MFBD.

grams call themselves, passing addresses and parameters. When they return, results are passed back to the caller, which is usually another occurrence of the same program.[16] Schematically a recursive program to process a METAMARC record looks like figure 3. The program consists of three parts: (1) an outer shell which reads a record and starts the recursive process, (2) the recursive algorithm which decomposes the hierarchical record, and (3) the contents processor which performs the needed work using the lowest level (i.e., nonhierarchical) segments of the record. The contents of the leader of a hierarchical record can be designed to map to MARC record leader fields in such a way that they can easily be distinguished and processed by similar software. A METAMARC implementation of Z39.2 would differ from MARC only in the definition of position eight in the leader (the third position of the legend), which is undefined in MARC. The leader fields as they could be defined for MARC and METAMARC record formats are shown in table 1.

PROCESSING MARC AND METAMARC RECORDS

The structural equivalency of MARC and METAMARC records assures that a great deal of the extant MARC processing software could be relatively easily modified to handle METAMARC (see figures 3 and 4). The outer program segment which in-

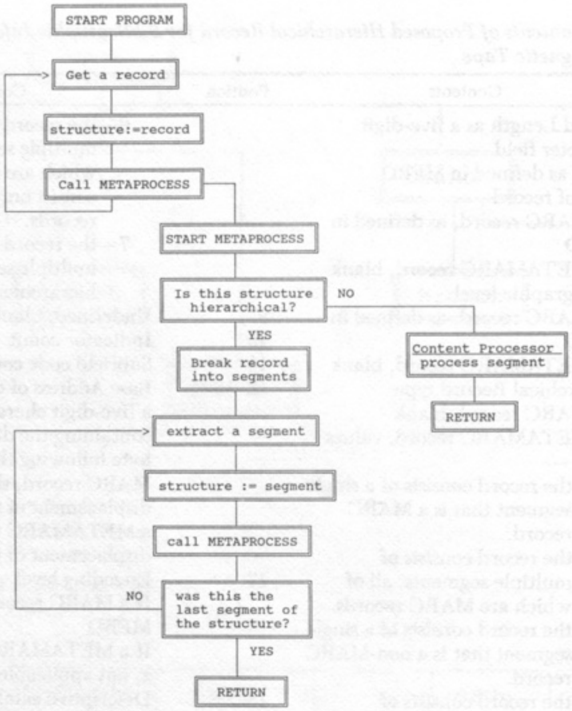


Fig. 3. METAMARC Record Processor.

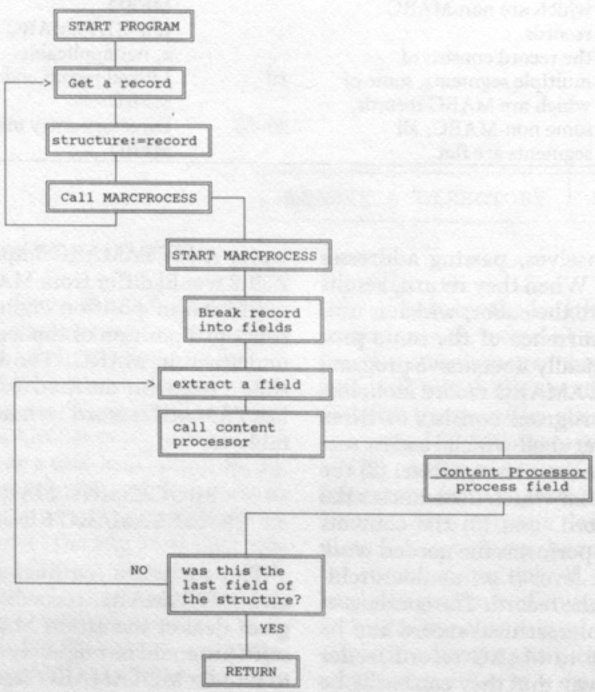


Fig. 4. MARC Record Processor.

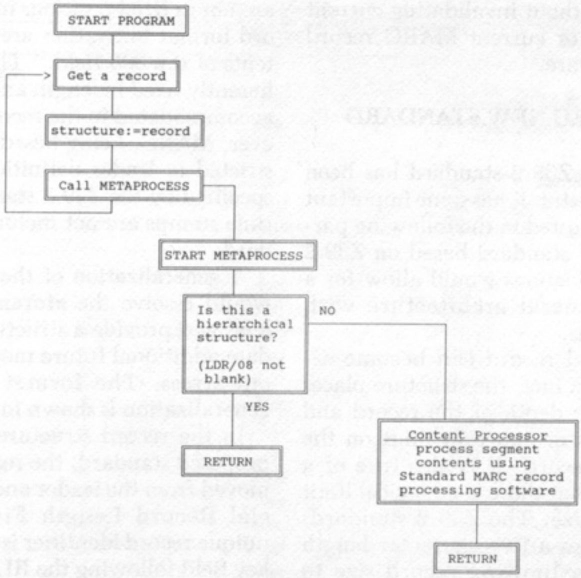


Fig. 5. Handling only Flat MARC Records.

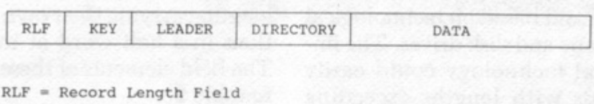


Fig. 6. Generalization of Z39.2 Structure.

puts the record differs only in the name of the module called to break a record into component parts. The content processor in the METAMARC case is simply another instance of the MARC record processor. The process of breaking a METAMARC record into segments is exactly the same as the process of breaking MARC records into fields. The test for the last segment in a METAMARC record is the same as the test for the last field in a MARC record. Clearly, very little new software is required to move from a MARC to a METAMARC processor, at least at the logical level; some implementations may be more difficult.

A minimal modification to existing software which would allow mixing MARC and METAMARC records on input but processing only the MARC records takes advantage of the leader structure equivalency. The standard MARC record format is a special case of the more general METAMARC format and could be processed as such. Similarly, a MARC record leader could be interpreted as a special case METAMARC leader. Processing software would recognize the record as being a single MARC record based on the blank in LDR/08 and process the record accordingly. In terms of the general logic flow shown in figure 3, a MARC record would not be treated as a hierarchical structure and would never enter the recursive process. The METAPROCESS part of the program in figure 3 becomes a simple check and pass-through, as shown in figure 5. If it were known that only flat MARC records would be encountered, no code changes would be required to existing software; hence the METAMARC format could be promulgated without invalidating current MARC records or current MARC record processing software.

A PROPOSED NEW STANDARD

Although the Z39.2 standard has been extremely successful, it has some important limitations, discussed in the following paragraphs. A new standard based on Z39.2 but with modifications would allow for a much more general architecture with fewer limitations.

A hierarchical record can become extremely large. In fact, the structure places no limits on the depth of the record and therefore places no inherent limit on the length of the record. It is also true of a MARC record that there is no logical limit to the record size. The Z39.2 standard, however, defines a five-character length field, explicitly limiting record size to 99,999 characters. While it is probable that no MARC record would exceed this limit, it is possible that a hierarchical record could. The specification of five characters is most likely historical and based on technological limitations of tape and disk drives. The impending optical technology could easily support records with lengths exceeding 99,999.

Practically speaking, online systems can only be designed using random access techniques. Random access generally requires at least one unique record identifier, placed in a fixed location in each record in the file and having the same length in each record in the file. As a standard for data interchange, Z39.2 does not specifically relate to online system record structure. The MARC record structure has become such a standard, however, that many online systems use the structure in their internal design. The lack of a unique record identification field, i.e., a key, in the MARC format causes additional processing at best and an inability to match records for update or replacement at worst. While keys normally must meet the fixed position and length criteria in any given file, there are no inherent length or position restrictions.

The processing of records is often controlled by the date and time of the record creation or last modification. These data are not in fixed positions in the MARC record format but rather are defined as contents of the 005 field.[17] These data are inherently fixed in length and could easily be accommodated in the record leader. However, MARC, being based on Z39.2, is restricted in leader definition to those data specified by the Z39.2 standard. Date and time stamps are not included in that standard.

A generalization of the Z39.2 standard would resolve the aforementioned problems and provide a structure to accommodate additional future modifications without stress. The format of a proposed generalization is shown in figure 6.

In the record structure defined by the proposed standard, the record length is removed from the leader and placed in a special Record Length Field (RLF); the unique record identifier is defined in a new key field following the RLF and preceding the leader; and the leader itself is generalized to allow for both predefined fields and implementation-defined fields. The RLF, the key, and the leader are all of variable length, carrying their own length specifications in a half-word at the front of each. The field elements of these fields are shown in table 2.

The Record Length Field is the portion of the record which specifies the length of the record. In the Z39.2 format, the five-character field at the beginning of the record leader serves this purpose. In the proposed standard, the RLF has been generalized as a variable length field prefixed by a 2-byte subfield specifying the total RLF (including prefix) length. This generalization allows for records of essentially any length, since the RLF itself could theoretically provide for a record length field that is 97 characters long.

The key is a unique record identifier. Z39.2 does not specify a key per se. Sometimes a MARC record will have a unique identifier in the 001 field. Often, the first step in processing a MARC record is to extract this datum and move it to a key field at the beginning of the record. By placing the key at the head of the record, the proposed standard more nearly approaches traditional data processing record structures. In the common case of a file in which all rec-

Table 2. Field Elements of the Proposed Standard Record Format

Element	Length	Type	Contents
Record Length Field			
RLFLEN	2	char	Length of RLF
RLFRLEN	V*	char	Record length
Key Field			
KEYLEN	2	char	Length of key field
KEYTYPE	2	char	Record type
KEYKEY	V*	binary	Record key, i.e., unique identifier
Leader			
LDR/00	5	char	Leader length
LDR/05	1	char	Record status
LDR/06	1	char	Record type
LDR/07	1	char	Bibliographic level
LDR/08	1	char	Hierarchical record type
LDR/09	1	char	Undefined
LDR/10	1	char	Indicator count
LDR/11	1	char	Subfield code count
LDR/12	5	char	Base address of data
LDR/17	1	char	Encoding level
LDR/18	1	char	Descriptive cataloging form
LDR/19	1	char	Linked record code
LDR/20	4	char	Directory entry map
LDR/24	8	char	Creation date (YYYYMMDD)
LDR/32	6	char	Creation time (HHMMSS)
LDR/38	10	char	Source identifier
LDR/48	4	char	Potential segment count
LDR/52	4	char	Actual segment count

*V = variable length

ords have the same RLF length and key length the key becomes a fixed location datum, easily used directly in online systems and in sorting processes.

The key field is not a fixed length field but rather consists of a two character key length element, a key type identifier element, and the variable length key. The key type would be implementation-defined and could be used to identify particular record types in a file containing multiple record types; each type could have a different key length since the key type would allow separation at an early processing step. A record that has no key would have a key field with a KEYLEN of '04', a blank KEYTYPE, and no KEYKEY element.

The leader in a Z39.2 record is the fixed length portion of the record. In the proposed structure, the leader would be of variable length, with a five character length prefix. In addition, the leader could contain both predefined and implementation-defined elements, with the predefined elements at the beginning of the leader and the potentially variable length implementation-defined elements following. This structure allows the implementor to place commonly used data in the leader, which can be processed more easily than data in variable fields.

The first part of the predefined elements is a direct reproduction of the Z39.2 leader with the hierarchical extension as defined in table 1, except that the first five characters are the length of the leader, not of the record. This mapping of the new leader to the Z39.2 leader would minimize modification of existing software. The rest of the predefined leader elements are new. The LDR/38 element allows for an implementation-defined identification of the origin of the record. NUC codes, OCLC profile codes, internal record numbers are all possible candidates for this element. Two segment counts are defined: LDR/48 contains the maximum number of segments it is possible for this type of record to contain and is implementation set; LDR/52 is the number of segments actually in the record. The variable length of the leader allows for implementation-defined elements to follow the predefined elements. The standard would make no assumption about such elements.

The directory would remain unchanged in the proposed structure. A directory entry would define a segment which may be a field or a record. If the record is a MARC record, i.e., LDR/08 is blank, then each directory entry would define a field, and the MARC standards for field tags would apply. If the record is hierarchical, i.e., LDR/08 is nonblank, then each directory entry would define a segment, and implementation-defined segment tag definitions would apply. As Z39.2 does not define valid field tags, the proposed standard would not define segment tags. Further study and analysis of applications could lead to standardized segment tags separate from this proposed standard, as the MARC tag definitions are separate from Z39.2.

APPLICATIONS OF THE METAMARC RECORD

The METAMARC record structure offers a solution to a number of problems that have arisen in library applications. The remainder of this paper will discuss a number of these problems and potential solutions using METAMARC. By no means should this be considered an exhaustive account. Some of the examples are of more immediate use; some are less so. Each example stands alone; I have made no attempt to make the examples consistent with each other.

A considerable number of problems amenable to the METAMARC solution are problems of connecting various types of records for transmission. While the METAMARC structure can meet many record connection needs using only a single level of hierarchy, the full power of the structure becomes apparent when considering possible applications of a multilevel record.

Monographic Series

Identifying the titles in a monographic series has been a concern for years.[18] Current cataloging practice uses the MARC 800-840 fields to attempt to solve this problem. This is a one-way solution, however, in that each title in the series identifies the series but there is no mechanism for the series record to identify all of the titles in the series. A METAMARC solution would be to physically connect the series record and the title records in a single METAMARC record. The hierarchical record would consist of a segment containing a record identifying the series followed by segments containing records identifying the titles in the series. The METAMARC record would be as shown in figure 7. Not only would this solution provide an exact definition of the relationship of the series record to the title records, it also could provide additional information if the title record segments are placed in the METAMARC record in a particular order, such as by publication date or volume number.

Parts

Identifying the component parts of an item is another problem encountered in cataloging. The 505 field containing contents notes has provided a partial solution.[19] There is, however, no way for the part to identify the host item and a component part often has no record defining it. A METAMARC solution to this problem is similar to that for the series problem in that the hierarchical record would be composed of a host item record in the first segment, followed by segments made up of records defining the component parts. The METAMARC record would look like that in figure 8. The order of the component part record segments could impute significant information.

Articles within Serials

The analysis of serial periodicals has been a major concern.[20] The METAMARC structure provides an elegant solution to this problem. The record would in a very real sense mirror the periodical issue. The METAMARC record would consist of a segment containing a record defining the issue of the periodical followed by segments containing records identifying the articles contained in the issue. The hierarchical record would look like that in figure 9. The order in which the article record segments appear in the hierarchical record could be alphabetic by main entry, by order of appearance in the issue, or some other order appropriate to the particular application.

Successive Entry Cataloging

Identification of a single publication throughout a life often involving multiple name changes is a major cataloging concern. This successive entry cataloging is a problem begging for the METAMARC solution. The METAMARC record would consist of the records for the publication under each of its various names. These records could be ordered in the METAMARC record either in chronological or inverse chronological order. Note that that there are neither "parent" nor "children" segments, as in prior examples; rather all segments are at the same level.

Authority Control

Authority control could be a major application of the METAMARC record structure. The connection of bibliographic and authority data could be performed in two ways. A METAMARC record could consist of a bibliographic record followed by the authority records for the controlled terms appearing in it. Alternately, the METAMARC record could consist of an authority record followed by the bibliographic records containing the controlled term. Which arrangement is chosen would depend on the particular purpose at hand. Either arrangement would have its uses and both are valid METAMARC structures.

Holdings Linkage

The MARC Format for Holdings and Locations (MFHL) provides the 004 field to connect the holdings record to the bibliographic record. This is a one-way linkage and depends on the consistent use of the bibliographic 001 and holdings 004 fields. The METAMARC structure provides a method of connecting bibliographic and holdings data that is intuitively pleasing. A METAMARC record containing a bibliographic record and the associated holdings records would look like figure 10. The definition of a METAMARC record does not preclude the mixing of MARC and non-MARC records in a single METAMARC record, so a system which does not support MARC holdings could still produce a METAMARC record of this form. Of course, the receiving system would have to know how to interpret the non-MARC holdings records, but the extraction of those records from the encapsulating METAMARC record would be exactly the same as for a METAMARC record containing only MARC records. This holds true even if the METAMARC record consisted of both MARC and non-MARC holdings records attached to a single bibliographic record.

Acquisitions Applications

Potential uses of METAMARC in the ac-

Fig. 7. Hierarchical Record Defining a Monographic Series.

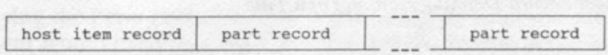


Fig. 8. Hierarchical Record Defining Component Parts.

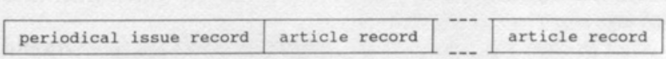


Fig. 9. Hierarchical Record Defining Articles in a Serial Issue.

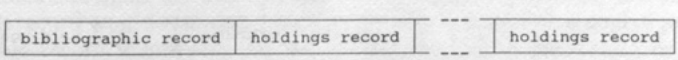


Fig. 10. Hierarchical Record Defining Holdings of a Series.

Fig. 11. BISAC/SISAC as a Hierarchical Record.

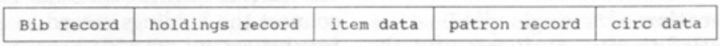


Fig. 12. Hierarchical Circulation Trace Record.

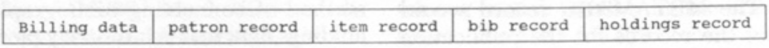


Fig. 13. Hierarchical Circulation Billing Record.

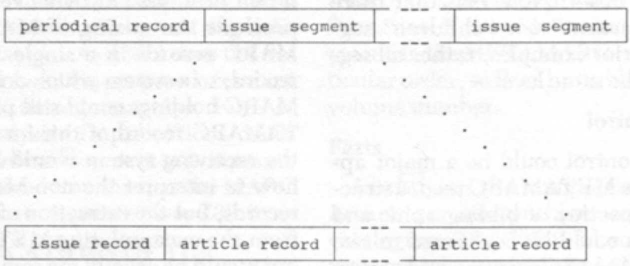


Fig. 14. Hierarchical Record Defining a Periodical.

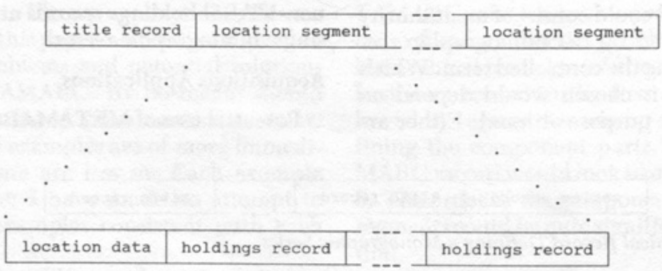


Fig. 15. Hierarchical Record Defining Holdings of a Title.

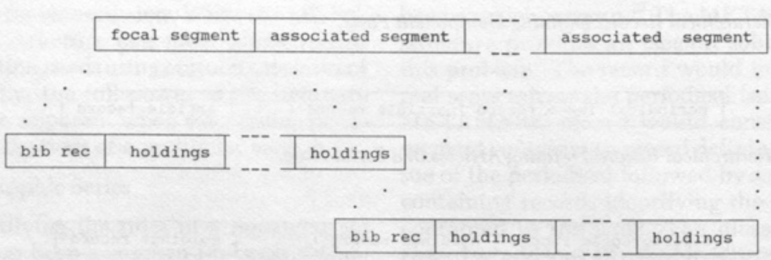


Fig. 16. Hierarchical Record for Multiple Versions.

quisitions area are numerous. The BISAC and SISAC standards use multiple records prefixed with a leader record and followed by a trailer record.[21] The METAMARC record structure is a perfect mechanism for communicating BISAC and SISAC information. The record would look like that in figure 11, with a segment containing the BISAC/SISAC leader record followed by segments containing BISAC/SISAC field level records followed by a segment containing the BISAC/SISAC trailer record.

Circulation Applications

Circulation consists, at its simplest level, of connecting an item with a patron under certain conditions. A circulation trace could be implemented using a METAMARC record containing a bibliographic record, a holdings record, an item record, a patron record, and a record containing circulation particulars, as shown in figure 12. Billing could be handled using a METAMARC record containing billing data, a patron record, bibliographic holdings, and item records, as shown in figure 13, thereby producing a complete snapshot of the circulation transaction for audit purposes.

Multilevel (Hierarchical) Applications of METAMARC

Hierarchical applications arise from the integration of the solutions of the previous section. For example, a METAMARC record could contain records for all of the articles in the issues of a periodical. Such a record would look like figure 14. The first segment of the record would be a MARC record identifying the periodical. Each remaining segment would itself be a METAMARC record consisting of a MARC record identifying the periodical issue followed by MARC records identifying the articles in the issue.

Another hierarchical application would be the collection of holdings in multiple locations for a title. The METAMARC record would look like figure 15. The first segment of this record would contain a MARC bibliographic record identifying the title. Each of the additional segments would contain location identification information followed by records defining holdings. These holdings records could be MARC or non-MARC. Note that this structure incorporates a different view of the relationship between MARC bibliographic and MARC holdings records than does the non-hierarchical solution shown in figure 9.

The problem of multiple versions of a work, now under study by the MULVER committee, could be elegantly solved using METAMARC. The first segment would identify the focal bibliographic record and its associated holdings records and would be followed by segments identifying associated bibliographic records and their holdings records. Such a record would have the structure shown in figure 16.

The transmission of order data could be performed with a METAMARC record containing an order record and the associated invoice records and perhaps also including the bibliographic and holdings records, as shown in figure 17. An alternative implementation of an audit trace could use a METAMARC record containing a single invoice record and the associated order records, again perhaps with bibliographic and holdings records, as shown in figure 18.

CONCLUSION

A straightforward extension of the existing Z39.2 standard would provide a structure which would meet many of the needs of the library community. Using a hierarchical record structure creates a situation in which current MARC record processing software could be used to process records in the new structure with a relatively simple addition of a recursive driver. This extension could be achieved simply by defining the eighth position of the leader, which under Z39.2 is undefined.

Further modifications of the Z39.2 standard would generalize the record structure to provide an environment for the transfer of virtually any type of data. This modification would define a record length field, a record key field, and an expanded leader. This paper is presented as a starting point for discussions with a goal of promulgating the changes to the standards herein outlined.

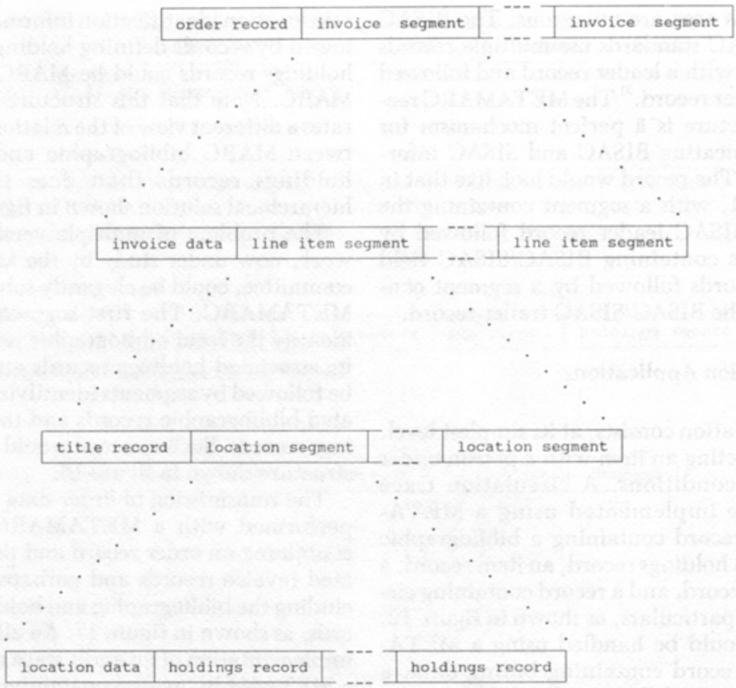


Fig. 17. Hierarchical Record for Ordering Materials.

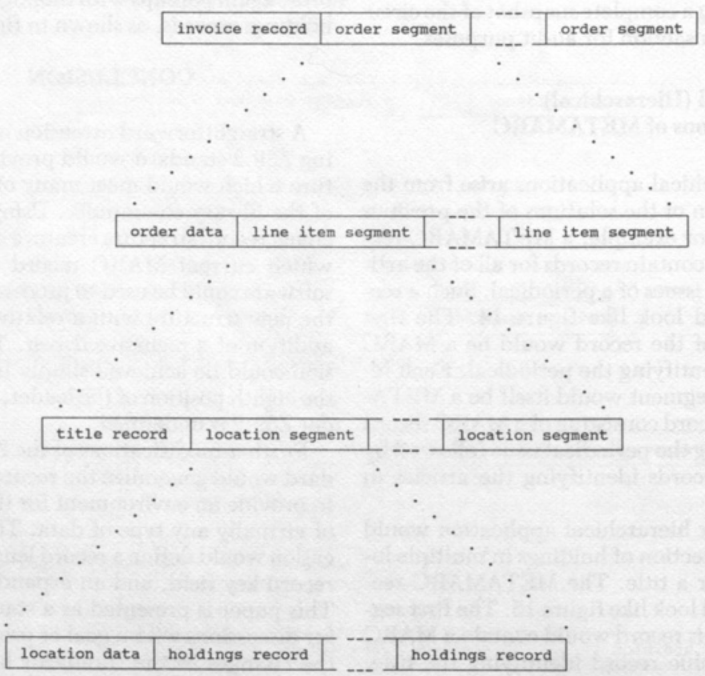


Fig. 18. Alternate Hierarchical Record for Ordering Materials.

REFERENCES

1. American National Standards Institute, *American National Standard Format for Bibliographic Information Interchange on Magnetic Tape* (New York: ANSI, 1985).
2. Library of Congress, Automated Systems Office, *MARC Formats for Bibliographic Data* (Washington, D.C.: LC 1980).
3. Salmon, Stephen R., *Library Automation Systems* (New York: Dekker, 1975).
4. Matthews, Joseph R., *Directory of Automated Library Systems* (New York: Neal-Schuman, 1985).
5. Library of Congress, *Proposal 80-5.1, MARC Subrecord Technique* (Washington, D.C.: LC, 1980).
6. McCallum, Sally H., "MARC Record-Linking Technique," *Information Technology and Libraries* 1, no.3:281–91 (Sept. 1982).
7. Library of Congress, Network Development and MARC Standards Office, *USMARC Format for Holdings and Locations* (Washington, D.C.: LC, 1984).
8. OCLC Online Computer Library Center, *Online Systems Book Format* (Dublin, Ohio: OCLC, 1984).
9. Library of Congress, *Discussion Paper No. 21: Communication of Records for Multiple Versions* (Washington, D.C.: LC, 1988).
10. Date, C. J., *An Introduction to Database Systems, Third Edition* (Reading, Mass.: Addison-Wesley, 1982).
11. IBM Corporation, *IMS/VS General Information Manual* (Mechanicsburg, Pa.: IBM, 1986).
12. Informatics General Corporation, *MARC V Concepts and Facilities Manual* (Canoga Park, Calif.: Informatics General, 1983).
13. Avram, Henriette D., *The MARC Pilot Project* (Washington, D.C.: LC, 1968).
14. Hansen, Per Brinch, *Operating Systems Principles* (Englewood Cliffs, N.J.: Prentice-Hall, 1973).
15. American National Standards Institute, *American National Standards Format for Bibliographic Information.*
16. Knuth, Donald E., *The Art of Computer Programming, Volume 1/Fundamental Algorithms* (Reading, Mass.: Addison-Wesley, 1969).
17. Library of Congress, Automated Systems Office, *MARC Formats.*
18. Smith, Lynn S., *A Practical Approach to Serials Cataloging* (Greenwich, Conn.: JAI Pr., 1978).
19. Library of Congress, Automated Systems Office, *MARC Formats.*
20. Wynar, Bohdan S., *Introduction to Cataloging and Classification* (Littleton, Colo.: Libraries Unlimited, 1976).
21. American National Standards Institute, *American National Standard Format for Information Sciences: Computerized Book Ordering* (New York: ANSI, 1985). ■■

Prospective Conversion: Data Transfer between Fossil and New Microcomputer Technologies in Libraries

Janet Vratny-Watts and Edward J. Valauskas

Over the past decade, libraries have been retrospectively converting a variety of files from manual storage to computer-based database management systems (DBMS) on a range of computers from personal to mainframe. Now, many of these libraries find that their systems are dangerously dated. Libraries using dated hardware and software face the task of evaluating their dependency on old DBMS and the usefulness of older computing hardware. Many libraries will face the challenge of prospective conversion over the next decade in light of technological change and demands from end-users. This paper addresses the problem of migrating a library's data from obsolete personal computers to newer models that feature radically different operating systems, providing three case studies that illustrate possible scenarios.

Over the past decade, libraries have been aggressively converting a variety of files from manual storage to computer-based database management systems (DBMS) on a range of computers from personal to mainframe. Retrospective conversion—the conversion of manual files in small- and medium-sized libraries—has focused largely on moving the circulation and cataloging records to personal computers. At the time of the original conversion, many of these libraries purchased state-of-the-art equipment. Now much of that original equipment is dated, and libraries face the prospect of upgrading hardware and software and transferring records to a new computing environment.[1] What are some of the prospects that libraries have faced during this form of data migration?

Retrospective conversion projects demanded a great deal of planning, based on extensive familiarity with the original manual files and the format of the computerized database. These conversion projects were extremely labor intensive, particularly if efforts were made to convert full records, ensuring transportability to future systems and databases. Planning was vital to preserve the integrity of the original records and to take advantage of the resources of the new computing environment.

Prospective conversion—the migration of computerized records from one computer to another to utilize new computer technologies—also requires a great deal of planning. Planning is required, partially because of hidden idiosyncrasies in operating systems and DBMS and inherent differ-

Janet Vratny-Watts is Senior Information Specialist at Apple Computer, Inc., Library 8C, 10381 Bandley Dr., Cupertino, California. Edward J. Valauskas is Assistant Director at Merriam Center Library, Chicago, Illinois.

ences between products. The actual transfer of data is accomplished much faster than the original manual conversion because of the electronic medium in use. But this also increases the risk of transmission error and field dislocation. Through prospective conversion, the value of detailed, full standard conversion increases because it minimizes staff efforts to overcome machine- and software-specific nonstandardized details.

Many libraries will face the challenge of prospective conversion over the next decade in light of technological change and demands from end-users. A changeover is never a complete one, with residual and possibly archaic technology often working in parallel with new and more efficient computing tools.[2] Personnel involved in a transition may find themselves no longer occupied with the usual maintenance of data files. Instead, they find themselves engulfed in the problems of moving data from one machine and operating system to another with minimal transmission loss. This situation implies that little original work is completed on the data, as all efforts are concentrated on coping with hardware and software incompatibilities and merely developing basic skills in managing several technologies at the same time.[3]

Problems associated with prospective conversion are especially notable when working with one of the most rapidly changing computer technologies of all, that of personal computers and workstations. Potential purchasers face an accelerated pace of change, and new models offer a variety of microprocessors operating over a wide range of clock speeds. Both managers and end-users may be discouraged from committing to a given computing solution because they are afraid of finding themselves in pursuit of new technology within six to twelve months. Nevertheless, the pluses of many recent computing advances in the realm of personal computers clearly outweigh one's loyalty to old systems. New hardware systems feature more sophisticated and graphic interfaces, more powerful processors, larger local storage, and more efficient DBMS. These hardware and software developments mean a more interactive role for the ultimate end-user, increasing local data manipulation and processing. New systems permit real time local co-processing, increasing the number of task options for the end-user and minimizing dependency on a remote mainframe.[4] Given the distinct hardware advantages of most personal computer and workstation options introduced in the past year, the ultimate problem is not one of choosing and defending a personal computer option. Instead, the problem is to create a plan for migrating data as effortlessly as possible from the old system to the new one.

This paper addresses the problem of migrating a library's data from obsolete personal computers to newer models that feature radically different operating systems. This situation has been recognized in business as one of the most cost-intensive operations, often outstripping the initial expenditures for new hardware and software.[5] Prospective conversion is complicated by the fact that many older software designers never anticipated data transfer, creating programs with idiosyncrasies that do not lend themselves to easy solutions. Prospective conversion for libraries means (1) enduring these potential costs to overcome hardware and software limitations created by ballooning databases; (2) ultimately improving staff and patron access; and (3) making record management more responsive to staff-initiated than programmer-induced changes.

DATA MIGRATION IN THE MICROCOMPUTER ENVIRONMENT: OVERVIEW

Data transfer involves answering three basic questions. First, is it physically possible to move data resident on one machine to another? In other words, the older system may be so dated that you no longer have and cannot purchase the necessary hardware or software for data transfer. Secondly, if it is possible to move the data to a new computing environment, is it feasible to use the data with the new local operating system and database management applications? It is possible that the transferred file will not be accepted by the new software, mainly due to: different protocols or data formats; unnecessary ASCII characters; or poor data from the original system. Third, if local software tools can be used, do they restrict the use of the original data; that is, are certain forms of data manipulation impossible in the new machine, even though they were possible on the old computer?[6] There may be cases where in choosing more limited database software for your new system, you loose some flexibility and data manipulation features, making your transferred data too complicated for the new environment.

During data conversion, it is important to retain the integrity of the original data as much as possible during the entire transfer process, minimizing the need for extensive re-keying and correction. Transferring data from one program designed for one machine to another program designed for a considerably newer computer can be difficult or easy, depending on the original nature of the database application. Text and data files may occur as machine-language or as ASCII (American National Standard Code for Information Interchange) files.[7] Data stored in programs based on machine-language files can be extremely difficult to translate to ASCII-based files. Conversely, if the original program containing the data stored on the old computer is ASCII-based, file transfer is considerably easier.

ASCII files are standardized with elements of the code numbered 0 through 127; high codes, from 128 through 255, are used by different hardware and software manufacturers for different purposes. For instance, high codes on IBM and Commodore computers equal different graphics characters; on Apple equipment, they report the first 0 through 127 characters.[8]

It is important at this stage to work with a duplicate of the file or a test subset, based on the original. A clear understanding of what form of ASCII code the receiving program accepts or requires is also essential. From working with the test file, the user determines if it will be possible to put the data into that required format.[9] An ASCII file can be created by asking the originating program to "print" the data to disk. However, it will contain extra, unnecessary characters that were meant as printer commands. They might need to be removed before the new software will accept the file. For example, this feature is possible using the print command within most Apple II or Apple III-based software or by using specific system commands within CP/M products. If the original program is sufficiently dated, it may not support this option.

With the data in the proper ASCII format, it is possible to test the physical transfer of files to new hardware. The following are some basic options to link systems for data transfer.[10] First, there are the simple hardware options, such as linking different external disk drives and using operating system commands to facilitate the translation of files from one physical format to another. Second, telecommunications hardware and software can be used to link the serial ports on two different computers, or modems could be used to directly connect via phone lines. Third, an outside, third-party service can be involved to complete the transfer for a fee, usually based on the size of the file and the degree of difficulty in moving the data between systems. Finally, an electronic mail service can act as an intermediary, with one computer sending ASCII files to an account established in a system on a mainframe, and another computer downloading the same set of files. This last form of transfer clearly prohibits the transfer of extremely large or sensitive files (see figure 1).

Of all of these options, the plan using telecommunications software and hardware is probably the easiest for most libraries. It does not involve the purchase of additional equipment, as in hardware transfer, and is the least expensive compared to the use of a third-party or electronic mail option. Most libraries will already own a modem and software for the original older system and will also have similar equipment for the newer computer. Using telecommunications software and hardware provides several options for the migration of files. One straightforward method is to connect both computers with a mutually compatible cable and to allow the telecommunications software to handle the transfer. This routine is usually possible when a common file transfer format or protocol such as XMODEM or YMODEM is available. This method is usually the most difficult to complete successfully, as the ports may vary in connector size or pin configuration. A second method involves using one modem attached to the new machine to call over phone lines to the other computer and modem waiting at the other end. If the older unit has the proper hardware and software to make the connection, text files can be sent and saved on the receiving, newer model in the appropriate application.

Once the data has been transferred, it is necessary to import it into the new database applications program. If it was possibile to manipulate the original data so that information is properly formatted, the file should "fit" properly in the new application using the 'import' or equivalent command. If the only solution was to print an ASCII-based rendition of the original, some manipulation of the data will be required to make the file work in the new program. Some word processing programs such as Microsoft Word will allow you to examine text files for extra tabs (ASCII code 10), carriage returns (ASCII code 13), and other ASCII debris. Editing these unnecessary characters from the text file will prepare it for a clean transfer in a properly formatted state.

Case Studies

Here are three case studies, based on events in two special libraries. These case studies illustrate in more detail some of the situations described above.

Case Study 1

Library: Apple Corporate Library. Apple Computer, Inc., Cupertino, Calif.

Obsolete System:

Hardware: Apple III running CP/M with the use of a special board; 5 MB hard disk.

Software: DataTrek's *DTI* software, which was a customized version of Ashton-Tate's *dBase II*.

Applications being transferred: Online catalog and circulation.

New System:

Hardware: Apple Macintosh

Software: *MLS* (Macintosh Library System) by CASPR

Transfer software:

Apple III: *IMP* (CP/M-based)

Macintosh: *Microphone* by Software Ventures

History: The Apple Library's needs for flexibility had long outgrown its version of software created by DataTrek. The library was one of the few institutions still using an Apple III CP/M-based system with an early DataTrek product. DataTrek had offered many upgrades of their product, but those were only for large customer groups, such as general CP/M & MS-DOS systems. Even though DataTrek was able to give limited support, loyalty to the software was unreasonable without software upgrades. On the hardware front, Apple had discontinued the Apple III.

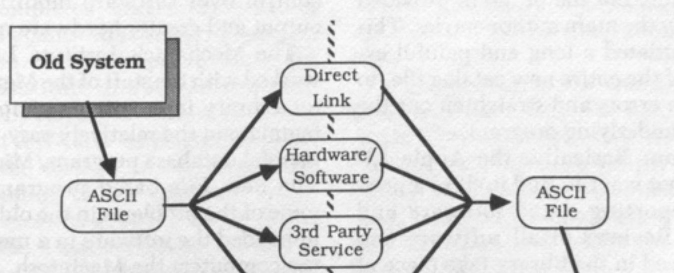


Fig. 1. Data Transfer Options between Old and New Systems.

The library staff felt some alternative would be better, given that no developers were interested in working with literally extinct technologies. The Macintosh had also made its appearance, but the library was not prepared to independently develop a Macintosh-based application incorporating cataloging and circulation functions. Eventually, a developer was located to create a Macintosh-based product, meaning that a data transfer was necessary from the Apple III-based system.

Data transfer occurred in a sequence of three steps. First, the original Apple III-based file was converted to a CP/M-based, tab-delimited ASCII text file with the aid of CP/M commands secured from DataTrek. The second step involved the migration of the data from the Apple III environment to the Macintosh. An Apple expert on data transfer recommended telecommunications products for both computers and customized each of their protocols. A cable was used to connect directly the Apple III and Macintosh. After the transfer, the transferred text files were examined and cleaned of visible errors induced by the migration technique. Finally, the converted text files were delivered to the developer, who worked on importing them into his customized Macintosh applications. The Apple Library eventually received the catalog files from the developer, and an extensive examination of the new database commenced. Some errors appeared in the new developer-created file. Main and secondary author entries were completely reindexed in the new file, but the program provided access to only the main author entries. This discovery initiated a long and painful examination of the entire new catalog file, to clean up the errors and straighten out the bugs in the underlying program.

Conclusion: Basically, the Apple Library does not want to find itself in a position of supporting dated software and hardware. Reviews of all software and hardware used in the library take place at least once a year. Based on the library's experience with both hardware and the developer, an effort will be made to completely understand the structure of a customized program, its inherent weaknesses and strengths, and its interaction with current in-house computer equipment.

Case Study 2

Library: Library, Mechanics' Institute, San Francisco, Calif.

Obsolete System:

Hardware: Digital Microsystems' computer, a CP/M-based product, using 8-inch floppy disks and a hard drive for storage.

Software: DataTrek's *DTI* software, based on Digital Microsystems' CP/M system and 8-inch floppy disk format.

Applications being transferred: Membership records for the Mechanics' Institute.

New System:

Hardware: Apple Macintosh SE equipped with a 20 MB internal hard disk.

Software: Microsoft *File* by Microsoft Corp.

Transfer software:

Digital Microsystems: *ASCOM* by Dynamic Microprocessor Assoc.

Macintosh: *Versaterm Pro* by Abelbeck Software

History: The Mechanics' Institute Library faced similar problems with the DataTrek software over a lack of available product upgrades. In addition, they had problems in obtaining hardware service because the Digital Microsystems' computer was no longer being manufactured. Only a few staff members were familiar with the system, and their knowledge of the intricacies of the software and hardware was far from complete. It seemed best to consider migrating the data off this extinct system, onto one that permitted greater library control over software modification and output and greater hardware reliability.

The Mechanic's Institute Library staff worked with the staff of the Merrriam Center Library in creating a simple database template in the relatively easy-to-use commercial database program, Microsoft *File*. The new *File*-based program corrected some of the problems in the old system and imbedded the software in a much easier to use computer, the Macintosh. The ease of use attracted the staff and administration, given that there would be little time for extensive training. Hardware and software familiarity would literally occur while the software was being used by the staff to create billing statements for the membership of the Mechanics' Institute.

The original file was converted to a CP/M-based,tab-delimited text file, with techniques borrowed from the Apple Library's experience in its migration from the CP/M Apple III environment. The most difficult step involved the transfer of the data from the 8-inch floppy format of the Digital Microsystems' computer to the Macintosh. The telecommunications software on the Digital Microsystems' computer had limited features and many outdated commands. Using a stripped-down version of *Versaterm Pro* (a telecommunications program for the Macintosh), most hand-shaking protocols could be ignored, and the transfer was successful. The computer at the Mechanics' Institute Library was accessed through the phone lines and controlled remotely from the receiving Macintosh. The originating files, containing some 6,000 records, were accessed and downloaded over a five-hour period. The file was broken down into two segments and copied onto several 3.5-inch diskettes. It was then sent to the Merriam Center Library for analysis and importation into the developed template in Microsoft *File*. Transfer into the template was not without problems, and several staff members checked every record, correcting errors that occurred in the course of the migration. All data were checked against a printout created by the originating computer at the Mechanics' Institute. The new database, which broke the membership into numerous categories, was shipped after all tests were completed, and record accuracy was verified against the original printout.

Conclusion: Like the Apple Library, the Mechanics' Institute Library learned a difficult lesson by relying on extinct hardware and difficult software. By migrating to the Macintosh, the library is in a better position to take advantage of new hardware developments to meet its growing needs. Using Microsoft *File*, it has some control over the nature of the membership database, with an ability to generate a number of subfiles, impossible on the original DataTrek program. When the staff outgrows the limitations of Microsoft *File*, its standard format and export features will ease the transfer to a more sophisticated database program designed for the Macintosh interface.

Case Study 3

Library: Apple Corporate Library, Apple Computer, Inc., Cupertino, Calif.

Obsolete System:

Hardware: Apple III with 5 MB hard disk.

Software: *PFS: File*.

Applications being transferred: Serials database.

New System:

Hardware: Apple Macintosh II

Software: *Filemaker Plus* by Forethought

Transfer software:

Apple III: *Access III* by Apple Computer

Macintosh: *SmartCom II* by Hayes Microcomputer Products

History: When the serials database was originally created, *PFS: File* was felt to be the best program to meet the Apple Library's needs using the Apple III as the basic hardware. With the growth of various database applications in the Macintosh environment, it was felt that a migration was appropriate to take full advantage of both this application's strengths and the graphic interface of the Macintosh. *Filemarker Plus* was chosen as the database application for the Macintosh, largely because responsibility for the design, transfer, and long-term maintenance of the serials file would reside with the staff rather than a consultant. *Filemaker Plus* was sufficiently sophisticated to handle the task, yet could be easily manipulated by a staff member with little programming experience.

The first stage in the transfer was relatively easy since *PFS: File* used traditional Apple III system commands. An ASCII text file was generated by simply printing a copy of all records to floppy disk. Transferring the Apple III-based file was straightforward as well. A hardware connection between the Apple III and the Macintosh failed, due largely to the lack of high-level protocols accessible in the Apple III telecommunications program. As an alternative, the entire serial file was electronically sent to an address on Ontyme, an electronic mail service. The Macintosh then accessed

Ontyme and downloaded the text-converted serials file. A word processing program was used to edit the file, preparing it for importation into *Filemaker Plus*. The edit process lasted about ten hours, due mainly to the size of the file. Thanks to some field order preparation in the Apple III environment, and the ability to specify incoming field order in *Filemaker Plus*, the final importation of the 800 records took about five minutes.

Conclusion: This migration was extremely successful due in large part to adequate planning and the transfer functions of both the originating and receiving software. It is anticipated that any future transfer will be blessed with equal success, thanks to *Filemaker Plus'* features and the existence of basic import/export commands in nearly all current database applications. It is clear that database software is vital to a library's operation. Frequent review of software in hand, software and hardware on the marketplace, and staff and patron needs will prevent a library from being caught in the position of making a difficult transition to new software, risking day-to-day operations.

FUTURE PROSPECTS

Data migration will become easier over the next decade as hardware and software manufacturers embrace basic standards for database management and systems software. Many current database and software problems are already being rectified by hardware developments of more powerful desktop personal computers equipped with new microprocessors and math coprocessors. Soon programmers will no longer need to create machine-specific code and be forced to make allowance for the characteristic idiosyncrasies of a give computer.[11]

Another part of the solution will come with the development of universally agreed levels of transparent connectivity for networks. The vehicle for transparent connectivity may be possible with the eventual adoption of protocols for IBM's System Network Architecture (SNA). Software and hardware "hooks" for both mainframes and personal computers will permit the access of database information in any computing location without the need for formal translation vehicles. These hooks will also permit the transfer of data across the entire computing environment with minimal loss of data. This level of computer dialog, formally called Advanced Peer-to-Peer Communications (APPC), is feasible with its two main components, called physical units and logical units. The latter are the software interfaces allowing communications to a mainframe or any other computer without the need for terminal emulation software. Physical units are the hardware intermediaries required to transmit information between various computers.[12] APPC has been viewed primarily as a tool to facilitate mainframe and personal computer interaction, but it also can be viewed as a way to permit different personal computers, operating in a network with a larger computer, to communicate with each other.

For the IBM OS/2 series of computers, these communications plans will permit the broad access and use of database management systems.[13] MacAPPC, which was released in 1988, will permit the integration of the Apple Macintosh family into the IBM mainframe and token-ring environment. The broadening support for APPC means that problems of transferring information from databases across differing computers will be a fading nightmare in the 1990s.[14]

CONCLUSION

Libraries using dated hardware and software need to evaluate their dependency on old database management systems and the usefulness of older computing hardware. This is especially true in light of recent advances in database design and the incorporation of multiprocessors with faster clock speeds in many high-end personal computers and workstations. Migration of data to new applications, operating systems, and hardware is rife with difficulties. In general, a data transfer project should be examined as a three-step process to minimize problems. It involves (1) placing the original data into a standard text file format, (2) physically transferring the file to its new hardware, and (3) placing the original data into its new database and checking for transmission errors. The frustrations and expenses involved in the entire procedure are compensated by long-term benefits of a more powerful hardware and software base. These benefits are realized by increased options for local data manipulation, easier-to-use hardware with reduced training costs, and accelerated basic functions in the database management program. Compared to the modest expenses for the entire upgrade and transferral, retaining an archaic database and accompanying hardware is too expensive to justify in any reasonable fashion.

REFERENCES AND NOTES

1. A survey conducted in 1984 of microcomputer use in U.S. libraries found that three brands of computers were most popular: Apple, IBM, and Tandy (Radio Shack). A number of units mentioned in the survey are no longer manufactured, including the Apple Lisa and the original IBM PC. See *Micros at Work: Case Studies of Microcomputers in Libraries*, Jim Milliot, comp. (White Plains, N.Y.: Knowledge Industry, 1985), p.11–14.
2. "[The task of libraries] is further complicated by the fact that the types of technologies found in present-day libraries vary widely. New technologies rarely replace previous ones. Libraries exhibit this mix of technologies perhaps more clearly than any other institution."—Barbara B. Moran, *Academic Libraries: The Changing Knowledge Centers of Colleges and Universities* (Washington, D.C.: Assoc. for the Study of Higher Education, 1984), p.5.
3. "In today's complex and highly competitive world, it usually takes all the time an individual can devote to a discipline or set of job functions just to maintain an adequate level of proficiency. Any additional work requirement—such as having to be directly involved with the operation, use or management of computer resources—not only reduces time available for basic job tasks but can also disrupt the fluidity of task execution necessary to perform well in a job."—William Kimmerly, "Toward Spontaneous Computing," *Computerworld* 21, no.41: 111 (Oct. 12, 1987).
4. Ross Altman, "A Cooperative Effort: Sharing the Processing Load," *Computerworld Focus* 21, no.44A:37–38 (Nov. 4, 1987).
5. Mark Goss quoted in Peter H. Lewis, "Ending the Pain of Data Transfer," *New York Times* 137, no.47:F12 (Nov. 29, 1987).
6. These questions were posed by Paul Saffo, a research fellow at the Institute for the Future in Menlo Park, Calif., in Peter Scisco, "Dreams of Access without Effort," *Computerworld* 22, no.13:S3 (Mar. 28, 1988).
7. Alfred Glossbrenner, "The Fine Art of Data Transfer," *Popular Computing* 4, no.11:70 (Sept. 1985).
8. Ibid.
9. Glossbrenner, p.118.
10. Bruce Brown, "From Port to Port: Navigating Data Transfer," *PC Magazine* 7, no.1: 187–89 (Jan. 12, 1988).
11. Gary Chapman, New York University, in Megan Jill Paznik, "Integration—It's Not Here Yet, But It's Coming On Strong," *Administrative Management* 47, no.11:39–40 (Nov. 1986).
12. Peter Lisker, "MacAPPC; Big League Connectivity," *Macintosh Today* 2, no.14:22 (June 6, 1988); Megan Jill Paznik, "The Micro-to-Mainframe Connection Grows Up," *Administrative Management* 48, no.10:36 (Oct. 1987).
13. Douglas Barney, "Putting OS/2 in the Big Picture," *Computerworld* 22, no.20:31 (May 16, 1988).
14. Lisker, "MacAPPC." ■■

Five PC-Based Expert Systems for Business Reference: An Evaluation

Richard G. Vedder, Maurice G. Fortin, Scott A. Lemmermann, and Ralph N. Johnson

In the past few years the commercial software market has witnessed the introduction of several expert-system tools called *shells*. At first these application generators were expensive and required special AI hardware that was also costly. This state of affairs naturally restricted the number of organizations participating in expert-system R&D as well as the range of potential applications deemed cost-effective.

Today a number of new shells, some costing as little as $100, are available for use with inexpensive personal computer hardware of the IBM type. These tools can build small- to medium-scale expert systems or serve as prototype environments for larger projects. The advent of PC-based expert-system shells offers more organizations (including educational institutions) the opportunity to participate in this new technology (see Williamson in bibliography). Many libraries are beginning to utilize this type of software to provide simple directional sources or sophisticated reference service (see Epstein, Rorvig, and Rada in bibliography).

This article examines five different PC-based shells used by teams of graduate students with knowledge engineering experience to build a prototype application. The examination demonstrates the strengths and weaknesses of a product for developing expert systems.

OVERALL PROJECT DESIGN

Ten M.B.A. and Ph.D. graduate students formed five development teams. All students were familiar with knowledge engineering techniques, having passed a graduate-level course in expert-systems development. The expert-system building tools, all commercial shell products with prices starting at less than $500 were EXSYS, 1st-CLASS, GURU, Personal Consultant Easy, and Personal Consultant Plus. (See appendix A for list of manufacturers.)

Each team worked with a different shell. Except for the team using EXSYS (all the other students were familiar with this shell), tool assignments were random. Development and delivery hardware were the same in all cases: IBM-XT PC clones operating at an onboard clock speed of 8 MHz with 640 Kbyte main memory, 20 Mbyte hard drives, and EGA monitors.

The teams spent the first 4.5 weeks of the 14.5-week research project learning how to use their tools. They relied only on product documentation, any other relevant documentation they could find, and past experiences. For example, the GURU team made use of the GURU-oriented *Business Expert Systems* (see bibliography). No vendor provided any training. At the end of this phase, each team submitted a sample application identifying an "animal" from a set of characteristics supplied by the user. The teams did not know the nature of the main development project during this phase.

The second phase lasted six weeks. A general meeting introduced the teams to the common application: building a prototype business reference adviser for the main li-

---

Richard G. Vedder of the BCIS department and Maurice G. Fortin and Scott A. Lemmermann of the library are at the University of North Texas, Denton. Ralph N. Johnson is with the library at the University of Arizona, Tucson.

brary at the University of North Texas. To maximize the availability of expertise, three librarians specializing in business reference served as human experts; these professionals coordinated their efforts to provide uniformity of opinion. In addition, other librarians providing business-reference service (the teams could observe but not interview them) prepared a set of background readings and other materials for use by each team. The reference librarians serving as experts did not interrupt their normal work routine to assist the development teams, with whom they spent approximately 5–10 percent of their workweek for the duration of the project.

During the final four-week phase the human experts evaluated the applications, and the teams prepared their final reports.

PROBLEM DESCRIPTION

The chosen domain was the business-reference activity at the University of North Texas (UNT) libraries, where as many as 50 percent of reference questions received by the staff at the general reference desk include requests for specific business information. The task is a complex one in that solving business-reference problems calls for different levels and types of expertise. Answers require not only an extensive knowledge of varied source materials but also an ability to judge the quality of information available from each source.

The user group (i.e., library patrons) for this service at UNT is heterogeneous, composed of undergraduate and graduate students and faculty. Some of these patrons are knowledgeable users of business-reference resources and library services, but most are not.

FUNCTIONAL SPECIFICATION FOR THE EXPERT BUSINESS LIBRARIAN PROTOTYPE

Each project team used its tool to build a prototype for assisting patrons with basic business-reference questions, regardless of their difficulty level. Given the time constraint on the project, the teams would address only three major request categories within the problem domain: general business, specific industry, and specific company information. For the last two categories, the prototypes would provide advice on both public and privately owned firms.

Patrons would directly access each system by keyboard with a minimum of keystrokes, and each prototype would outline what subjects the patron could consult. Based on responses to a series of questions, the prototype would suggest possible sources to examine; hence the functional specification for these prototypes was diagnostic/prescriptive in nature. Finally, the system would refer the patron to a reference librarian if it could not answer a given question.

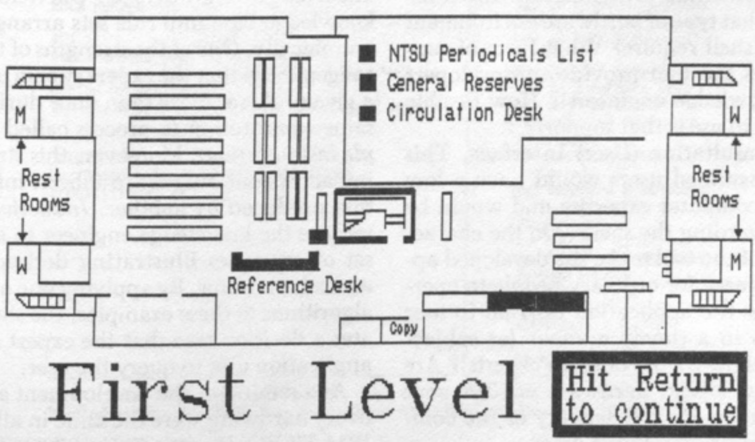


Fig.1. Dr. HALO II-Generated Map.

To test each shell's ability to access external software, the prototypes were to use graphics for directional information. For this purpose, all the teams had access to the same floor plans, drawn using Media Cybernetics' Dr. HALO II (see figure 1). With some shells the teams had to use both a graphics and a text integration package—American Programmers Guild's INSET—to reformat the HALO graphics for display.

EVALUATION DESIGN AND PERFORMANCE CRITERIA

It was important to agree on the evaluation design before initiating work because there were two distinct interest groups involved, each having different goals, motivations, and expectations. The expert system developers were interested in testing the limits of each shell. On the other hand, the librarians serving as project experts were interested in applying expert systems technology to the library environment. They wanted an effective prototype for testing by library patrons. Ease of use, clarity of advice, and thorough patron support were important advantages for them. Consequently, the following issues (not in rank order) formed the basis for the evaluation process. One or more sample questions provide illustration.

1. Power and flexibility. Does the method that a shell uses to represent and store knowledge impact the size and nature of applications possible at the personal computer level?

2. Application Development Environment. What type of hardware environment does the shell require? What form of support does the tool provide to developers (i.e., knowledge engineers)? How flexible and easy to use is that support?

3. Consultation (User) Interface. This project assumed users would have a low level of computer expertise and would be naive regarding the shells. Do the characteristics of the tool make the developed application easy for ordinary people to operate? Does the application respond to user demands in a timely manner (as subjectively judged by the domain experts)? Are there any special hardware environment requirements for the delivery of the completed application to the user?

4. System Connectivity (with external software). Should the need arise, can the shell invoke and/or exchange data with external software? For this application, how well does the tool support the use of external graphics?

5. Training Requirements and Documentation. What percentage of total project time does learning the shell take? Can developers learn how to use the tool effectively and efficiently without formal training? How well does the documentation supplied with the tool assist the learning and application development processes?

6. Hidden Costs. Are there any costs that may not be apparent?

7. Ease of Changing or Updating the Program. Could an organization employing an application written using this shell (in this case the domain experts and their staff) make day-to-day changes, or would this require a trained knowledge engineer?

Obtaining good evaluations was not easy—no team succeeded in building a complete prototype in the allotted time span. Therefore, the evaluation process dealt with partial implementations only.

TOOL EVALUATIONS

This section classifies each product according to the taxonomy found in Harmon, Maus, and Morrissey (see bibliography), which defines *simple rule-based tools* as shells that record knowledge in rule form and store those rules in a single knowledge base. *Structured rule-based tools* permit knowledge engineers to partition the knowledge base into rule sets arranged hierarchically. One of the strengths of this arrangement is that the expert system can use a given rule set more than once during the same consultation (a process called *multiple instantiation*). Moreover, this structuring allows one rule set to inherit information produced by another. *Inductive tools* require the knowledge engineer to enter a set of examples illustrating desired facts and relationships. By applying one or more algorithms to these examples, the shell creates a decision tree that the expert system application uses to query the user.

As a reminder, the development and delivery hardware were the same in all cases: IBM-XT PC clones operating at an onboard clock speed of 8 MHz with 640 Kbyte main memory, 20 Mbyte hard drives, and EGA monitors. Dr. HALO II, the product used to draw the floor-plan graphics, does not offer an easy (i.e., a transparent, quick turnkey) way to show these graphics when called by another program; consequently all shells had to reformat the graphics before they could display them. Some shells were supplied with a utility program for this purpose; others had to call the software product INSET to perform this task.

1. EXSYS.

A. *General Description*. EXSYS belongs to the class of simple rule-based tools. Written in the *C* programming language, the product has three major modules: a demonstration and tutorial system; a development module (EDITXS); and a runtime module (EXSYS). The shell also includes three utility programs that can improve execution speed or combine two knowledge bases. This product supplies a runtime version of the program for use with completed applications.

B. *Developer Interface*. Judged easy to learn by the development team, EXSYS includes online demo and tutorial software as well as sample knowledge bases. Documentation consists of one manual and a card containing a distilled version of the command set and operating instructions. The team considered the documentation unevenly written in that, while offering many examples, some issues (such as external interfacing) received less explicit attention. In addition, EDITXS offers the developer considerable online help, which is usually just one keystroke away. Although the team judged EXSYS itself easy to learn, the external interface issues proved difficult—approximately 25 percent of the team's effort was devoted to learning the system, and most of that was spent on external interface issues.

EDITXS first asks the developer a series of questions that customize the general appearance and operation of the intended application. These questions include selecting a likelihood or probability scale to use with the reasoning process and creating a list of "choices," i.e., goals or conclusions for EXSYS to pursue. Note that the developer must choose a probability scale (the three possible alternatives are 0–1, 0–10, and 0–100) and that this scale is the only part of any application that cannot be altered later. EDITXS then presents a screen divided into three windows: the right window is where the developer builds the components of each rule; the left window displays the current status of the rule under construction or revision; the bottom window displays a list of available commands (see figure 2). To build a rule of the If/Then/Else format, the developer usually builds and/or selects rule components from lists of *qualifiers* (sentences having a variety

```
RULE NUMBER: 34                                        | 1  Map to General Reserves
                                                       | 2  Map to Circulation Desk
IF:                                                    | 3  Map to MTSU Periodicals
Stock Trends yes                                       |      List
The level of investor interest in the Stock is        | 4  For further information,
Medium or not known                                    |      go to Reference Desk
The depth of stock information required is High        | 5  Use General Company
                                                       |      Information Reference
                                                       |      Materials
THEN:                                                  | 6  Use Ratio Reference
Use Moody's Handbook of Common Stock -                 |      Materials
Probability=8/10                                       | 7  Use Industry & Economic
                                                       |      Data Reference
                                                       |      Materials
                                                       | 8  Use Moody's Manual
                                                       | 9  Use Moody's Handbook of
                                                       |      Common Stock
                                                       | 10 Use Q Data File
                                                       | 11 Use Standard and Poor's
                                                       |      Corporation Record
                                                       |
                                                       |    <more choices>
-----------------------------------------------------------------------------
Input choice number, New choice <N>, Typo correction <T>, cancel <ESC>,
Delete/reorder <D>, Find <F>, Help <H> or any other key for more choices:
```

Fig. 2. EDITXS' Rule-Building Windows.

of possible endings or *values*) or from the list of choices. This cut-and-paste technique reduces keystroking to a minimum; after building a few rules, the developer can experiment with the work without leaving EDITXS.

The development team did not believe anyone with more than an introductory level of knowledge about expert systems needed formal training in order to use EXSYS.

C. *Consultation Interface*. Qualifiers not only serve as rule constructs but are also the primary medium for interrogating the user. During a consultation EXSYS usually queries the user for information by displaying a given qualifier with its associated range of possible values. (Recall that this construct consists of a sentence fragment with a list of possible endings.) Users may select one or more answers from this multiple-choice menu. Alternatively, EXSYS can prompt the users for specific numeric or string data stored as a variable; after reaching a conclusion, they can change any part of their input data and immediately rerun the application. The program also permits them to save input data and/or results for later consultations (see figure 3).

EXSYS offers various modes of online help. After loading a knowledge base, the program asks if the user wants instructions on how to operate the system. At the bottom of most screens, two lines of instructions on command alternatives are offered; more detailed help is only a keystroke away. Developers can also provide custom help files.

D. *External Interface*. EXSYS' Run command calls external programs at any point during a consultation session, permitting the team to display the floor-plan graphics whenever desired (but the team had to reformat the HALO graphics with INSET before EXSYS would display them). EXSYS also can export and import data either in file form or as a parameter string accompanying the Run command.

E. *Domain Experts' Evaluation*. The business-reference experts believed users could quickly learn how to operate the EXSYS application, rated its response time as fast, and liked the menu-driven user interface. The EXSYS team used a confidence factor scale of 0–10, and the experts felt this did not allow enough discrimination among the choices. The team also designed the application to segregate users at the outset on the basis of familiarity with business-reference sources, a strategy that made system operation more efficient. The detailed help screens would assist novice users in operating the system.

2. 1st-CLASS.

A. *General Description*. 1st-CLASS belongs to the class of inductive tools. It comes with user's manual and two diskettes, one for development work and one that provides a runtime environment. It is written in Microsoft Pascal and assembler.

B. *Developer Interface*. The development team found 1st-CLASS easy to learn, estimating that the time spent learning this shell at less than 5 percent of the total time involved with the project. One reason for this is that 1st-CLASS presents the various development screens in a format similar to that found in electronic spreadsheets. A window at the top of the screen displays the list of major commands (the F9 and F10 keys allow movement among them). Highlighting any one of these commands reveals the associated group of second-level commands, each accessible by a single keystroke. Moreover, the ordering of the major commands reflects the logical sequence one would follow in building an expert system application. A third line of text indicates what the function keys will do and provides other status information. Much of the actual construction process is oriented to column (attribute) and row (value), another spreadsheet characteristic.

Although the software lacks an online tutorial, it does contain a large demo program. Both diskettes also come with a wealth of sample knowledge bases to explore, and the F1 key always provides extensive and scrollable online help.

1st-CLASS guides the developer through the building process, first identifying attributes, supplying values, and creating examples. After applying the algorithm to the matrix, the developer can examine the decision tree, which 1st-CLASS presents in left-to-right format. A chapter in the manual discusses testing and debugging strategies (see figure 4).

1st-CLASS requires the use of confidence or certainty factors called *weights*, with each example. The 0–100 scale may include real numbers, and the shell assigns default weights of 1.0. 1st-CLASS also includes a utility program that aids creation of pop-up windows for extensive user help; a report generator is available for storing session data.

The manual assumes little or no previous knowledge with expert systems. While beneficial for novices, this approach exasperated the team—as experienced knowledge engineers, they thought this material got in the way of locating needed information and suggested reorganizing the manual to divide and concentrate instructional material from reference and advanced materials. The manual does have an index, which was judged usable but not as detailed as one might wish (e.g., references to function keys are missing). The manual also contains several helpful appendixes, including one devoted to error messages.

The development team liked 1st-CLASS' support for text editing and decision-tree tracing, as well as the visual feedback provided by the decision trees themselves. But

```
RULE NUMBER: 7
IF:

      (1)       Map required no
and   (2)       Type of business information required is Industry ratio information

THEN:

                Use Ratio Reference Materials - Probability=10/10
and             For further information, go to Reference Desk - Probability=5/10

IF line # for derivation, <K>-known data, <C>-choices
↑ or ↓ - prev. or next rule, <J>-jump, <H>-help or <ENTER> to continue:
```

Fig. 3. EXSYS' Consultation Interface Features.

```
new_Example, Replicate, Change, Activate, Move, Delete
         Files  Definitions  Examples  Methods  Rule  Advisor
[F1=Help]      40 Examples in WILLIS        [F9=Definitions] [F10=Methods]
                                                              weights-->
        WHAT-IS-IT  MAP      I-WANT     lit-searc-f bus-que    Acct-que
>  20:  DIRECTIONS  YES      *          *           *          *
   21:  DIRECTIONS  YES      *          *           *          *
   22:  DIRECTIONS  NO       INFOTRAC   *           *          *
   23:  DIRECTIONS  NO       INFOTRAC   *           *          *
   24:  OTHER       *        *          *           *          *
   25:  LIT-SEARCH  *        *          business    Acct       tax
   26:  LIT-SEARCH  *        *          business    Acct       audit
   27:  LIT-SEARCH  *        *          business    Acct       other
   28:  LIT-SEARCH  *        *          business    fin-ratio  *
   29:  LIT-SEARCH  *        *          business    fin-ratio  *
   30:  LIT-SEARCH  *        *          business    Econ       *
   31:  LIT-SEARCH  *        *          business    Econ       *
   32:  LIT-SEARCH  *        *          business    Econ       *
   33:  LIT-SEARCH  *        *          business    Govern     *
   34:  LIT-SEARCH  *        *          business    Govern     *
   35:  LIT-SEARCH  *        *          business    Govern     *
   36:  LIT-SEARCH  *        *          business    Co-info    *
   37:  LIT-SEARCH  *        *          business    Co-info    *
   38:  LIT-SEARCH  *        *          business    Co-info    *
What can I help you with ?
```

Fig. 4. First-Class Decision Tree Matrix.

they faulted 1st-CLASS for limiting the size of a matrix to thirty-two attributes/factors and thirty-two values per attribute. More complicated problems (such as the main project) require a segmented solution effort using a series of forward- or backward-chained 1st-CLASS matrices. The development team did not believe this was easy to do.

This discussion highlights a major weakness of 1st-CLASS: its lack of flexibility in accommodating increasingly complicated knowledge bases. New rules (i.e., additions or other changes to the decision tree) require building a new matrix and rerunning the algorithm. When a complex application is needed, other expert system solutions may well be necessary.

C. *Consultation Interface.* The team praised 1st-CLASS' user interface, which makes applications very easy to learn and operate. Users merely highlight their value choices for each queried attribute. At the end of a consultation users can, with a keystroke, replay the finished consultation (changing responses as desired), start a new consultation, or quit. During consultations a status line informs them of special functions assigned to keys. 1st-CLASS supports creation of data-entry forms, if developers and users prefer this to the menu approach.

1st-CLASS offers extensive facilities for providing online help. Developers can customize up to seven help hot keys (F2–F8), each of which displays different help texts for each consultation screen.

D. *External Interface.* 1st-CLASS can call and move data to and from DOS, Lotus 1-2-3, and other outside programs, but external programs have a limited ability to make calls to 1st-CLASS for analysis purposes. The team did not find it easy to incorporate graphics into their 1st-CLASS application. The shell comes with Pascal source and compiled-code versions of a program called SLIDE1, which will display picture files made with PC-Paintbrush. Since the library floor plans were drawn with Dr. HALO II, the application had to call a DOS batch file to have INSET show the graphics.

E. *Domain Experts' Evaluation.* The business-reference experts were impressed by the ease with which a user could move back to the beginning of a consultation without having to wade through any intermediate screens. The domain experts judged the response time of the project application to be slow.

3. GURU.

A. *General Description.* GURU provides developers with an integrated software environment for building expert system applications. In addition to a shell, it offers an electronic spreadsheet, relational database management system, natural-language data retrieval system, fourth-generation programming language, data communications package, business graphics, and many other features. The project's time limitations forced the GURU team to work with only the expert systems shell and the fourth-generation language. Thus much of GURU's power, stemming from the interrelationship of its many components, remained unexplored.

GURU is written in *C*. The product consists of four manuals and five diskettes (requiring about 1.5 Mbytes of hard disk storage), excluding demo and tutorial software. The vendor sells a runtime version of GURU separately for use with completed applications. The product can access expanded or extended main memory.

GURU belongs to the class of structured rule-based tools. Its structure centers on the role of variables in the rule base. The variables within each rule set form a hierarchical network of dependencies (for obtaining their values). Storing values for these variables in GURU's database makes inheritance possible. Moreover, database commands can repeatedly apply different values to a rule's variables, thus supporting multiple instantiation.

B. *Developer Interface.* The inherent power and flexibility of GURU makes it hard to master. The development team spent approximately 60 percent of its time trying to learn how to use just a small part of this product. As a result, the team strongly recommended attending the three-day training seminars offered by the vendor (something the team had neither the time nor opportunity to do). GURU's online tutorial is helpful but not a substitute for the training sessions.

GURU's documentation, while massive,

is extremely poor in examples and proved difficult to use. It was not always easy to find information required for the expert system shell amid instructions pertaining to the spreadsheet, database, or other components in this integrated software product. Moreover, the manuals were written for persons familiar with MDBS' other products and do not employ language familiar to knowledge engineers and others experienced with AI. While online development help is abundant, the team found the information value of these screens limited. Also, the simple knowledge bases supplied with the demo software were not detailed enough for learning much. The team did find helpful Holsapple and Whinston's *Business Expert Systems*, which uses a scenario knowledge base written for GURU (see bibliography).

GURU supports both a command- and a menu-driven development interface; the software has no practical limits on the number of rules or variables used. Both the developer and the user can save and replay consultations. The use of certainty factors (default scale 0–100) is optional.

Although developers can test individual rules or the entire expert system application in a stand-alone mode (i.e., without involving the other parts of the integrated packages), the GURU team did not find the package's debugging aids easy to learn or particularly useful.

For building its application, the GURU team did not prefer to use the menu-driven text editor supplied with the package. To speed up development the team relied instead on an external, full-screen text editor, then copied the work into GURU (see figure 5).

C. *Consultation Interface.* Users can initiate a consultation as a stand-alone activity, a menu pick, or from other parts of GURU (e.g., interpretation of data values in a GURU spreadsheet). Methods for user input are menu picks, numeric or data strings, and natural language. The development team chose and implemented a menu-driven format. Time constraints prevented examination of the natural-language software.

D. *External Interface.* The integrated software components within GURU reduce the need for external calls. When they prove necessary, though, a Run command will invoke any external software that can coexist with GURU inside main memory. It can also import and export data with external software.

Because GURU supports only a limited range of business graphics, the team used the Run command to access the Dr. HALO images via INSET.

E. *Domain Experts' Evaluation.* The business-reference experts did not find the GURU application easy to learn or use, judging the response time to be fair. GURU allows users to view and amend each search. The development team used short

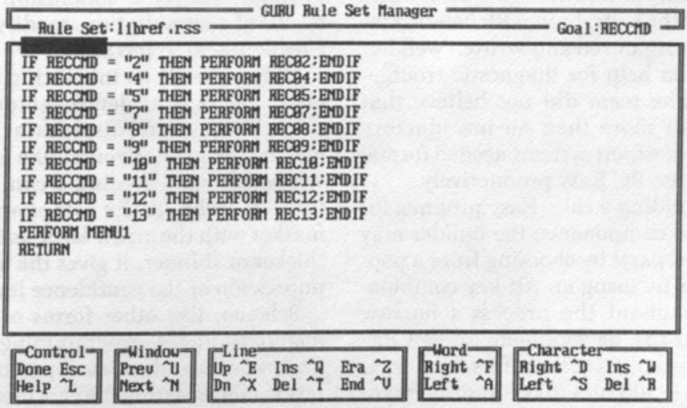


Fig. 5. GURU Text Editor Screen.

abstracts to describe the contents of each source, and this would prove helpful to users of the system. Here again, searchers must pass through intermediate screens in order to initiate additional search inquiries.

4. Personal Consultant Easy (PC Easy).

A. *General Description*. PC Easy belongs to the class of structured rule-based tools. Written in PC Scheme, a dialect of Common Lisp, the product comes with two manuals, "Getting Started" and "Reference Guide," and has two software modules for development and runtime.

Texas Instruments (TI) views its product as a simplified introduction to developing expert systems. Knowledge bases created under PC Easy are upwardly compatible with the more versatile PC Plus. The runtime diskette converts the finished application into a protected format for distribution.

B. *Developer Interface*. Judging this shell easy to learn, the team's estimate of the time spent learning PC Easy was less than 10 percent of total project time. The "Getting Started" manual specifies two learning paths, one for novices and a second for veteran knowledge engineers. Most of it is written for those with no experience in knowledge engineering, and a useful list of reference works is included. The "Reference Guide" manual is the main source of help for knowledge engineers who just need information on the nature of PC Easy. There is no online tutorial, but this volume does provide a written one. (The software includes sample knowledge bases.) The team found the texts, both with detailed indexes, to be organized and written well but very short on help for diagnostic troubleshooting. The team did not believe that anyone with more than an introductory knowledge of expert systems needed formal training to use PC Easy productively.

When building a rule, Easy prompts for the required components; the builder may add optional parts by choosing from a popup menu or by using an Alt key combination. Throughout the process a narrow window on the development screen displays clear prompts. Easy offers developers the use of a -100 to $+100$ certainty factor scale. Unlike EXSYS, however, the use of certainty factors is optional.

Response time began to suffer as the rule base approached and then exceeded 100 rules. Consequently the team felt Easy was suitable for developing prototypes and full expert systems with knowledge bases of less than 100 rules.

In development mode, the F1 function key provides limited general help with the various components available for use in the rule base. The F2 key displays a list of possible options available to the developer at any stage during the construction process (see figure 6).

C. *Consultation Interface*. The Easy user interface provides for either a menu format or prompts for numerical or string data entry. Developers can customize each menu presentation for single or multiple choices. Users merely highlight menu items using the arrow keys, thus saving keystrokes and lessening the chance of error; they can save their input and the system's output for later use.

Easy offers several forms of help. Each screen requesting a user response includes brief instructions at the bottom. The F1 function key can access Help screens customized by the developers for the given application. Unless a graphic is used, however, the small size of the available help screen restricts the amount of textual assistance that developers can offer at any given time. The F2 key provides the user with a menu of functions that can explain the purpose of a question at issue, display what the system currently knows about the problem at hand, justify a conclusion, store the results of a consultation, modify and rerun a session, and so forth.

Easy supports an interesting method for obtaining a confidence factor from the user. The system displays a graphic in which a scaled horizontal line is associated with each item needing a confidence factor. On each line the user simply moves a marker with the arrow keys; as the line gets thicker or thinner, it gives the user a visual impression of the confidence level.

Scheme, like other forms of LISP, is a memory-hungry programming language. To operate in either development or consultation modes within the confines of 640K, it must employ "garbage collection" routines for discovering main memory locations that contain data no longer needed by Easy. The program then reallocates those addresses to hold new data. Whenever this process occurs, execution of the application stops and the Active message flashes in the upper right corner of the screen. Obviously this slows the performance of any significantly sized application (see figure 7).

D. *External Interface*. The development team appreciated the ability of Easy to display graphics at any point in a consultation, wherever a graphic image might be appropriate. The shell includes a utility program called SNAPSHOT, which reformatted the Dr. HALO graphics for use by Easy.

Easy can invoke external software, such as DOS commands and dBASE II, III, and III Plus, and exchange data with dBASE.

E. *Domain Experts' Evaluation*. The business-reference experts liked the user interface and found it easy to learn (if a little tedious to use), but they judged the response time of the Easy application as slow. While response time is a drawback, its relative ease of use does make PC Easy a user-friendly system.

5. Personal Consultant Plus (PC Plus).

A. *General Description*. PC Plus, a

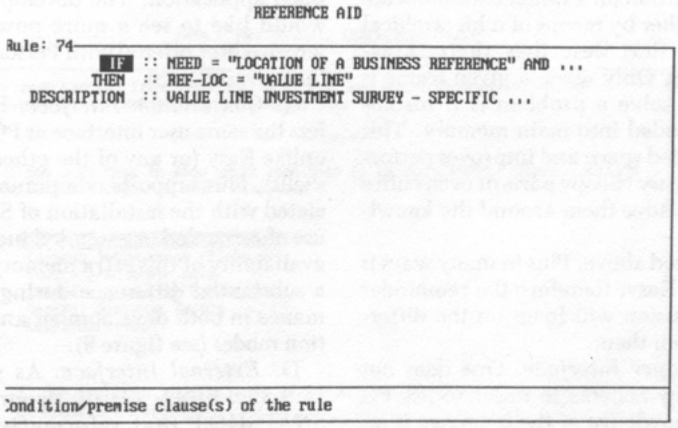


Fig. 6. PC Easy Rule-Base Reference Aid.

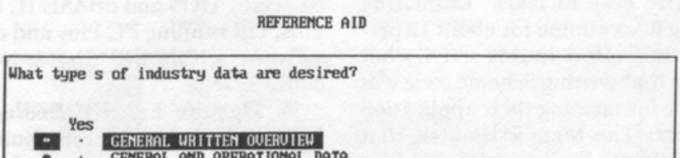


Fig. 7. PC Easy User Help Screen Utilizing LISP.

greatly enhanced version of PC Easy, is a structured, rule-based tool. Like Easy, Plus is written in PC Scheme and has two modules for development and runtime. Plus comes with four manuals (PC Scheme and PC Plus each have a user's guide and a reference manual) and five diskettes requiring about 1.5 Mbytes on a hard disk. TI offers two extra utility packages, PC On-line (for electronic process monitoring) and PC Images (for incorporating animated graphics, e.g., a moving dial indicator). Neither was available for this project, nor are they required in order to build complex applications.

Plus uses rule-set structures called *frames*, each of which addresses a part of the overall problem. Frames communicate with each other by means of a hierarchical frame tree that describes their interrelationships. Only when a given frame is required to solve a problem is it instantiated, or loaded into main memory. This reduces wasted space and improves performance. It is easy to copy parts or even entire frames and move them around the knowledge base.

As indicated above, Plus in many ways is identical to Easy; therefore the remainder of this discussion will focus on the differences between them.

B. *Developer Interface*. One does not have to know Scheme in order to use PC Plus, but knowledge of the language is essential for employing the full capabilities of this shell. The PC Plus team judged the package fairly easy to learn, estimating that learning it accounted for about 15 percent of the total effort in this activity but emphasizing that writing Scheme code was not necessary for meeting their application requirements. The team estimated that mastering Scheme for use with PC Plus would take a significant amount of time—novice developers would probably benefit from attending one of the training seminars offered by TI.

The team rated Plus' documentation adequate but wished for more detail in examples and a reference guide. Plus is so powerful that there is more than one way for developers to accomplish a given task. The team wished for material describing the advantages and disadvantages (e.g., impact on system performance, knowledge base complexity) of these various methods. As with Easy, there is a written but no online tutorial, and there are sample knowledge bases.

The development interface for Plus is very similar to Easy's but has several differences. A significant one, as noted above, is the ability to enter and program directly in Scheme, which provides greater customizing ability and more control over the application than Easy provides. Another important difference is the Tree On/Off command, which can display the growing frame tree graphically from left to right. A third difference is the Build command, which creates a runtime version of the finished application. The development team would like to see a more powerful full-screen editor offered with Personal Consultant (see figure 8).

C. *Consultation Interface*. PC Plus offers the same user interface as PC Easy, but unlike Easy (or any of the other evaluated shells), Plus supports (via parameters associated with the installation of Scheme) the use of expanded or extended memory. The availability of this extra memory can make a substantial difference during its performance in both development and consultation modes (see figure 9).

D. *External Interface*. As with Easy, Plus includes a utility program called SNAPSHOT that reformatted the Dr. HALO graphics. This shell can call and move data to and from external software, especially DOS and dBASE II, III, and III Plus, but running PC Plus and other major software within 640 Kbytes is very difficult.

E. *Domain Experts' Evaluation*. The business-reference experts found Plus one of the easiest to operate, evaluated its response time as good, and considered it a more powerful tool that PC Easy.

CONCLUSIONS

Expert systems technology takes some effort to learn but can pay handsome dividends when applied wisely. Assuming that one has found a good topic for an expert system, this experiment provides several important lessons on choosing a shell tool. A major criterion must be the level of experience that the application developers have with building expert systems, and as this project showed, some shells either assume a novice developer or provide novices a path for learning. Some shells, however, assume nothing of the sort, and their mastery can present a difficult challenge. A second factor to consider is the nature of the required (or at least preferred) hardware environment for development and/or delivery. The trend is clearly in the direction of PC-XT or PC-AT machines with EGA-level graphics, at least 640 Kbytes of main memory, and hard disks. Finally, simple rule-based or induction tools are best suited for relatively simple problems. Structured rule-based tools are a good choice when dealing with logically subdividable or complex problems requiring many rules. PC Plus and GURU are probably the only two shells that proved suitable for implementing a workable business-reference adviser.

The investment in hardware and development time may seem excessive at first glance; however, librarians working with expert system developers can provide users with powerful new tools to enhance reference service in virtually any subject area. In the course of this project, the three librarians who served as human experts did not have to invest a considerable portion of their time in work with the various development teams.

These software tools offer libraries pow-

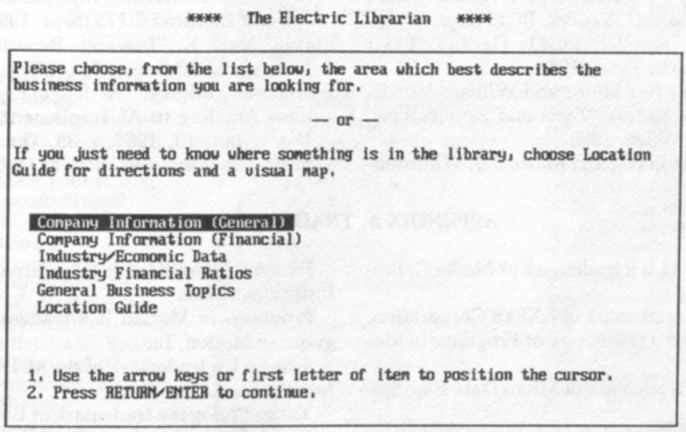


Fig. 8. PC Plus User Aid Utilizing Scheme.

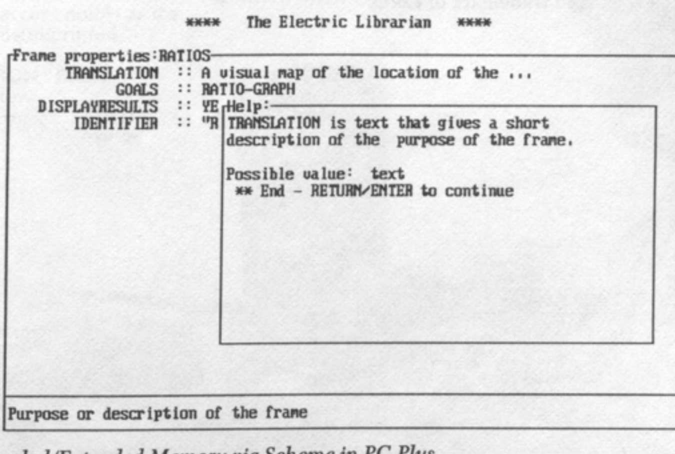


Fig. 9. Expanded/Extended Memory via Scheme in PC Plus.

erful new options for providing service during times of peak use or low staffing. The integration of graphics and color monitors can create readable and user-friendly aids, even to those who are unsophisticated about libraries or computers. Expert systems could relieve pressure at service points by providing directional and information assistance or, when fully developed, simple and complex reference advice. Systems suitable for smaller knowledge databases could be utilized for educational purposes or for intelligent computer-assisted instruction. These types of systems could then be applied to teaching use of specific reference tools or library services such as the online or card catalogs. The experience gained in this project indicates what can be produced in minimum time by librarians working with expert system developers. Updating the resulting products would require a staff member to become knowledgeable about use of the system software.

BIBLIOGRAPHY

Epstein, Hank. "An Expert System for Novice MARC Catalogers." *Wilson Library Bulletin* 62:33 (Nov. 1987).

Evaluation of PC-Based Expert System Shells, Ed. by Richard G. Vedder. BCIS Dept. Working Paper, no. WP-88011. Denton, Tex.: Univ. of North Texas, 1988.

Harmon, Paul, Rex Maus, and William Morrissey. *Expert Systems Tools and Applications*. New York: Wiley, 1988.

Holsapple, Clyde W., and Andrew B. Whinston. *Business Expert Systems*. Homewood, Ill.: Irwin, 1987.

Rada, Roy, and others. "Computerized Guides to Journal Selection." *Information Technology and Libraries* 6:173 (Sept. 1987).

Rorvig, Mark E. "Research Reports." *Library Journal* 112:60 (Sept. 15, 1987).

Williamson, Mickey. "At DuPont, Expert Systems Are Key to AI Implementation." *PC Week*, Jan. 13, 1987, p.35. Documents the heavy use of low-cost shells at DuPont.

APPENDIX A. TRADEMARKS

Dr. HALO II is a trademark of Media Cybernetics, Inc.

EXSYS is a trademark of EXSYS Corporation.

1st-CLASS is a trademark of Programs in Motion, Inc.

GURU is a trademark of Micro Data Base Systems, Inc.

INSET is a trademark of American Programmers Guild, Ltd.

Lotus 1-2-3 is a registered trademark of Lotus Development Corp.

Personal Consultant is a trademark of Texas Instruments, Inc.

Programs in Motion is a trademark of Programs in Motion, Inc.

Scheme is a trademark of the M.I.T. Artificial Intelligence Laboratory.

Turbo Prolog is a trademark of Borland International, Inc. ■■

Learning about CD-ROM Technology: An Educator's Perspective on Sources, Issues, Criteria, Breakthroughs, and Research

Richard S. Halsey

CD-ROM technology is unsettled but promises cost savings and improved access to information. Literature on CD-ROM varies in clarity, authoritativeness, and currency; therefore librarians must develop procedures for acquiring information directly from CD experts. Ways to do this are suggested. Characteristics of reference works that make them amenable to CD-ROM conversion are described. Besides environmental factors, such as budget and available supporting technology, other critical factors are intrinsic value and usefulness, size of text, ideational density, range of subjects, and obsolescence rate. Breakthrough areas to watch include networking, multimedia, and hypertext. Researchers should concentrate on content classification, ergonomics, and the educational impact of hypertext.

CD-ROM technology is alluring but generally overpriced at this stage of its development. It can, however, offer substantial cost benefits and accelerated access to information seekers if used with full knowledge of its capabilities and limitations. Advantage will flow only to those individuals and institutions that can comprehend, weigh, and predict the economics of use that distinguish this particular information carrier from online, microform, print, and other storage-access media; adversity will afflict those who are not as well informed.

It is difficult for a librarian to acquire adequate knowledge just from reading about CD-ROMs because the technical, library, and promotional literatures vary in reliability, clarity, authoritativeness, and currency. The technical literature, including such sources as the *Proceedings of the Society of the Photo-Optical Instrumentation Engineers*, various IEEE Transactions, *Applied Optics*, *Journal of Electronics Engineering*, *Journal of Imaging Technology*, *Journal of Applied Physics*, and reports issued by the Matsushita Electric Industrial Company, Philips, and GE give state-of-the-art information. Because at least 90 percent of the CD research is being conducted in industrial labs and schools of engineering, the technical literature is on the leading edge and can provide the best insights regarding such factors as dimensional stability, abrasion resistance, standards, error correction, resolution, filing systems, servers, and specifications for other CD formats. For most librarians, however, these sources, because they presume their readers' command of mathematics and engineering, remain out of

---

Richard S. Halsey is Dean at the School of Information Science and Policy, State University of New York at Albany.

bounds. But is this really a critical deficiency? How many chefs know how to build and wire microwave ovens, electric ranges, and refrigerators? The technical publications are still tops in timeliness and reliability, while the derivative articles in the library literature focus on user experience and comparative costs and gloss over specifications instead of providing hard data.

Finally, and least dependable because of its self-touting nature, is that all-too-familiar literature, the writings of those affiliated with the anti- or pro-CD-ROM camps. Online promoters see problems galore with CD-ROM and vice versa.

Consequently, librarians should supplement their reading with observations, which means attending (with checklists in hand) such events as the annual CD-ROM Expo and the LITA conferences. These meetings introduce attendees to strategies and applications for library and information services and also reveal the activity in product development. Of course, wariness on the sales floor is crucial: the sellers' enthusiasm can mask the products' shortcomings.

The buyer's library is of course the best site for bringing together CD-ROM experts and librarians. The application of CD-ROM can best be judged in a real library setting, and librarians on their own turf can more freely compare the library's needs to the wonderland of technology.

LEARNING FROM EXPERTS

Librarians cum information professionals who want to get dependable guidance from CD-ROM experts must learn how to employ two techniques, *interrogation* and *confrontation*.

The first of these uses the skill of posing questions to solicit information that is relevant and valid. What can the expert say that will justify a decision to go with or reject CD-ROM? (For example, is online, CD-ROM, print, or a mixture preferred for a particular reference source and why?) And how much confidence can be placed in what experts say? This same modus operandi applies to testimony in the courtroom, where expert witnesses are often found to be as fallible or ignorant in some respects as the questioners.

The purpose of the second technique, confrontation, is to promote discussion among two or more respected experts that generates knowledge, facts, and true cost projections rather than arguments that result merely in the exchange of more ignorance. The point is to observe how contending experts interact, incorporating information gleaned from previous interrogations. Where and why do they agree? Where and why do they disagree? The quality of answers can be assessed only if the right questions are posed, which is why this must be a two-step process. Fortunately, these skills are learnable and are common currency in our culture (see "Perry Mason," "L.A. Law"). This adversarial game may not be entirely transferable or comfortable because of librarians' inclination to be trusting, but it does make a good start and, if played right, could increase severalfold the effectiveness of technology-driven decisions in situations where dollars and knowledge of current optical disk technology and systems are in short supply.

CRITERIA FOR CD-ROM

In any case, when considering CD-ROM products, it is absolutely essential to think carefully about compatibility of the content with the disk format as well as the context and economics of use; local capacity to support CD-ROM's potential; quality and quantity of available hardware and readiness for networking; size and flexibility of budgeting; perceptions and preparedness of users, staff, and administrators; and anticipated frequency of demand. Factors relating to content are size of the knowledge base, complexity and range of subject matter, and susceptibility of the material to obsolescence. Beyond libraries, examples of successful utilization of CD-ROM include the U.S. Postal Service's national address file with 109 million delivery addresses arranged within 25 million address data records and associated ZIP + 4 information. The new system will be considerably more reliable than the older online system that depended upon temperamental telephone links. Microfiche will be retained as a distribution medium.

NYNEX, which has converted its 10 million telephone directory listings in Boston and New York City onto a single CD-ROM and makes the disk and search and retrieval software available, is betting that corporate customers, because of the projected gain in efficiencies of operation, will be willing to pay the annual $10,000 fee for this service, as will law enforcement and marketing agencies and firms that process credit applications for which sorting and report generation are critically important functions.

The use of CD-ROM systems for resource sharing is also becoming increasingly common. In New York State, a CD-ROM union catalog has been developed for the Long Island Library Resources Council and Nassau Library Systems to accommodate access to machine-readable bibliographic records and collections of member libraries. And in California, the Los Angeles Unified School District hopes to boost library services with an automated computer system for management, enhancement of library services, and student instruction. Similar utilization to networking and management has been reported in Missouri, Hawaii, and New York. The Fred Meyer Charitable Trust, besides issuing state-of-the-art reports, has committed money to the Washington Library Network so that nine high schools can purchase this technology. The Albany-Schenectady-Schoharie Counties School Library System (New York) is converting its union catalog to CD-ROM.

The number of basic reference works (e.g., atlases, dictionaries, encyclopedias, biographical directories); abstracting and indexing services (e.g, *Wilsondisc*, *Science Citation Index*, *PsychLIT*); and major databases (e.g., *AGRICOLA*, *Books in Print*, *ERIC*, *MEDLINE*, *NTIS*, *CD-MARC*) being converted to CD-ROM continues to expand. More than 500 titles for business and reference are on the market or in the planning stages. To date, several considerations taken as a whole tend to prime publishers' willingness to develop CD-ROM products. Most important among these is the assurance that urgency, irrefutable need, and purchasing power will generate sufficient income to justify investment. The medical and legal professions, financial markets, military, and, to a lesser extent, public education have the incentive and dollars to pay for the product. Without such assurance, publishers will not move into a high-risk area in which much more work must be done in the development of front-end systems minds, indexing, software programs, editing, color and resolution, networking and provision for simultaneous access, negotiation of rights and releases, and sophisticated multiformat programs. As publishers move into the CD-ROM arena, librarians should assist them in determining which reference sources are most congenial with the medium. For the reference specialist, the following three factors stand out as being critically important:

1. Ease of conversion to CD-ROM. The less the producer has to index, classify, organize, or reorganize the content, the better.

2. Longevity of the information's value. Treasuries of quotations, *Harvard Classics*, *Hortus II*, and *Oxford English Dictionary* contain knowledge with a very slow decay rate; in contrast, airline guides, real estate listings, social studies texts, and commodities exchange price quotations are outdated quickly. The cost of updating a database frequently is high.

3. Size of the database. Because of the extraordinary compression achievable by CD-ROM, databases exceeding 2 gigabytes are especially compatible with this medium.

Figure 1 illustrates the relative appropriateness of various reference tools for CD-ROM as a cube showing a least-likely product—a regional weather advisory service—on the lower left, inside corner; a grouping of most promising products clustered around the upper right, outside corner; and other items with varying qualifications.

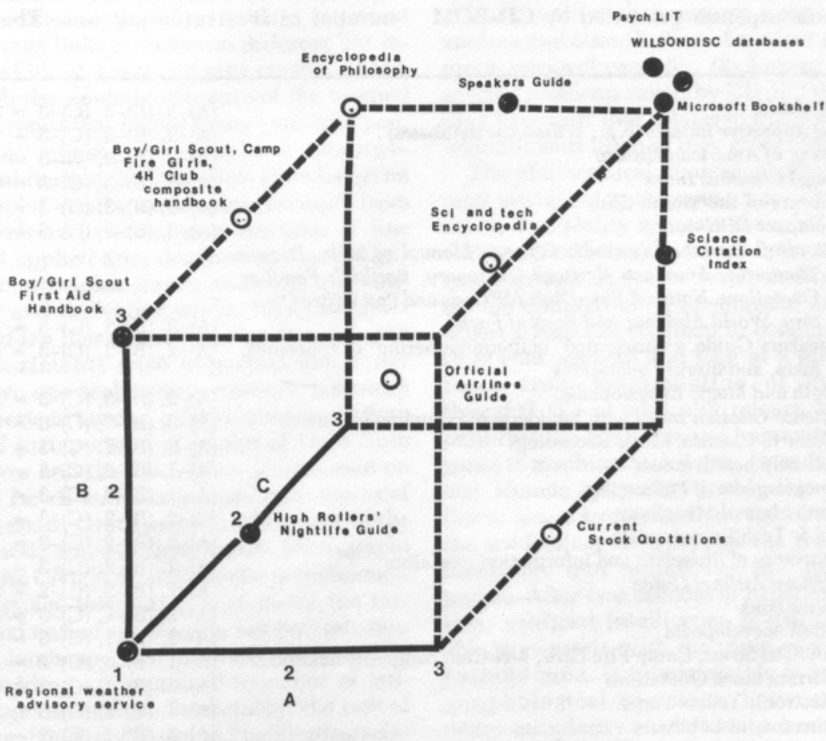
Figure 2 includes some of the CD-ROM candidates shown in figure 1 as well as several additional examples of items, some already published in this format and others still in the planning stage or existing only in the mind of this writer. A rating of 8 or 9 for a real or hypothetical product indicates a promising candidate for the format; items with lower scores may also be justified if the targeted audiences are affluent and strongly motivated to use the product.

Each item is given a score of 1, 2, or 3 across each criterion dimension: ease of organization (A), usable lifetime (B), and size of database (C). The U.S. Postal Service *Zip Code Directory*; NYNEX telephone directory of Boston and New York City; a cumulation of recipes in major cookbooks with a master index of ingredients, nutritional data, nationality, and region; and large retrospective databases such as *AGRICOLA*, *CC/MEDLINE*, *CDMARC-Subjects*, and *NTIS*—all earn perfect 9-point scores, 3 being credited to each criterion.

BARRIERS TO REALIZATION OF CD-ROM'S PROMISE

The progress of CD-ROM is being hindered for a number of reasons so that it may not remain viable. Although lack of uniform standards is a barrier, it is a lesser one than the certain prospect of substantial research and development costs. Because of this and the attendant financial risk, CD-ROM products are much more pricey than they would be if the technology were settled into the marketplace. For that to happen, the following missing ingredients should be in place: (1) cheap disk servers that preclude the necessity for manual changing and the resultant wear of the me-

CD—ROM CUBE OF APPROPRIATENESS



A-axis Ease of conversion or organization
B-axis Longevity of value to user
C-axis Size of database

Fig. 1. CD-ROM Cube of Appropriateness.

dium; (2) CD-ROM drives priced below $500; (3) development of network-server capacity in a distributed online environment, allowing equilibrated access to current and retrospective information in the same and cognate databases; (4) fast access time; (5) revision programs that keep reference works on CD-ROM up-to-date; (6) training programs for users; (7) facilitation of longitudinal searches of large, chronologically arranged databases; (8) search capability that matches the capacity of the medium and is at par with that of online systems (e.g., all indexed fields can be simultaneously searchable, field-defined searching can be accomplished); (9) pricing that is within the reach of the typical school or public library; (10) willingness of libraries to impose a charge system that takes into account the many hidden, special-purpose expenses generated by CD-ROM (paper, printer fluid, leases, etc.); (11) equipment that can accommodate simultaneous use by six or more patrons; and (12) incorporation of an orientation to CD and other emerging information technologies into the curricula of all accredited schools of library and information science.

TECHNOLOGICAL BREAKTHROUGHS

As these ingredients are added, three breakthroughs should follow: networking, multimedia, and hypertext.

1. Networking. No new system should be brought into a library environment without the vendor's assurance of provisions for compatibility and extension of connective capacity after the system is installed. One must plan to integrate the system effectively and judiciously for multiple use, rapid response time, and a minimum number of frustrated patrons. The file

PsychLIT	(A)–3, (B)–3, (C)–3 = *9*
Retrospective indexes (e.g., *Wilsondisc* databases)	(A)–3, (B)–3, (C)–3 = *9*
Ency. of American History	(A)–3, (B)–3, (C)–3 = *9*
Legal Citations Index	(A)–3, (B)–3, (C)–3 = *9*
Library-of-the-Month Club	(A)–3, (B)–3, (C)–3 = *9*
Compact Disclosure	(A)–3, (B)–3, (C)–3 = *9*
Microsoft Bookshelf (includes *Chicago Manual of Style, Roget's Thesaurus, American Heritage Dictionary, Bartlett's Familiar Quotations, National Five-Digit ZIP Code and Post Office Directory, World Almanac and Book of Facts*)	(A)–3, (B)–3, (C)–3 = *9*
Speakers' Guide, an integrated "omnium-gatherum" of aphorisms, jokes, and literary quotations	(A)–2, (B)–3, (C)–3 = *8*
Myth and Magic Encyclopedia	(A)–2, (B)–3, (C)–3 = *8*
Science Citation Index with bibliographic coupling feature	(A)–3, (B)–2, (C)–3 = *8*
Guide to Classical Music Recordings	(A)–3, (B)–2, (C)–3 = *8*
Self-help health guides	(A)–1, (B)–3, (C)–3 = *7*
Encyclopedia of Philosophy	(A)–1, (B)–3, (C)–3 = *7*
Auto Manual Miscellany	(A)–2, (B)–2, (C)–3 = *7*
Sci & Tech Encyclopedia	(A)–2, (B)–2, (C)–3 = *7*
Directory of librarians and information specialists	(A)–2, (B)–2, (C)–3 = *7*
Official Airlines Guide	(A)–2, (B)–2, (C)–2 = *6*
News Bank	(A)–2, (B)–1, (C)–3 = *6*
Adult encyclopedia	(A)–2, (B)–1, (C)–3 = *6*
Boy/Girl Scout, Camp Fire Girls, 4-H Club composite handbook	(A)–1, (B)–3, (C)–2 = *6*
Current Stock Quotations	(A)–3, (B)–1, (C)–2 = *6*
Electronic Yellow Pages	(A)–1, (B)–2, (C)–3 = *6*
Directory of Lobbyists	(A)–3, (B)–1, (C)–2 = *6*
Boy/Girl Scout First Aid Handbook	(A)–1, (B)–3, (C)–1 = *5*
High Rollers' Nightlife Guide to Baltimore, New Orleans, Atlanta and Miami	(A)–1, (B)–1, (C)–2 = *4*
Regional weather advisory service	(A)–1, (B)–1, (C)–1 = *3*

Fig. 2 Real and Hypothetical CD-ROM Candidates.

server should be programmed to queue up requests.

2. Multimedia. There is a movement underway to transform CD-ROM presentations into an interactive audiovisual form akin but superior to television. Publishers have long had the hardware to accomplish this but have been stymied by (a) educators who have not designated the subject areas that have high priority and could best benefit from this approach to instruction; (b) high production costs imposed by the difficulty of devising ways in which to mix audio, video, and text in a structured mosaic that can be broken into pieces, reconstructed, and recalled by users; and (c) the "low-tech" environment of most libraries and schools.

3. Hypertext. The most significant innovation, a dramatic departure from the standard linear approach to reading, hypertext permits linkages between different but related ideas. Users can play creative games with the syndetic structure of the original reference that brings them—via light pen, icon, mouse, or megaphone symbol—combinations of recorded knowledge on people, events, and topics; excerpts from novels and poetry; reproductions of fine and applied arts; demonstrations of scientific principles; music; readings; and variant and derivative works. Apple's *Hypercard* is a first step in this direction. Diverse experiments with hypertext are under way: one dealing with nineteenth-century American history, offers immediate call-up and juxtaposition of photos of Gold Rush towns and the Civil War, information on the Lewis and Clark expedition, voices of Ulysses S. Grant and Lincoln, music of the period, and Matthew Brady photographs of the Civil War. Another involves foreign-language instruction that melds the targeted nation's literature, history, art, music, landscape, and historical figures, while providing an organized sequence of language instruction. Predictably, the task of organizing, defining, and calibrating relationships between subjects and appropriate access routes is difficult and can only be accomplished by thinking, uncommonly cultured, highly skilled editors. Because editors are chronically underpaid, the talent shortfall that has developed must be addressed before hypertext can achieve its potential. Because it allows instant access to an assemblage of audio, visual, and textual presentations specified by the user, hypertext is indeed different. There are no long menus, just a mouse or light pen. The interaction is so personalized that any barrier between child and recorded knowledge seems to evaporate. This is because the system does not anticipate, coach, or force the users' moves but leaves them free to move extemporaneously. Again, as with multimedia, the dearth of current technology (e.g, lack of superior color, high-resolution monitors) in the schools could be an inhibitant to progress.

RESEARCHABLE AREAS

It is suggested that research focus on (1) analysis and classification of content to improve retrieval capacity, (2) human engineering problems raised by CD-ROM, and (3) a hypertext that relates to the different ways students learn as they grow older.

The philosophical dimensions should be most intriguing to the researcher. To what extent does reality generate thought? Or is reality a product of the mind? Is there a direct correspondence between reality and the mind's internal representation? Should expert systems be confined to communications enhancement and kept at a safe distance from manipulation of human thought? We periodically relearn at our peril that the attribution of human intelligence to machines is delusive, even dangerous. Human rationality is only skin deep. Below lies a secret, restive region, where the wild things are, beyond the reach of machines and numbers—the human psyche. After two decades of experimentation, artificial intelligence is still barely able to compete with the mind of a four-year-old child. It is indefensible to assume morphological correlates between the neuro-complexity of the human brain and even the most sophisticated computer.

The point is to place foremost the mission of the librarian—to transmit, interpret, and advocate recorded knowledge. Whether CD-ROM technology can be a major part of this remains to be seen.

BIBLIOGRAPHY

Arnold, Stephen. "Electronic Information on CD—a Product or a Service?" *Online* 11, no.6:56–60 (Nov. 1987).

Carrier, Harold D., and William A. Wallace. *An Epistemological View of Decision Aid Technology with Emphasis on Expert Systems.* Decision Sciences and Engineering Systems, Technical Report no.37-87-140. Troy, N.Y.: Rensselaer Polytechnic Institute, 1987.

Crane, Nancy, and Tamara Durfee. "Entering Uncharted Territory: Putting CD-ROM in Place." *Wilson Library Bulletin* 62, no.4: 28–30 (Dec. 1987).

DeBuse, Raymond. "So That's a Book. . .Advancing Technology and the Library." *Information Technology and Libraries* 7, no.1: 7–18 (Mar. 1988).

Garfield, Eugene. "Current Comments." *Current Contents/Social & Behavioral Sciences* 22:3–9 (May 30, 1988).

Herther, Nancy K. "CDROM and Information Dissemination: An Update." *Online* 11, no.2:56–64 (Mar. 1987).

Lowry, Charles B. "A Convergence of Technologies: How Will Libraries Adapt?" *Library Administration & Management* 2, no.2: 77–84 (Mar. 1988).

Newhard, Robert. "Coverting Information into Knowledge: The Promise of CD-ROM." *Wilson Library Bulletin* 62, no.4:36–38 (Dec. 1987).

Pearce, Karla J. "CD-ROM: Caveat Emptor." *Library Journal* 113, no.2:37–38 (Feb. 1, 1988).

Peters, Charles. "Databases on CD-ROM: Comparative Factors for Purchase." *The Electronic Library* 5, no.3:154–59 (June 1987).

Quint, Barbara. "How is CD-ROM Disappointing? Let Me Count the Ways." *Wilson Library Bulletin* 62, no.4:32–34,104 (Dec. 1987).

Reese, Jean, and Ramona Steffey. "ERIC on CDROM: A Comparison of DIALOG On-Disc, OCLC's Search CD450 and SilverPlatter." *Online* 11, no.5:42–54 (Sept. 1987).

Rietdyk, Ron J. "Creation and Distribution of CD-ROM Databases for the Library Reference Desk." *Journal of the American Society for Information Science* 39, no.1:58–62 (Jan. 1988).

Schwerin, Julie B. "CD-ROM: Potential Markets for Information." *Journal of the American Society for Information Science* 39, no.1: 54–57 (Jan. 1988).

Stewart, Linda. "Picking CD-ROMs for Public Use." *American Libraries* 18, no.9:738–40 (Oct. 1987).

Valk, Anton. "Compact Disc Interactive." *Electronic and Optical Publishing Review* 7, no.2:64–68 (June 1987).

Vandergrift, Kay E., and others. "CD-ROM: An Emerging Technology. Part 2: Planning & Management Strategies." *School Library Journal* 33, no.11:22–25 (Aug. 1987).

Watson, Paula D. "Cost to Libraries of the Optical Information Revolution." *Online* 12, no.1:45–50 (Jan. 1988).

Zoellick, Bill. "CD-ROM Software Architecture to Promote Interchangeability." *Journal of the American Society for Information Science* 39, no.1:47–53 (Jan. 1988). ■■

Shanghai Jiao Tong University Pao Sui-Loong Library: Toward an Integrated Online System

Sharon Chien Lin

This paper reports on China's first high-tech library. Early activities in automation and difficulties encountered in computerization of library operations are summarized, various systems now in service are described, and new developments utilizing CD-ROM technology are discussed.

Shanghai Jiao Tong University, founded in 1896, is one of the oldest and best engineering schools in China. For a long time before 1949, it was known as "the M.I.T. of China." The old library building was built in 1919, when it had a collection of more than 31,000 volumes in Chinese and at least 3,600 volumes in English. By 1985 the collection had grown to 1,275,382 volumes, including 143,166 bound periodicals. Library services had also expanded considerably. The new Pao Sui-Loong Library was built in 1985 with a donation of ten million dollars from Pao Yue-kong, an alumnus of the university, in his father's name. At present it is one of the largest research libraries in China, and the completion of the new building was a milestone not only in the library's history but in the development of Chinese libraries as well. As the first computerized library in China, it marks the entrance into an automated environment from the traditional manual operation. It is to this aspect that the present paper will devote its effort. A description of the new library building, its organization, collection, and various services is presented in a separate paper.[1]

Though initial interest in computers began in 1956, when the Chinese Academy of Sciences formed the Institute of Computing Technology in Beijing, it was not until almost two decades later that this interest began to gain momentum. In 1974 the 748 Project, a national plan under the auspices of the State Planning Commission, gave high priority to the computerization of Chinese script processing. In a nation whose bibliographic records are mainly in the Chinese language, this is undoubtedly an important prerequisite for the eventual library automation and computerized information retrieval. At the 1978 National Science Conference, Fang I, vice-premier for Science and Technology, asserted that "automation of key research centers" and "national networking" should be the top priority goals in China's modernization plan for science and technology. In response, many organizations that deliver information began to study the application of computers.[2] Generally, during the early stage, the utilization of computers in library information services in China was focused primarily on information retrieval services that mounted foreign (imported) bibliographic databases on Chinese-made computers.

At Shanghai Jiao Tong University (SJTU), the first effort toward moderniza-

Sharon Chien Lin is with the Serials Department, University Libraries, State University of New York at Buffalo.

tion began in 1978 when a delegation including librarians was sent to the United States to commence cultural exchange activities. Between 1983 to 1985, in preparation for building the first automated library for Chinese higher education, three different groups of university personnel came to the U.S. to study modern techniques and operations at over twenty large university libraries and major library and information networks.[3] These groups included architects and engineers, as well as librarians. The focus of these visits was on two themes: one was on library buildings and the layout therein, and the other dealt with the automation of library activities, management, modern equipment, and new library technology. Besides these study tours, two small groups of library administrators and staff participated in three-month internships at Brandeis University and the University of Pittsburgh. Those at Pittsburgh also studied to receive an advanced degree in library science.

Meanwhile, back at SJTU, foreign library delegations and experts visited the library and introduced modern concepts of librarianship to help in the development of library automation activities.

EARLY DIFFICULTIES

Pao Sui-Loong Library realized from the beginning that, due to the complexity of Chinese characters and the need for advanced scientific and technical information from foreign countries (especially periodicals published during the Cultural Revolution), initial automation efforts should begin with Western-language materials; however, three major difficulties were encountered in the early stages:

Shortage of Equipment and Limitations of Computer Storage Space. The library did not have its own computer when research and experiments on automation first started. Software designs for various systems were developed on computers in the university's computing center. The limited capacity of the computer storage often restricted the library from building databases with a large number of information records. This was a relatively simple problem, however, since money could solve it fairly quickly. For example, a $250 million loan from the World Bank in the mid-1980s for university automation projects provided SJTU and many other universities with powerful computers by IBM, Honeywell, DEC, Data General, etc.[4] SJTU was also one of the four key universities that benefited from a donation of 100 IBM 5500 Chinese-language multifunction workstations.[5]

Shortage of Trained Personnel. People who were knowledgeable in computers as well as in library work did exist but were not easily available. The SJTU Library therefore created an automation research unit that worked jointly with an assigned staff in the information network unit of the university's computing center. The new team was in charge of planning the library automation research work. Related library personnel participated in the design of software for each specific library system, e.g., circulation. They prepared the specifications for functions required of that system while the staff at the information network unit wrote the programs. This kind of cooperation resulted in the development of several successful software applications.

Lack of Standardization. Although there were several national standards for automation (e.g., the codes for Chinese character sets for information interchange and record format for bibliographic exchange on magnetic tapes were available) cataloging rules for various types of publications were not issued until later. The *Descriptive Cataloging Rules for Western-Language Materials*, published by the China Society of Library Science in 1985, greatly facilitated the building of databases in Western languages. This has been the primary area of computerization efforts in library information services so far. Major projects for Chinese databases have begun only in the last few years, when the difficulties of Chinese character processing were essentially solved.[6] There is not yet an official MARC format for libraries of mainland China to follow.[7] With the eventual cooperative cataloging, resource sharing, and networking in mind, the absence of a Chinese MARC format and other relevant standards hindered the design of software.

In spite of these unfavorable circumstances, the SJTU Library did forge ahead and put forth a number of automated systems in cooperation with the university computing center and sometimes with other units on campus. Some of these systems, (e.g., acquisition and cataloging for Western-language monographs, serials control), though they passed experimental tests, have never been put into service. The following discussion covers only operative systems.[8]

SYSTEMS IN OPERATION

Shanghai Jiao Tong University Circulation System (SJTUCS). A circulation system was first developed on a WANG MVP-2200 minicomputer, but due to the lack of equipment at the time of design, information about the borrower and the book to be charged out could not be automatically keyed in. Furthermore, there was no terminal available at the circulation desk. Work was therefore stalled.

In late 1985 a new system known as SJTUCS was developed on a HP-3000/39 minicomputer using COBOL/77 and the IMAGE/3000 database management system.[9] The capacity of the mainframe computer was 1 MG, 65 MB for the hard disk, and 67 MB for each cassette tape unit. One central processing unit, four terminals equipped with light pens, two local printers, and one HP-150 personal computer with a light pen were included. For software, there was a MPE/V multiuser operating system, IMAGE/3000 network management system, VPLUS/3000 data input and screen format management system, query/3000 inquiry system, etc. This realized the first phase of an integrated online system for the Pao Sui-Loong Library; it was probably the first full-scale circulation system as well as the first computerized library operation in China.

When SJTUCS was implemented on May 15, 1986, the database had only the information about the readers, with no bibliographic records. When a book were being charged out, a pair of barcode strips bearing the same numbers were placed on the book and on the old charge slip in the book pocket. The light pen was then run over the barcodes on the borrower's card and on the book. The system would automatically examine the eligibility and the privileges of the borrower.

At the end of the day, based on the charge slip, brief bibliographic information, including the call number, was fed into the computer, and the light pen was then run over the barcode on the charge slip. The bibliographic data was thus gradually established along with the borrowing records.

This has worked amazingly well. It eliminates much of the nearly impossible task of preparing a complete bibliographic database before the system could go into operation. It also saved computer space from being clustered with records for books that may never be charged out.

This system has two backup devices supported by the IMAGE/3000 database management system, and a HP-150 microcomputer to safeguard any breakdown of the mainframe computer. It has more than thirty functions in two major categories: the first includes borrowing, discharge, renewal, hold, fine, reader's file maintenance, and various levels of searching by readers and public and technical services personnel; the second category maintains the bibliographic file, prints announcements and statistics, and offers backup capability.

SJ-II Book Security System.[10] In use since March 2, 1981, this device has detectors on either side of the exit. Magnetic needles are placed in books at an inconspicuous place. When a book passes the electronic gate without being charged out and desensitized, the detectors trigger the operator, causing a red light to flash, the alarm to ring, and the exit gate to lock. This system is now commercially available.

SJTU Online Information Retrieval System. The library began in early 1979 to develop this system on a Wang MVP-2000 minicomputer. It passed test in April 1981.

A terminal equipped with a printer and a 10 MB disk with a maximum capacity for 10,000 entries was installed in the library to test the system. The database of more than 6,000 entries contains the *Government Reports Announcements (GRA)* (i.e., PB, AD, NASA, USDOE reports) since 1976 and the *Journal of Shanghai Jiao Tong University* from 1978 on. Information can be accessed by related subjects, author, corporate body, subject/author, and report number. It can generate cards, compile subject headings and new book announcements, and perform automatic check-in. The easy-to-learn system has been used as a laboratory tool for a course in teaching scientific and technical information retrieval offered by the library. Due to limited disk storage space and the restricted library access time on the computer, the system has so far been available only for training purposes. The existing database will constitute a part of the integrated system.

Dialog International Online Retrieval System. With the approval of the Ministry of Education, SJTU signed a contract with Lockheed's International Online Dialog Information Services. Through the telex terminal installed in the library, users may have direct access to more than eighty million records in its more than 190 databases in all major disciplines via the international communication satellite Tymnet and Telenet lines. Since December 1, 1983, the library has been offering services to users on campus as well as to those from other institutions. It has also introduced the technique of computer searching and conducted training sessions for the faculty and personnel in the field of science and technology.

Aside from the systems mentioned above, the library has also been involved in a few joint computerized projects, such as in creating databases and networking systems on a local or national basis.[11]

NEW DEVELOPMENTS

In June 1987 Pao Sui-Loong Library imported a laser audiovisual system and a complete CD–ROM laser controlled disc system.[12] In addition, a number of databases for bibliographic records and information retrieval on CD–ROM, e.g., Bibliofile, ANYBOOK, NTIS, LISA, Medline, etc., were also acquired.[13] More importantly, the library imported in early 1988 an HP-3000/930 super minicomputer with 32 MB main memory. Its peripheral accessory includes two tape subsystems and six fixed disc drives with a capacity of 571 MB each.[14] (Figure 1 shows the configuration diagram of the HP-3000/930.) More terminals and workstations that can display both

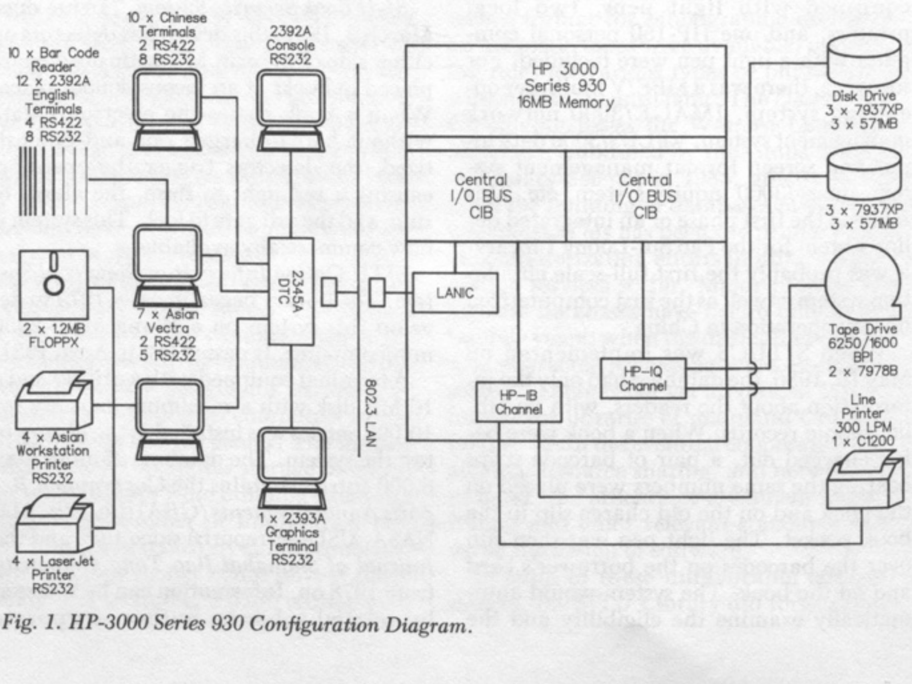


Fig. 1. HP-3000 Series 930 Configuration Diagram.

English and Chinese were installed. The SJTU circulation system and the MINISIS software were both transferred to the new computer system in March 1988, thereby expanding the functions of the SJTUCS system.[15]

The SJTUCS automated circulation system described above brought SJTU into the first phase of an integrated online system. With the entry of many databases on CD-ROM (providing large amounts of machine-readable bibliographic records), the larger-capacity computer, and other related accessories, Pao Sui-Loong Library entered into the second phase of automation in early 1988. The integrated online system, being developed on the HP-3000/930, will encompass subsystems for acquisitions, cataloging, circulation, serials control, online public catalog, and library management when completed in the early 1990s.[16]

Taking advantage of the compatibility of the HP-3000/930 with the IBM system, an IBM PC/XT is interfaced with the HP-3000 series along with two CD-ROM drives to transfer the disc data to the HP-3000 database efficiently. Since 1987 cataloging for Western-language books has been done online using the bibliographic records found on Bibliofile,[17] which consists of the entire LC MARC database stored on three CD-ROMs with a monthly update that includes a cumulative replacement disc. The database will make it possible for SJTU to start its retrospective conversion project in the near future, and this system also prints catalog cards and labels. When online copy/original cataloging is completed, cards are produced and the newly modified/created bibliographic records become a part of SJTU's database. Bibliofile also offers services to eight other institutions in Shanghai, providing bibliographic records, catalog cards, etc.

For acquisitions, SJTU is now using ANYBOOK to process its acquisitions activities.[18] Developed in July 1985 by Ingram Book Company and the Library Corporation of the United States, ANYBOOK is a union list of about 1,500,000 bibliographic records on CD-ROM for computerized selection and ordering of books. In early 1986 Library Corporation merged the ANYBOOK and Bibliofile CD-ROMs. Books received in the acquisitions department are searched on Bibliofile by ISBN number (provided in the ANYBOOK database) for bibliographic records. These records are automatically downloaded into the database of the library and eventually become a part of the integrated system.

At present the library is developing its integrated system on the MINISIS, using Bibliofile and ANYBOOK for its sources of machine-readable data.[19] Furthermore, Pao Sui-Loong Library is also expanding its information retrieval services through ALLBASE software, using LISA, NTIS, and Medline databases on CD-ROM.[20]

CONCLUDING REMARKS

Due to the lack of equipment available to the library and the limited storage capacity in the early stage, only several of the systems developed and tested were actually put into operation. Library automation at SJTU moved very slowly until early 1987. Nonetheless, these efforts did train a group of personnel for software design and system operation. Some of the systems developed had served as first steps for information input and retrieval. Databases were built and expanded when the larger-capacity HP-3000/930 computer became available. Other systems, notably the SJTU circulation system and the Dialog retrieval system, have been in operation and constitute an integral part of library service. Pao Sui-Loong Library is therefore well on the way toward complete automation.

Moreover, rapid advances in information and computer technologies, and the deployment of microcomputers with more sophisticated capabilities have made it possible to store even greater amounts of information in smaller and cheaper equipment. The optical technologies, such as CD-ROM, offer easy access to larger bodies of data without the anxiety of connect time cost. In addition, various standards essential to the computerization of information are gradually being established. It appears reasonable then to expect that the faculty and graduate students of SJTU will enjoy the luxury of searching for library holdings on microcomputers located in their own departments during the next decade.

ACKNOWLEDGMENTS

The author wishes to thank SJTU for the warm hospitality extended to her during her stays on campus in 1985 and 1987. She is particularly grateful to Yang Zongying, deputy director, and Ji Luqin, assistant to the director, for generously sharing their time and information with her.

This research was supported in part by two grants from the New York State/United University Professions, Professional Development and Quality of Working Life Committee.

REFERENCES AND NOTES

1. Sharon Chien Lin, "The New Library of Shanghai Jiao Tong University," submitted for publication.
2. For a discussion on this subject, see Sharon Chien Lin, "Development of Computer Applications in Chinese Information Centers," *Journal of Information Science* 14:195–203 (1988); and "Library Automation in China," *Information Technology and Libraries* 7:230–42 (Sept. 1988).
3. Luqin Ji, "Developing International Contacts and Promoting Academic Exchange" (in Chinese), *Bulletin of Shanghai Jiao Tong University* (hereafter referred to as BSJTU) 4:35–36 (1986).
4. John H. Maier, "Library Automation in the People's Republic of China," *Information Technology and Libraries* 5:107 (June 1986).
5. Emily Gallop Fayen and Vladimir Slamecka, "China Modernizes Scientific Communication System," *Information Processing & Management* 22:73 (1986).
6. Lin, "Development of Computer Applications in Chinese Information Centers," p.197–98.
7. It is noted, however, that the Chinese MARC format in Taiwan was published by the National Central Library in 1981. This is widely used by the academic and research libraries in Taiwan. Cf. *Chinese MARC Format*, 2d ed. (Taipei, Taiwan: National Central Library, 1984).
8. Unless otherwise mentioned, information for these systems came primarily from two sources: my personal visits to the university library in 1985 and 1987 and He Dayong, "A Brief Report on the Computer Applications at the Shanghai Jiao Tong University," *Daxue Tushuguan Tongxun* 4:46–48 (1984).
9. Zongying Yang and Shumin Shi, "Modern Facilities in Pao Sui-Loong Library" (in Chinese), *Bulletin of the China Society of Library Science* 2:13 (1986); Shumin Shi, "The Multiuser Circulation System in Pao Sui-Loong Library" (in Chinese), *BSJTU* 4:18–20 (1986).
10. Shanghai Jiao Tong University, Research Group no. 34 and the University Library, "The Development of the SJ-II Book Security System" (in Chinese), *Library and Information Service* 6:28–29 (1981).
11. Lin, "Development of Computer Applications in Chinese Information Centers," p.198–99,201; "Library Automation in China," p.235–37.
12. Zongying Yang and Qiaoying Zheng, "CD-ROM Technology Enters Pao Sui-Loong Library" (in Chinese), *Xiandai Tushu Qingbao Jishu* 2:1–5 (1988).
13. Ibid.
14. Zongying Yang and Xiaoling Shi, "HP-3000/930 and the Integrated Online System at the Pao Sui-Loong Library" (in Chinese), *Electronic International* 82:31–33 (1987). Zongying Yang and Shumin Shi, "A Multiuser and Online Library Circulation System with Barcode Reader," *Proceedings of International Symposium on New Technologies and Applications in Libraries*, Xian, Shaanxi, China, Sept. 8–11, 1988.
15. Yang, "HP-3000/930 and the Integrated Online System."
16. Ibid.
17. Yang, "CD-ROM Technology."
18. Ibid.
19. Yang. "HP-3000/930 and the Integrated Online System."
20. Ibid.

Communications

Imaginative Terminal Design for Online Public Access Catalogs

Norman D. Stevens

The rapid proliferation of the ubiquitous online public access catalog (OPAC) throughout American academic libraries in the late 1980s has brought with it a host of new and unanticipated challenges to the imaginations of librarians, as they struggle to design, select, implement, promote, and even justify their efforts and expenditures toward replacing the ever-popular card catalog with the OPAC, which is only the latest in a long line of whiz-bang devices ostensibly designed to save staff time and effort and to make collections more readily accessible to users. Or so we say, and so we hope. The challenges of the design and/or selection of an appropriate system at the least possible cost and the trials and tribulations of the implementation of an OPAC have been overabundantly dealt with in the literature. There is little new to be said on that subject. That is not to say, however, that new challenges, some of them not yet fully understood and explored, do not lie before us and that there are not some significant aspects of OPACs that have not yet even been considered in the literature.

Many of our past challenges have been the usual mundane, behind-the-scene challenges that have been, in most respects, transparent to the user and sometimes, but not always, even to the college or university administration. As OPACs increasingly become a reality, libraries are often faced with new and unforeseen difficulties as they seek to tread their way carefully through uncharted waters.

Norman D. Stevens is Director, the Molesworth Institute, Storrs, Connecticut.

One such challenge, which has received considerable mention in the library press but little in the way of analytical review to provide guidance, has to do with the naming of OPACs. The general idea behind this ritual naming process seems to be that since machines are semihuman, they need to be given an appropriate name, even though there is no recorded instance of a card catalog having been named. The naming process varies, but in most cases it appears either to be approached casually as an administrative decision—perhaps the only decision having to do with automation that can be entrusted to a library administrator—or as a contest that may even have a prize attached to it. No matter what the process, the results to date have been distressingly similar and unimaginative. Most academic libraries have settled on a narrow range of options, most often opting for a name that utilizes some local name element, such as the university's athletic mascot or the name of the library, coupled with the term *cat*. In other cases names have been chosen largely because they have some local significance, such as affiliation with a long-departed institutional or library administrator or benefactor, revered in memory to a greater degree than she or he ever was in real life. Above all, cute and catchy names are in, as though in some fashion such names will enhance the quality of the product.

Another even more significant area of choice in relationship to the implementation of OPACs, which is only now becoming apparent to astute observers, is the design of the terminals to be used in those devices. To date most academic libraries have been unaware of the importance of this issue and have foolishly been content simply to settle for the standard terminals supplied or recommended by the system supplier. The color of the display and sometimes of the terminal itself have been the only apparent areas of concern and choice to date. Those simple choices pale in comparison with the true issues of terminal selection that most academic librarians have not yet even begun to think about.

There are several excellent reasons why much more careful attention deserves to be paid to the important question of terminal design. In most academic libraries computer terminals have multiplied to such an extent, just as they have on the entire campus, that one terminal is indistinguishable from another. The card catalog had a unique appearance unlike anything else in the library or on the campus that gave it a status of its own. The OPAC deserves no less an honor. As libraries place OPAC terminals, which are an essential element of access to the vast wealth of information to be found in their collections, in public places, those terminals must, in some fashion, stand out in a way that clearly sets them apart from their poor country cousin ordinary terminals found elsewhere in the library and on campus. The location of those OPAC terminals in public areas within the academic library also raises critical questions about how well they match the architectural style and decor of the library. Most academic libraries have vast spaces once designed to hold impressive and massive card catalogs so that they would blend nicely into the solemn and majestic decor of the library. The replacement of those carefully designed and placed card catalogs with the typical ugly, modernistic, metal terminals with harsh blinking screens is not a pleasant prospect, especially if a relatively few terminals are given the honor of filling the cavernous space vacated by a card catalog that has been exiled to a basement storage space. Their appearance invariably changes the appearance of the library, most often in a negative fashion. That is, unfortunately, an issue that most librarians and most libraries have simply ignored to date, perhaps electing to believe naively that their patrons will be so overjoyed with this new electronic toy that they will ignore its visual appearance and the ugly change that has taken place. Several recent incidents, including user demonstrations at several major libraries, have now indicated that this issue can no longer be ignored.

As a result of one of those 1968-style demonstrations—coupled with serious expressions of concern from the university administration about the potential desecration of the heart of the university and the pride of the wealthy alumni—the Molesworth Institute was recently commissioned by a prestigious university library, which shall remain nameless here, to consider whether site-specific terminals might be designed for that library. The commission called for a design that would blend appropriately with the architectural style of the library building and the decor of the area in the building where those terminals were to be located.

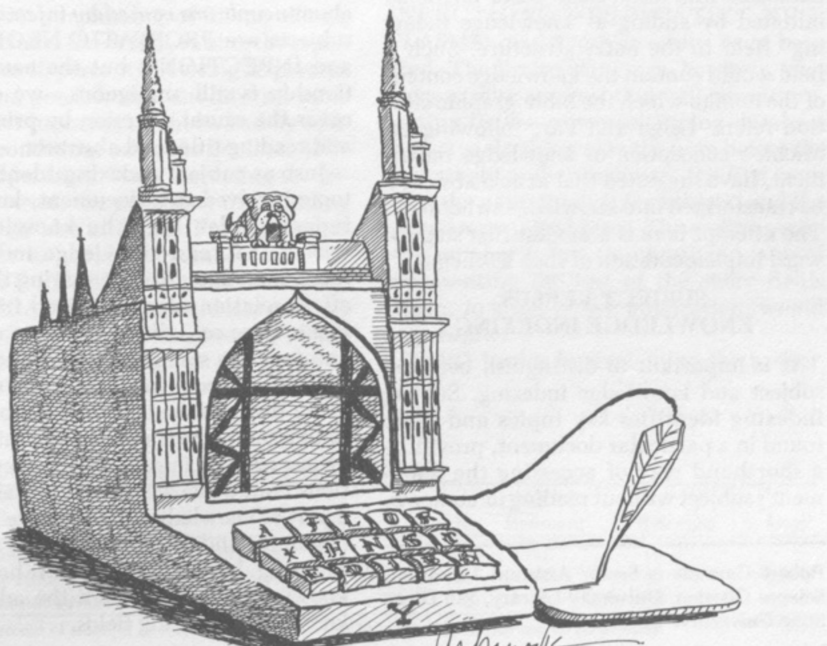
A special, highly skilled design team, which included one of the nation's premier library cartoonists, was selected to work on this unusual problem. After a review of a wide variety of postcards of that library, several site visits, and two blue-sky design sessions, the design team outlined a set of terminal design specifications for review with the library and university administration as well as for presentation to the user community. Those specifications took fully into account the fact that the library building is of restrained gothic design with spires and towers, stained glass windows, and the like and that the terminals were to be located in long, narrow, dimly lit, stone cathedral-like passageways. The medieval nature of the library building was a determining factor in the development of the final design specifications. In addition other important factors, such as the use of the university mascot and other peripheral concerns, were incorporated into the design.

With the unanimous approval of that imaginative concept, the design team then developed a specific terminal design document that incorporated the following elements: a dark oak wooden case with spires and/or towers rising from the top; a display screen with leaded glass panels delineating various fields; a gothic letter keyboard and screen display; a light pen in the shape of a quill pen but with durable plastic feathers; a small bulldog in the shape of a gargoyle perched on the terminal; and a blue *Y* in a white background discreetly placed on the side of the terminal. The only design feature that was rejected in a final review by the community was the wrought iron candelabras on both sides of the terminal that would have provided additional lighting in the dim, religious atmosphere that exists in the OPAC area. The fire marshall, as well as the continued maintenance costs for candles and matches, scorched that brilliant idea. In its place the user will be offered the option of hitting a special "candle" key that will produce a row of flickering orange candles around the perimeter of the display screen. The terminals have also been designed to be placed on long, dark, slanted, oak tables—with precut initials and other graffiti carved into the surface—at a height that will allow matching dark oak wooden stools to be supplied for users to sit at as they use the terminals. The imaginative design, which is shown in black and white in the accompanying illustration, has been enthusiastically accepted by all involved. The terminals are now in production and will be installed shortly. Even though each of these unique terminals costs five times as much as an ordinary run-of-the-mill OPAC terminal, this has been adjudged to be a creative and acceptable solution to the particular problems generated by the installation of an OPAC in this setting.

Emboldened by the success and profit of that design effort, the Molesworth Institute is now prepared to offer its terminal design services to other libraries faced with similar problems. Inquiries, preferably accompanied by postcards that indicate what the library looks like, along with the usual $1,000 nonrefundable design retainer fee, may be addressed to the Director, the Molesworth Institute, 143 Hanks Hill, Storrs, CT 06268.

ACKNOWLEDGEMENT

This article was written after a pleasant luncheon discussion with several administrators of the Yale University Library that touched upon the kinds of issues dealt with here. While some wild ideas for terminal design were bandied about in that conversation, the content of this article reflects only the views of the author and not necessarily those of the administration of the Yale University Library. The design of an equally improbable OPAC terminal for the Homer Babbidge Library at the University of Connecticut is on the drawing board. ■■



Database to Knowledge Base

Robert Carande

Present-day commercial databases may become the foundation for future commercial knowledge bases. The addition of knowledge-index fields to citations permits the searcher to identify references to items that contain the knowledge sought. A small computer program, ZENO, is used to demonstrate the conceptual feasibility of searching knowledge-indexed records from the Medline database.

Recently suggestions have been made for developing national knowledge bases with universal access.[1] Such knowledge bases would not be subject to any particular "goal" but would be large and robust enough to support multiple users and knowledge requests simultaneously.

This paper suggests that the current bibliographic databases already mounted on powerful vendor software systems such as Dialog and BRS might serve as the basis for universally accessible and multiple goal-tolerant knowledge bases in the future. Transformation of bibliographic databases into knowledge bases might be initiated by adding a "knowledge indexing" field to the entry structure. Such a field would contain the knowledge content of the item to which the bibliographic citation refers. Leigh and Paz, following D. Michie's conception of knowledge refinement, have suggested that article abstracts be transformed into knowledge structures.[2] The attempt here is a modest first step toward implementation of such a scheme.

SUBJECT VERSUS KNOWLEDGE INDEXING

It is important to distinguish between subject and knowledge indexing. Subject indexing identifies key topics and ideas found in a particular document, providing a shorthand way of accessing the document's subject without reading its contents.

Robert Carande is Senior Assistant Librarian, Science Division, University Library, San Diego State University.

Typically, the terms used to index the subject content are derived from an authority list, standardizing both points of access and future indexing practice. Authority lists such as *Library of Congress Subject Headings (LCSH)*, *Medical Subject Headings (MeSH)*, and the *ERIC Thesaurus* are well known.

Sophisticated as these thesauri are, subject organization of documents in card catalogs, periodical indexes, and printed bibliographies remained accessible via simple alphabetical ordering until the advent of machine-readable data files. Thereafter, Boolean logic could be employed on multiple-subject terms to deliver a set of citations corresponding more or less to the searcher's precise subject need.

While subject indexing is good for locating topics, it is less efficient and effective in finding exact cause-effect relationships between topics. For instance, a set of Medline citations constructed by the MeSH headings PROSTATIC NEOPLASMS and INFECTION will always create a set in which the cancer and an infection figure. The precise way they relate to one another, however, is still an open question. We can't know if any given document is about *infections caused by prostatic neoplasms* or about *neoplasms caused by infections*. The subjects are PROSTATIC NEOPLASMS and INFECTIONS, but the causal relationship is still ambiguous—we only discover the causal direction by printing out and reading titles and abstracts.

Just as subject indexing identifies the topic involved in a document, knowledge indexing identifies the knowledge involved. I intend knowledge indexing to mean something like identifying the cause-effect relationships contained in a document. Here *causality* is understood loosely and includes statistically relevant correlations. Knowledge indexing in this context is to be distinguished from the more ambitious attempts to translate the content of an article or book into knowledge representations.[3] But while directional causality is just a part of knowledge, it does serve as a clear beginning and shows the capacity for transforming bibliographic databases into knowledge bases through the addition of knowledge-indexing fields.

INDEX MEDICUS EXAMPLES

If we search Medline (search conducted August 18, 1988) using PROSTATIC NEOPLASMS and EXP INFECTION from 1980 to 1988, we find forty-six "hits." These may be divided into three groups: *(A)* those that treat of infections caused by prostatic cancer, *(B)* those that treat of prostatic cancer caused by infections, and *(C)* those that contain no cause-effect relationship between prostatic cancer and infection. In this latter group fall all studies treating infection as a complication of prostatectomy or other forms of surgical intervention. From a reading of titles and abstracts, the forty-six hits are distributed as follows:

A	*B*	*C*
4	7	35

The distribution is this author's indexing, and there is always the possibility of error. A search for information about prostatic cancers caused by infections achieves a 15.2 percent rate of precision. Of course Medline lets us be more specific by modifying our subject headings with *MeSH* topical subheadings. Thus we can redo the search using PROSTATIC NEOPLASMS and INFECTION/COMPLICATIONS. This would drop our number of hits to eleven, which includes five *B* hits, raising the precision rate to 45 percent but losing 28 percent of the total possible relevant hits (see table 1). Further precision might be had by adding new conditions using the Boolean AND or NOT operators. With increasing set criteria, however, the possibility of excluding relevant hits also increases. For instance, if we sought to avoid all members of *A* by AND NOT PROSTATIC NEOPLASMS/COMPLICATIONS we would lose another relevant *B* hit. This phenomenon of increased precision bought at the price of comprehensiveness, or "recall," has long been known from studies of online searching.[4] Consider the advantages to be had, if, at this very simple level, the cause-effect knowledge structure were indexed in an accompanying searchable field.

To demonstrate the feasibility of such an index a small database program, ZENO, was written to search surrogate Medline entries that have been knowledge indexed. The format of Prolog, the popular AI language, is used to represent cause-effect relationships propositionally. Prostatic-neoplasms-causing-infection is indexed as CAUSE (PROSTATIC NEOPLASMS, INFECTION). Infections-causing-prostatics-neoplasms is indexed as CAUSE (INFECTION, PROSTATIC NEOPLASMS). Medline entries are represented by their accession numbers so that, for example, CAUSE (PROSTATIC NEOPLASMS, INFECTION, "86315561") is how ZENO stores the information that the article entitled "Morphology and Diagnosis of Acquired Bladder Deformities" contains knowledge about PROSTATIC NEOPLASMS causing INFECTIONS. There is nothing sacrosanct about the Prolog structure. The more traditional rendering of first-order predicate calculus (CAUSE, INFECTIONS, PROSTATIC NEOPLASMS) could just as easily have been used. The Prolog structure, however, provides us with a syntax that will permit processing by the powerful Prolog database language. In this syntax the term before the parenthesis is the predicate, the first term after the parenthesis is the subject, and the second term is the object. The ending accession number is just the shorthand way of representing the rest of the entry fields which, in a full-blown application, would be present.

ZENO begins by explaining the proper

Table 1. Three Searches with Increasing Inclusion Criteria

	Hits	No. of Relevant	% Relevant	% Loss*
1. Prostatic Neoplasms and Exp Infection	46	7	15.2	
2. Prostatic Neoplasms and Exp Infection/CO	11	5	45	28
3. Set 2 and not Prostatic Neoplasms	6	4	66	42

*Loss as a percent of number of relevant hits in set 1 lost through subsequent search refinement.

format to enter search requests. It then displays a Cause-List and prompts for a selection. The user inputs the cause term and is then shown an Effect-List and is prompted for input. Having both cause and effect terms, ZENO then moves through its knowledge base in search of a match. For instance, the terms INFECTION and PROSTATIC NEOPLASMS produce a set whose members have as their topic prostate cancer and infections and whose cause-effect knowledge structure is INFECTION *causing* PROSTATIC NEOPLASMS. All other citations have been dropped from the set. The set is comprised of Medline accession numbers that represent the entries so indexed.

ZENO simply shows how Prolog's matching capabilities can be made to render knowledge when the "match" is a syntax that represents a knowledge relationship. The database can be enlarged by adding related cure-knowledge from pertinent entries. If the interest is in infections cured by cephalsporins, the input CURE (CEPHALSPORINS, INFECTION) would construct such a set. Suppose we were interested in superinfections of the intestines caused by cephalsporins used to fight infection. Adding location-knowledge from relevant entries would permit CAUSE (CEPHALSPORINS, INFECTIONS) ANDED to LOCATION (INFECTION, INTESTINE), identifying articles concerning cephalsporins as a cause for intestinal infection. The sophistication of the knowledge representation would depend upon what other predicates and syntax rules could be used in the knowledge-indexing field.

A sufficiently sophisticated (but by no means highly complex) knowledge index would permit the application of inductive reasoning to derive potentially new information. This possibility would be particularly enhanced if knowledge terms were drawn from a hierarchically organized thesaurus. Such logical arguments as categorical and hypothetical syllogisms, modus ponens, and modus tollens could then be applied to achieve search sets.

Medline has been used to illustrate how the simple addition of a knowledge-index field can increase both online search precision and the opportunities for useful application of inductive reasoning. Knowledge indexing can be done on any database. Complete knowledge indexing of a database the size of Medline would be labor intensive, and no doubt small test bases would be needed initially to ferret out problems and determine ultimate feasibility. A thoroughly knowledge-indexed Medline, however, might speed up the rate of medical diagnosis and research. The same could be said of other discipline-related databases including empirical sociology, anthropology, economics, chemistry, materials engineering, etc. In the final analysis, perhaps knowledge bases will not so much succeed databases as evolve out of them.

REFERENCES AND NOTES

1. U.S. Congress, House, Committee on Science and Technology, Task Force on Science Policy, *The Impact of the Information Age on Science*, 99th Cong.,1st sess., 1985, p.210.
2. William Leigh and Noemi Paz, "Conventional and Knowledge-Based Information Retrieval with Prolog," *Information Technology and Libraries* 7,no.1:67–72 (1988).
3. R. P. Futrelle and Linda C. Smith, "Expert Systems Using Semantic Representations of Full Text for Question Answering," in *Proceedings of the 45th ASIS Annual Meeting*, v.19, 1982 (White Plains, N.Y.: Knowledge Industry, 1982), p.103–5.
4. F. W. Lancaster and C. J. Gillespie, "Design and Evaluation of Information Systems," in *Annual Review of Information Science and Technology*, v.5, 1970 (Chicago: Encyclopaedia Britannica, 1970), p.45–47. ■■

The Circulation Control Facility at the Library of Congress

Thomas E. Smith and Virginia J. Vitucci

The Library of Congress (LC) has been lending materials since 1802. Originally loan privileges were limited to the president, vice-president, and members of the United States Congress. During the nineteenth century these privileges were extended to justices of the Supreme Court, the diplomatic corps, government agencies, and the Smithsonian Institution. In 1901 they were further extended to include scholars, both in the United States and abroad.

Lending activities were centralized in 1944 when Librarian of Congress Archibald MacLeish created the Loan Division, which was part of a massive reorganization of the library's reference functions. It was charged with handling outside loans as well as internal LC circulation and with establishing policies and procedures related to such loan activity. After nearly fifty years, this charter still holds.

The number of outside loan requests filled by the division currently exceeds 135,000 per year. Items lent, primarily to members of Congress and their staffs, LC staff, government agencies, and libraries throughout the world, include books, microfilm, and maps from both the general and special collections. The division also handles 90,000 internal circulation transactions each year for charges to LC work units and scholars' research facilities.

THE LOAN DIVISION AND AUTOMATION

Automated circulation systems, among the most common commercial data processing applications for libraries, are particularly valuable to large institutions such as LC, which has as many as 15,000 collection items on outside loan at any given time. Accurately tracking so many loans by hand is difficult and expensive; cutbacks in funding and personnel exacerbate the problem.

An attempt was made in the mid-1970s to develop an in-house automated circulation system that could accommodate the lending patterns at LC, but limited automation resources caused this effort to be abandoned. By the early 1980s the Loan Division's need for automated circulation control had become acute. Rather than rework the original attempt to develop a system, the decision was made to start anew, which allowed developers to take advantage of advances in programming languages and database management technology that had occurred in the intervening years.

To lighten the work load until an in-house system could be developed, the Loan Division investigated commercially available software that could support its interim needs and began time-sharing the Integrated Library System (ILS) in March 1983. ILS was developed at the National Library of Medicine and marketed commercially by Avatar, Inc., until 1985; since then it has been marketed by OCLC under the product name LS2000.

INTERIM USE OF LS2000 TIME-SHARING

The LS2000 service used by LC is housed on a Data General MV-6000 minicomputer owned by OCLC and operated by Sentient Systems in Kensington, Maryland. While this product provides several layers of access security, there was concern about confidentiality of congressional records, particularly loan information stored outside LC's computer environment; consequently, the decision was to use LS2000 for noncongressional charging only. Loans to members of Congress, about 25 percent of all LC charging, would be done manually, and LS2000 would be used to handle the remainder, including interlibrary loans, government accounts, and staff members.

Thomas E. Smith is Head, Circulation Section, Loan Division, and Virginia J. Vitucci is Senior Systems Analyst, Automated Systems Office, at the Library of Congress, Washington, D.C.

THE CIRCULATION CONTROL FACILITY

LS2000 time-sharing eased some of the strain of handling LC's circulation operations, but the need for an online circulation system capable of supporting loan activities as well as interfacing with other automated systems remained. In the spring of 1988 the library's Automated Systems Office (ASO) delivered the LC automated Circulation Control Facility (CCF) for Loan Division use; it was expected to run parallel to LS2000 for three months and fully replace it in November 1988.

CCF handles traditional circulation functions. Information on loan patron accounts as well as on individual items in the collections is created and maintained online. Loan charges, discharges, renewals, and recalls are also handled online, and extensive access authorization checks maintain data security and borrower privacy.

Machine-readable piece identification number (PIN) labels, which are affixed to items as they circulate, increase the accuracy and speed at which loan charges and discharges are performed. Overdue and recall notices, statistical reports, and borrower ID cards are generated via regularly scheduled batch processing.

The collection inventory database used by CCF will, in time, be shared with other collection control–related software applications. As a primary source of information for that database, CCF is key to the development of a comprehensive system for automated collection control, which is an enormous undertaking. To understand the magnitude of this task and the significance of CCF's role, it is important to know something about how the library's physical collection inventory is currently controlled.

COLLECTION CONTROL

The Loan Division is part of Research Services, one of seven departments within LC. This department supports scholarly research, creates bibliographies and other research aids, and provides direct reference services in reading rooms. It is the primary means by which the library fulfills its role as a major archival and research institution serving Congress, libraries, the scholarly community, and the nation at large. The cornerstone upon which these services rest is the immense collection—more than eighty-four million items in various formats (e.g, books, maps, prints and photographs, microfilm).

Physical support of the vast LC collection is also the responsibility of Research Services. This involves proper storage and delivery of collection material as well as the tracking of material lent to all borrowers (see Figure 1).

Special collections such as rare books, music, and manuscripts represent a large portion of the holdings and are maintained by custodial divisions. Responsibility for physical support of the general collections is shared by the Loan and Collections Management divisions.

Given the sheer volume of materials, managing and tracking the use of the general collections are arduous tasks. In the

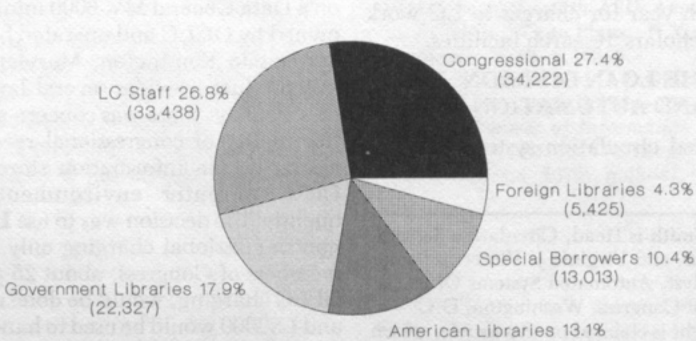


Fig. 1. 1987 Circulation Activity by Patron Category.

past the Loan and Collections Management divisions did not have the kinds of automated tools that permit tracking and monitoring of this collection. CFF is the first step toward support of these inventory control activities. In the Loan Division, it will support traditional circulation activities. In the Collections Management Division, use of bar codes (PIN labels) will make it easier to report and retrieve location and other general collection inventory information once the system is fully implemented.

The basic design objective for CCF has been to develop an automated tool that effectively supports circulation activities but is structured flexibly enough to evolve into a more generalized inventory tool. It is, in effect, the first step toward the online Collection Status Monitor designated by the LC Committee on Automation Planning as a primary automation goal.

TECHNICAL ASPECTS

CCF runs on an IBM 3084 mainframe computer. Most data entry and retrieval are performed online under the Customer Information Control System, a teleprocessing system marketed by IBM. Loan, patron, and item status updates and statistical tallies are performed nightly in batch mode.

CCF was developed by ASO using the DATACOM/DB relational database management system (DBMS) and the IDEAL procedural programming language marketed by Applied Data Research, Inc. of Princeton, New Jersey. The use of a commercial system rather than the attempt to enhance "homegrown" DBMS capabilities represents a significant departure from previous ASO software development. Moreover, it marks the library's first major use of a relational data model and a fourth-generation programming language in a mainframe teleprocessing environment.

The primary advantage of using a commercial DBMS is that developers can focus on application software rather than on the need to create and maintain underlying system software. Problems or updates required by teleprocessing system changes are the responsibility of the vendor, and developers are free to concentrate on software for meeting end-user needs.

A relational DBMS such as DATACOM/DB is well suited for efficient storage and retrieval of inventory data such as those used in CCF. Data are stored with a minimum of duplication. Different applications can have different views of the same data; that is, only data relevant to an application need be accessed, regardless of whatever else is available in a data structure. Efficient data sharing among online applications is feasible. Further, all data fields defined under DATACOM/DB can be searched via commercially available retrieval software such as DATAQUERY or INTELLECT.

CCF also marks the first successful online use of machine-readable LC bibliographic data by a nonbibliographic LC application. A "read only" interface is used to retrieve author and title information dynamically from the library's Multiple Use MARC System (MUMS) database for inclusion in CCF online displays and hard-copy reports. Future LC application releases may well include an interface that will allow users of bibliographic applications to retrieve and display item-specific information created by CCF or similar inventory applications.

IMPLICATIONS FOR THE LIBRARY COMMUNITY

How will CCF affect the work of librarians and information specialists within and outside LC? Initially, the only noticeable changes will be the presence of a PIN label on loaned materials and a slight reformatting of the charge slip. Turnaround time for processing interlibrary loan requests will stay about the same.

Over time, however, there will be several significant changes. Congressional borrowing will be automated for the first time. LC employees, the second largest group of borrowers, as well as government and other loan patrons, will have machine-readable borrower cards. Use of these cards, which will carry the patron's borrower ID in Bar Code 39 format, will enable Loan Division personnel to access patron information quickly and thus further expedite loans to LC patrons who borrow in person.

Finally, and perhaps most importantly,

as the new collection inventory database grows in size, better physical control of the library's general collections will be achieved. Given their size, this will take time, but CCF is the first step toward making such physical collection control a reality. In the future it will be possible to determine online whether LC has a given title and where its available copies are to be found. While information about the loan status of materials will be available, no information concerning names of individual or organizational borrowers will be revealed.

PHASED IMPLEMENTATION

Prior to delivery of the production version of CCF, name and address data for LC staff borrower accounts were loaded from the machine-readable employee file into the CCF borrower database. Similarly, name and address data for borrower accounts that had been entered into LS2000 were converted from a tape provided by OCLC and loaded into the CCF borrower database. Just under 4,900 LS2000 patron accounts were converted to CCF patron accounts, and roughly the same number of CCF accounts were created using LC employee file data.

The first phase of the Loan Division's implementation of CCF involved reviewing the information that had been loaded into the borrower database for accuracy and completeness. This editing was completed in late May, and on June 6, 1988, the Loan Division's charge records unit began handling interlibrary loans via CCF. In mid-July CCF was expanded to include government accounts, its first "in person" borrowers. In August, CCF was further expanded to include LC staff accounts. In 1989, when the 101st Congress convenes, expansion of CCF to include congressional borrowers will complete the implementation cycle.

This incremental approach was adopted for several reasons. First, it was essential that Loan Division staff who would be using CCF regularly be comfortable with the new system. Although these individuals had received considerable online training and had participated in the final round of system testing, there is no real substitute for hands-on experience with the production version of an automated system. Gradual implementation of CCF enabled staff to become familiar with the system with a minimum of stress.

Phased implementation also gave ASO personnel the opportunity to fine-tune online CCF performance with a minimum of user inconvenience. It likewise helped the Loan Division identify needed changes and enhancements and mitigated some of the strain associated with site preparation and changes in hardware configuration. In summary, it was felt that gradual implementation would cause less disruption to work flow and reader service than an "all or nothing" approach.

MANAGING CCF DEVELOPMENT AND IMPLEMENTATION

Making CCF a reality was a joint effort. ASO provided the analytical and technical skills to identify user needs, specify required capabilities, and develop necessary software. The Loan Division created the impetus for development of the system and oversaw user participation in the implementation process. The division's Circulation Section head coordinated user testing, scheduled training for affected personnel, kept staff informed, and planned the phased introduction of CCF into the circulation control work flow.

While ASO and the Loan Division played distinctly different roles in CCF development, their relationship in this effort was complementary. As with most human endeavors, there were occasional misunderstandings and disagreements; however, shared commitment to a single objective and mutual appreciation of the fact that neither could work effectively without the other fostered cooperation and, ultimately, the delivery of a successful, usable system.

CCF AND THE FUTURE: WHAT'S NEXT?

CCF is the first step toward integrated, automated control of the LC collections. Interfaces with administrative systems to share name, address, and personnel information as well as interfaces that enable the library's bibliographic search systems to display item status information are among the next steps. Given the complexity of LC's computer applications and the number of collection items for which inventory information must be captured, much remains to be done. Yet the successful implementation of CCF bodes well for what lies ahead.

APPENDIX A. BORROWER CATEGORIES AT THE LIBRARY OF CONGRESS

- Members of Congress
- Congressional staff
- Former members of Congress
- Statutory borrowers (president, cabinet members, Supreme Court justices)
- Libraries of the U.S. federal judiciary system
- Diplomatic corps
- Interlibrary loan—American libraries
- Interlibrary loan—government libraries inside and outside the Washington, D.C., metropolitan area
- Interlibrary loan—foreign libraries
- Research facilities (scholars using materials at LC)
- LC work units
- Materials on exhibit
- Contributing editors to *Handbook of Latin American Studies*
- LC staff
- Retired LC staff
- Special accounts
- Temporary accounts ■■

Automated Control of Subject Headings at the OSU Libraries

Sally A. Rogers

As more libraries acquire online catalogs, the interest and attention given to automated authority control continues to increase. Baer and Johnson recently did a survey of the authority control literature produced since 1974. They found that most of it (some seventy references) did not support the idea that advanced computer-searching capabilities are making authority control less necessary in online catalogs. In addition, their survey of the uses of authority control in American college and university libraries inspired "a number of comments to the effect that maintaining authority control online was just as time-consuming as maintaining authority control manually."[1] If automation has not made authority control unnecessary or even less time-consuming, what *has* it done in this area? Baer and Johnson express the hope that automation has enabled libraries to provide *better* authority control even if it has not allowed them to save time.[2] Such has been the case at the Ohio State University Libraries (OSUL). What follows is a description of how automation has recently been used to improve authority control of subject headings in the online catalog.

Since December 1981, the OSUL online catalog (LCS) has included an authority file, which also serves as an index to the bibliographic records in the catalog[3] and consists of four headings indexes that provide access by author, subject, series, and uniform title. The Library of Congress name and subject authority tapes have been used to add headings and cross-references to the LCS authority file and to update the data for consistency with current LC practice.[4] (Markey and Vizine-Goetz's research report *Characteristics of Subject Authority Records in the Machine-Readable Library of Congress Subject Headings*[5] should prove invaluable to any library planning to use the LC subject tape.)

A tape containing OCLC records created or edited by OSUL is added to LCS each week by batch processing. At that time, headings on the records being added to the catalog are machine-matched against headings already in LCS—headings tagged 600, 610, 650, etc. are matched against the LCS subject index; those tagged 100, 700, etc. are matched against the name index; and so forth. Certain types of headings are reported for human review. If errors are found, LCS is corrected manually.

Regular review of these reported headings indicated that some kinds of errors were repeatedly occurring. Since few of the headings on OCLC records used for copy cataloging are checked before the records are added to LCS, the same outdated headings and subdivisions kept reappearing in

---

Sally A. Rogers is Catalog Planning Librarian, Ohio State University Libraries, Columbus.

the subject index. If an incoming heading exactly matched a cross-reference in LCS, it was automatically "flipped" to the correct form; but there was no provision for automatically correcting only *part* of a heading when it was added to the catalog.

OSUL staff responsible for heading maintenance kept track of the errors that occurred most frequently and wrote specifications for a computer program to correct as many of them as possible automatically. Several lists of outdated geographic names and topical subdivisions were compiled and incorporated into the program. Since November 1987 this headings control program (HCTL) has been applied weekly to headings being added to LCS. In March 1988 it was retrospectively applied to all headings in the LCS subject index.

The HCTL program represents an extension of the automatic control made possible by the use of the LC authority tapes. It provides the capability to delete or replace heading subdivisions as specified on three lists that are part of the program. "List A" includes subdivisions to be deleted from all headings (e.g., —Addresses, essays, lectures). "List B" includes more than thirty geographic names to be replaced by newer forms, whether they appear as main headings or subdivisions (e.g., Zaire replaces Congo). "List C" includes more than sixty abbreviated or outdated subdivisions to be replaced (e.g., —Pol. & govt. becomes —Politics and government and —Yearbooks becomes —Periodicals). These lists can be amended as additional problems are identified and as LC policy changes.

HCTL makes it possible for part of a heading to flip to another form while any subdivisions following it remain the same. For example, "Insurance, Social—Florida" can be partially flipped to "Social security—Florida" if the cross-reference "Insurance, Social, see Social security" is already in LCS. If the cross-reference points to more than one heading, HCTL reports the possible partial flips so that library staff can decide which one is most appropriate.

Even when automatic correction is impossible, HCTL assists OSUL staff by identifying headings likely to be incorrect and reporting them for human review. A fourth list ("List X") was used in the retrospective application of HCTL to identify headings with "city flip" subdivisions (i.e., subdivisions no longer authorized for use after names of cities). These headings were reported for review and manual correction.

The program also isolates problems by looking for a partial match when an incoming heading does not exactly match any heading already used in LCS. It truncates the heading from the right end by dropping the final subdivision (e.g., Agriculture—Research—United States becomes Agriculture—Research). The search for an exact match is then repeated using the truncated form, or "grandparent" heading. If no exact match is found, the heading is further truncated by dropping a second subdivision from the right end (e.g., Agriculture—Research becomes Agriculture) and so forth, until no subdivisions remain. The headings for which there are still no exact matches found at the end of the truncation process are said to have "No grandparent in LCS" (i.e., the unsubdivided heading does not match any other LCS heading). The headings that wind up in this category are the ones most likely to contain errors, particularly typographical ones (e.g., Agric*lu*ture).

HCTL further isolates headings likely to contain errors by looking at their verification status during the truncation process. Headings in the LCS indexes are coded as either verified or unverified; those that are unverified are flagged with an asterisk. Headings become verified by meeting certain criteria; for example, any heading on an OCLC record created and input by LC according to *AACR2* is automatically verified when it is added to the file. Headings already in LCS become verified when they match those on the LC authority tapes or when they are manually verified by OSUL staff.

Separate reports are generated for headings that have verified and unverified grandparents. For example, if no exact match is found in LCS for the incoming heading "Dogs—Breeding," the HCTL program truncates the heading and looks for "Dogs." If an exact match is found and "Dogs" is verified on LCS, then "Dogs—

Breeding" is said to have a verified grandparent in LCS. Headings with unverified grandparents are more likely to contain errors than those with verified ones.

In a limited number of cases, headings with verified grandparents become verified themselves without being reported for review. OSUL staff identified 14 very general subdivisions (e.g.,—Periodicals) that may be added to virtually any heading. If a heading consists of a verified grandparent plus one of these subdivisions, it is automatically verified without being reported.

While the HCTL program was written primarily to control subject headings, it has also been applied to personal name headings to identify and, in some cases automatically combine, two headings that are identical except for the lack of a death date in one. The program looks at the verification status of both headings and corrects one of them or simply reports them, according to specifications.

When the HCTL program was retrospectively applied, there were approximately 766,000 headings (including cross-references) in the subject index. These headings came from cataloging done by OSUL on OCLC since 1972 and from catalog records of the State Library of Ohio, Center for Research Libraries, and the U.S. Government Printing Office that had been added to LCS. Table 1 shows how many of these headings were automatically corrected and/or reported for review.

These figures show that 6 percent of the headings in the LCS subject index in March 1988 were automatically corrected by the HCTL program. An additional 7 percent were reported as potential problems needing review and manual correction. The status of 18 percent of the headings on the file changed from unverified to verified because the grandparent headings were verified.

On a weekly basis an average of 1,250 OCLC records are added to LCS, and the HCTL program reports an approximate average of 300 headings with verified grandparent, 15 with unverified grandparent, and 125 headings with no grandparent in the LCS subject index.

In their recent study of the compatibility between LCSH and subject headings in the catalog at the University of Michigan, Frost and Dede considered how easily conflicting headings might be automatically controlled and what types of corrections would require human intervention. They concluded that the LC Subject Authority File (SAF) and automated authority control systems could effectively be used to control main subject headings and topical and chronological subdivisions but that geographic subdivisions posed more of a problem.

Frost and Dede suggested that lists of free-floating subdivisions from the *LC Subject Cataloging Manual: Subject Headings* might be added to the SAF and a "parts-matching" approach used to authorize more headings than could be authorized by the use of the SAF alone.[6] OSUL has used a list approach and a parts-matching process in a different way to accomplish a similar purpose.

Table 1. Headings Automatically Corrected and/or Reported for Review

Automatic Corrections	
Subdivision deleted	30,521
Subdivision replaced	8,398
Heading partially flipped	4,429
Reported headings	
No grandparent*	21,354
Unverified grandparent	28,934
Verified grandparent	139,840
City flip	4,879

*These headings all have at least one subdivision.

To summarize, HCTL has improved our ability to correct outdated headings automatically by providing us with a means of globally updating subdivisions as well as main headings. It also makes more extensive use of the cross-references in LCS by partially flipping headings when the grandparent portion matches a reference. It has improved our ability to identify problems by automatically grouping headings according to their verification status and relation to other LCS headings.

The HCTL program prevents many variant headings from being added to LCS so that fewer corrections have to be done manually, and correct data is displayed to the public right away. It helps OSUL make the best use of limited human resources by grouping data with common characteristics so that attention can be focused on the areas where it is needed most. This type of programming is particularly valuable when OSUL does retrospective conversion projects and processes cataloging backlogs—activities that tend to reintroduce outdated, incorrect headings to the file. The specifications of the HCTL program can be modified as LC policy changes and can greatly assist library staff in maintaining the quality and currency of the LCS database.

REFERENCES

1. Nadine L. Baer and Karl E. Johnson, "The State of Authority," *Information Technology and Libraries* 7:148 (June 1988).
2. Ibid.
3. Lorene E. Ludy and Susan J. Logan, "Integrating Authority Control in an Online Catalog," *Proceedings of the ASIS Annual Meeting* 19:176–78 (1982).
4. For descriptions of the use of the LC name and subject authority tapes at OSUL, see Lorene E. Ludy, "LC Name Authority Tapes Used by Ohio State University Libraries," *Information Technology and Libraries* 3:69–71 (Mar. 1984); and Lorene E. Ludy, "OSU Libraries' Use of Library of Congress Subject Authorities File," *Information Technology and Libraries* 4:155–60 (June 1985).
5. Karen Markey and Diane Vizine-Goetz, *Characteristics of Subject Authority Records in the Machine-Readable Library of Congress Subject Headings* (Dublin, Ohio: OCLC Online Computer Library Center, Aug. 1988).
6. Carolyn O. Frost and Bonnie A. Dede, "Subject Heading Compatibility between LCSH and Catalog Files of a Large Research Library: A Suggested Model for Analysis," *Information Technology and Libraries* 7:288–98 (Sept. 1988). ■■

INSTRUCTIONS TO AUTHORS

Information Technology and Libraries welcomes manuscripts related to all aspects of library and information technology. Some specific topics of interest are mentioned on the masthead page. Feature articles, communications, letters to the editor, and news items are all considered for inclusion in the journal. Feature articles are refereed; other items generally are not. All material is edited as necessary for clarity and length.

Manuscripts must be typewritten and the original submitted with one duplicate. Do not use onion skin. All text must be double-spaced, *including footnotes and references*. Manuscripts should conform to *The Chicago Manual of Style*, 13th ed., rev. (Chicago: Univ. of Chicago Pr., 1982). Illustrations should be prepared carefully as camera-ready copy, neatly drawn in a professional manner on separate sheets of paper. Manuscript pages, bibliographic references, tables, and figures should all be numbered consecutively.

Feature Articles consist of original research, state-of-the-art reviews, or comprehensive and in-depth analyses. An abstract of one hundred words or less should accompany the article on a separate sheet. Headings should be used to identify major sections. Authors are encouraged to relate their work to other research in the field and to the larger context of economic, organizational, or management issues surrounding the development, implementation, and use of particular technologies.

Communications consist of brief research reports, technical findings, and application notes. An abstract need not be included.

Letters to the Editor may offer corrections, clarifications, and additions to previously published material, or may be independent expressions of opinion or fact related to current matters of concern in the interest area of the journal. A letter commenting on an article in the journal is shared with the author, and a response from the author may appear with the letter.

News and Announcement items may announce publications, conferences, meetings, products, services, or other items of note.

Book Reviews are assigned by the book review editor. Readers wishing to review books for the journal are invited to contact the book review editor, indicating their special areas of interest and expertise.

Names and addresses of the journal editors may be found in paragraph three on the masthead page. In all correspondence please include your own name, institutional affiliation, mailing address, and phone number.

News and Announcements

Thomas W. Leonhardt Named New *ITAL* Editor

The Library and Information Technology Association (LITA), a division of ALA, has announced that Thomas W. Leonhardt has been named editor of its journal, *Information Technology and Libraries* (*ITAL*), beginning with the 1990 volume. For the remainder of the 1989 volume, Leonhardt will serve as editor-designate and work with the current editor, William Gray Potter.

Leonhardt is dean of Libraries at the University of the Pacific in Stockton, California. Prior to his current position he was assistant university librarian for Technical Services at the University of Oregon and earlier was head of Acquisitions at Duke University. He has also held positions in the libraries at Boise State and Stanford universities. He holds an M.L.S. from the University of California at Berkeley.

Thomas W. Leonhardt

Leonhardt has been on the *ITAL* Editorial Board since 1985 and is a member of the LITA Publications Committee. He is currently editor of the *RTSD Newsletter*, a publication of the Resources and Technical Services Division of ALA, and has held many other appointments in that division.

Leonhardt's other editorial experience includes service on the editorial boards of *Library Acquisitions: Practice and Theory* and *Library Resources & Technical Services.* He is also editor of the Library and Information Science Series published by JAI Press and has published widely in the areas of acquisitions and technical services in such journals as *The Serials Librarian* and *Library Acquisitions: Practice and Theory.*

His address is University Libraries, University of the Pacific, Stockton, CA 95211.—WGP ■■

Rosary College Graduate School Hosts National Institute on Library Network Management

A national, week-long institute addressing the problems of library networks, consortia, and utilities will be presented by the Rosary College Graduate School of Library and Information Science, River Forest, Illinois, on May 15–19.

The institute, the first of its kind nationally, will bring together key opinion leaders in the library community who currently hold, or who are likely to hold, advisory positions in the rapidly changing library networks field. Through seminars and elaborate role-playing scenarios, participants will identify, analyze, and—where possible—predict the problems libraries and networks will face in the next decade of networks.

Faculty for the institute are William Welsh, deputy director emeritus of the Library of Congress; Rowland Brown, recently retired president of Online Computer Library Center; Richard Dougherty, professor at the Graduate School of Library Science and Information Studies at the University of Michigan and former director of the University of Michigan Library; and Pat Molholt, associate university librarian at Rensselaer Polytechnic Institute and author of a recent study, "Library Networking: The Interface of Ideas and Actions," commissioned by the Department of Education.

Institute director is Michael Koenig, dean of the Graduate School of Library and Information Science at Rosary College.

Library networking, in both the technical and nontechnical sense, is at a crossroads. Resource sharing is in danger of losing out to isolationism, said Koenig. Every time librarians choose to purchase cataloging data from vendors and load it into stand-alone systems, the nation loses the ability to share the item represented by such records.

Experts in network management and planning agree that libraries are moving from traditional bibliographic networks to broad information delivery and access support systems and from a national to a global focus. Because of this transition, library professionals face major challenges in the twenty-first century.

Funded by a grant from the U.S. Department of Education, the institute is tuition-free, with lodging and meals at Rosary College provided from Sunday evening, May 14, through Friday afternoon, May 19. Participation is limited to seventy-five library professionals with relevant networking experience. The selection committee is chaired by Beverly Lynch. All interested professionals should submit an application by April 1 to Dr. Beverly Lynch, University Librarian, University of Illinois at Chicago, Box 8198, Chicago, IL 60680.

For a free brochure, contact the Rosary College Graduate School of Library and Information Science at (312) 366-2490, ext. 302, or write: Library Network Management Institute, Rosary College Graduate School of Library and Information Science, 7900 W. Division St., River Forest, IL 60305. ■■

Ohio Announces Plans for Statewide Library Information System

The State of Ohio and the Ohio Board of Regents have appropriated $2.5 million as the initial capital budget to begin the process of implementing a statewide integrated library and information system.

A statewide project is aimed at developing the Ohio Library Information System, which will initially link the libraries of fifteen state-supported universities. The concept of a library access system grew out of a 1987 Ohio Board of Regents' Library Study Committee's recommendation that "the State of Ohio implement as expeditiously as possible a statewide electronic catalog system." The proposed system, currently referred to as OLIS, is intended to increase the research effectiveness and productivity of faculty and students at Ohio's colleges and universities. OLIS will integrate traditional catalog and circulation functions for university libraries, incorporate a document-delivery service, and serve as a multidimensional information system that provides shared access to a wide range of databases and other informational sources well beyond those found in a card catalog.

To further demonstrate the importance of OLIS to the Ohio Board of Regents, William Coulter, chancellor, has recently incorporated OLIS into the regents' selective excellence initiatives—a series of interrelated challenge programs that has already garnered considerable national recognition.

Two directors have been appointed for the OLIS project, and inquiries should be directed to them: Greg Byerly, Director of Library Systems for OLAS, Head of Systems, Kent State University Libraries, Kent, OH 44242, (216) 672-2962; J. Carroll Notestine, Director of Computer Systems, OLAS, 759 Drummond Court, Columbus, OH 43214, (614) 395-1551 or 451-3783. ■■

Paris Award Gives CLSI Two Largest City Library Automation Projects

CLSI has announced that its office in Paris, France, has been selected to automate Le Reseau des Bibliotheques de la

Ville de Paris (RBVP). This five-million-dollar agreement represents the largest city library automation project in Europe and is the second largest automation contract for CLSI.

With the Paris award, CLSI holds the two largest city library automation projects in the world. The automation of the New York City/Brooklyn Public Library, which CLSI began in 1986, is the largest city contract in the library and information industry.

The RBVP consists of forty-nine general public libraries and five special libraries, with 3.2 million volumes and an annual circulation of seven million. The Technical Services Center manages all orders for all the libraries, including receiving, cataloging, and circulation, and annually processes 300,000 documents.

Current implementation plans call for placement of fifty terminals by August 1989 to demonstrate the City of Paris' automation system during the International Federation of Library Associations meeting in Paris. CLSI will install Sequent's Symmetry 81 Parallel Processor, 5 Altos Series 2000 superminicomputers, and more than 500 terminals, of which 200 will be for CL-CAT, CLSI's online public access terminal. ■■

120 NOTIS Installations

NOTIS Systems, Inc., has set another record: NOTIS Systems, Inc., has a total of 120 customers, after installing 30 library management systems in 1988 and 34 in 1987.

NOTIS installations now include many of the major academic research libraries in the United States. Its management software is also the basis for two statewide library networks and the choice of corporate, special, and public libraries throughout North and South America.

NOTIS library management software is now in release 4.5. A new product, GTO (Generic Transfer and Overlay), was recently made available to customers. It permits the online transfer of RLIN, OCLC, and UTLAS bibliographic and authority records into the local NOTIS database. Access to locally mounted external databases through the interface of the NOTIS online public access catalog will be available in 1989.

Walter Stine Receives 1988 LITA/CLSI Scholarship

Walter D. Stine has been awarded the 1988 LITA/CLSI Scholarship, a cash award of $1,500 made to a beginning student at the master's level in an ALA-accredited program in library and information science with emphasis on library automation. The LITA/CLSI scholarship is supported by a contribution from CLSI, Inc., Newtonville, Massachusetts, and is administered by the LITA Education Committee.

Stine, currently enrolled in the School of Information and Library Science at the University of North Carolina at Chapel Hill, received his B.A. from the University of California at Santa Cruz, where he majored in history and religious studies. He worked for eight years as a paraprofessional in public and academic libraries. As a cataloger at the Southern Oregon State College Library, he received experience with a bibliographic utility and participated in automating the student payroll and printed serials list.

As Stine explains his view of the role of library automation: "My interest in automation goes back to some of my first work experiences in libraries. What inspired me then and continues to now is the role of the library in supplying the information needs of people of all backgrounds and many levels of sophistication. Library automation is the logical extension in our mission to supply these needs more accurately and efficiently."

Stine clearly fulfills the LITA/CLSI Scholarship requirements of academic excellence, leadership, and evidence of a commitment to a career. For further information, write or call LITA, 50 E. Huron St., Chicago, IL 60611-2729; (312) 944-6780. ■■

PBS Announces UNIX Version of Pro-Cite

Personal Bibliographic Software, Inc. (PBS), has announced the development of a UNIX version of Pro-Cite, its bibliographic database management program. The UNIX version will initially run on Sun workstations.

Pro-Cite is designed for organizing and formatting bibliographic information. Users can create personal databases by entering information into twenty different work forms, ranging from those for journals and books to conference procedings and art works, or create work forms of their own. Records can also be transferred automatically into Pro-Cite from online database searches using PBS Biblio-Links or imported from other database programs. Once in Pro-Cite, records can be searched, sorted, indexed, and formatted automatically in any punctuation style. With Pro-Cite, users can create accurate, attractive, correctly formatted bibliographies with no knowledge of the rules of the format itself.

Other PBS products for IBM personal computers and compatibles include Pro-Search, a specialized front-end search aid program for easy searching of BRS and Dialog online database services, and Biblio-Links, reformatting programs that transfer records downloaded from online database systems directly into a Pro-Cite database. Biblio-Links are available for BRS, Dialog, MEDLARS, USMARC, STN, OCLC, NOTIS, and RLIN. Pro-Cite and Biblio-Links to BRS, Dialog, and MEDLARS are also available for the Macintosh. ■■

HARLiC CD–ROM Union Catalog

The Houston Area Research Library Consortium (HARLiC) will use a $100,000 HEA Title II-D grant from the Department of Education to develop a CD–ROM catalog of the consortium's combined collections of books, journals, and other materials. Marcive, Inc., San Antonio, will produce the catalog. The implementation of a union catalog will enable the HARLiC libraries to provide access to more than nine million items and to share these resources more effectively through improved interlibrary loan and coordinated collection development activities. The catalog will contain approximately 2.1 million bibliographic records on multiple CD–ROMs. It is anticipated that the catalog will be delivered and installed shortly before the beginning of the 1989 fall semester.

The following libraries are HARLiC members: Houston Academy of Medicine–Texas Medical Center, Houston Public, Prairie View A & M University, Rice University, Texas A & M University, Texas Southern University, University of Houston, and University of Texas Medical Branch at Galveston. All of these libraries, except Rice, will contribute their records to the CD–ROM union catalog project. ■■

University of Michigan Conversion Project

The University of Michigan Library and Utlas International signed agreements in May 1988 for the retrospective conversion of approximately 1,200,000 titles to machine-readable format over an eighteen-month period. SAZTEC is participating in the project as a subcontractor to Utlas to handle all document preparation, microfilming, search key creation, and original record creation in the massive conversion effort.

The project has been funded by a $1.4 million grant to the university from the W. K. Kellogg Foundation, with the university matching funds providing support. In addition to mainstream research collections, library records will be converted for transliterated versions of many nonwestern language materials, including Chinese, Japanese, and Korean titles. Some special collections are planned for conversion, though non-book materials and musical scores have been excluded from this project.

Under the terms of the agreement, Utlas and SAZTEC will provide complete retrospective conversion services. The library's shelflist cards have been microfilmed on site for use in the conversion process, assuring the security of the original cards. Search keys, with specific local data elements added, are being prepared by SAZTEC for matching against the Utlas database. After the search is carried out by Utlas, SAZTEC will provide original entry services for any record unmatched in the Utlas search. All records will undergo authority control before loading to the library's system.

The records will be delivered to the library in bimonthly shipments of 200,000 records, joining 900,000 records already in the local NOTIS system. By 1990 researchers at the University of Michigan will be able to locate records for nearly 100 percent of the library's collections acquired and cataloged over the course of more than 150 years. ■■

Michigan State University Signs with Blackwell for Authority Control

Michigan State University Libraries has signed a contract with Blackwell North America for application of authority control to Michigan State's present MARC file of approximately 830,000 records. As part of Michigan State's ongoing retrospective conversion project, an additional 490,000 records will be processed through Blackwell authority control. Blackwell will deliver the library's initial edited bibliographic file in March 1989, whereupon it will be loaded into the library's NOTIS system. Subsequent edited recon records will be sent to the library periodically, with completion of authority control on all records by September 1989.

Blackwell specializes in the supply of North American books and bibliographic support services worldwide. Its service provides Library of Congress authority control for names, series, uniform titles, and subject in MARC records. In addition, library-specific LC authority files are provided for use as cross-references in online catalogs. ■■

Carlyle Systems Moving to UNIX

Carlyle Systems, Inc., has announced plans to offer a UNIX-based version of its library automation systems. The migration will enable Carlyle to take advantage of the ongoing research and development efforts of commercial UNIX system suppliers and of the numerous software products and software tools available on UNIX.

The Carlyle system utilizes a multiple processor, distributed architecture with a special purpose bibliographic database server. The new design places UNIX at the heart of the system, running the processors that handle transaction functions, such as circulation, acquisitions, and serials control. Under UNIX, these processors will support a larger number of simultaneous users, will adhere to open systems interconnection (OSI) standards, and will include optional drivers to CD-ROM products and external dial-up database services. They will also support a wide range of commercially available UNIX-based products including report writers and electronic mail systems. ■■

Faxnet Brings Rural Montana into Information Mainstream

A permanent, statewide facsimile network is being used by residents of Montana to quickly communicate information to and from bases of financial and political power such as New York; Washington, D.C.; and Chicago. The facsimile network, available in more than twenty-five Montana public, law, education, and government libraries, was conceived as a pilot program by librarians in Montana's state law library in 1986.

During the pilot, funded by a trust, some facsimile machines were placed at specific library sites, while others were rotated among municipal libraries and state agencies. Fax machines were also rotated among select legal, legislative, and medical organizations for periods of four to six months. By temporarily placing the machines in these organizations, the libraries believed they could reach a large number of Montana residents, teaching them how to use facsimile technology to benefit them in their professional, educational, and social interests.

Among the varied applications performed with the fax machine, they were used by Montana legislators to transmit testimony to their congressional delegation in Washington, D.C.; legal offices obtained research information; and health organizations received documents from major research libraries. The Department of Agriculture used the network extensively to communicate with Japan to market Montana beef in that country. The pilot was deemed a success by the state's residents and became permanent in October 1988.

"People who have never been to the library come here to use the fax machine—county administrators, local government officials, physicians, private consultants, lobbyists, lawyers, insurance representatives," says Linda Brander, Faxnet project director. ■■

Colorado Libraries Begin Phase Two of IRVING Project

Four libraries using Dynix systems in Colorado have been granted $65,000 in LSCA funds to complete the online connection to the IRVING Library Network. Phase one of this project, which began in 1987, used $50,625 in LSCA funds to develop software for making this interconnection possible. Phase two funding will provide the actual hardware and telecommunications equipment to bring the project to completion. These four (Adams County Public Library, Douglas County Public Library, Englewood Public Library, and the University of Southern Colorado Library) are working with IRVING and Minicomputer Systems, Inc., the Boulder company that developed the original IRVING communications software, on the project. Once phase two is complete, the IRVING Library Network will allow for transparent communications between five different vendor systems (Dynix, CLSI, Eyring, CARL, and Pueblo) and will provide all users with copy status, network statistics, and interlibrary loan capability to the more than 5,000,000 items in the combined systems. The Dynix-IRVING grant project is administered by the Colorado State Library.

Carlyle Offers Online Services

Access to more than 900 online databases is now available from Carlyle Systems, Inc. Carlyle has signed an agreement with Telebase Systems, Inc., to offer its EasyNet service to libraries under the name Carlylenet.

Carlylenet is a gateway to thirteen host systems, including Dialog, BRS, Orbit, Newsnet, Wilsonline, ADP Network Services, Data-Star, DataTimes, G. Cam Serveur, Pergamon Financial Data Services, Profile Information, Questel, and Vu/Text. It serves as an intermediary between the user and the host system, translating the user's search statement into the language the host system understands. Carlylenet provides both the ease of menu searching for end users and the power of command searching for online professionals.

Libraries can subscribe to the Carlylenet service in several different ways:

Dial-in Service. Existing hardware and software can be used, and the library pays for service on a usage basis.

Answer Machine Service. Libraries can install the Answer Machine, a keyboard/CRT/modem unit housed in a Brodart study carrel and intended for heavy public use.

Direct Connect Service. Libraries with online book catalogs will be able to offer Carlylenet to users through online catalog terminals. The opening screen will give users the option to search either books or articles. When they select books, the system searches the library's own in-house catalog; for articles, the system searches Carlylenet. ■■

Recent Publications

Book Reviews

End-User Searching: Services and Providers. Ed. by Martin Kesselman and Sarah B. Watstein. Chicago: American Library Assn., 1988. 230p. paper, $26 (ISBN 0-8389-0488-2).

For many academic and public libraries, end-user searching is a return to their usual way of doing business: self-service access to information. Until recently, the complexity and expense of time-sharing systems required librarians to conduct searches on behalf of their clients. Today, menu-driven systems, proliferation of personal computers, after-hours price reductions, and availability of laser disks have all combined to make self-service access to machine-readable data in the library possible and attractive.

Despite the great interest in the topic in the past few years, *End-User Searching: Services and Providers* is the first book to deal with the subject from the perspective of librarians. The editors have brought together writings that emphasize practical issues: starting a service, training end-users, and choosing appropriate systems and databases. They include extensive annotated lists of articles, books, and journals for end-users, and a directory of online products and services as well. More than half the book is devoted to describing systems and databases and because this topic is approached by vendor (BRS, Dialog, Wilson); by subject (business and science); and by format (laser disk), there is some overlap of information. The redundancy is probably useful for the reader who is unfamiliar with the subject, however.

Ida Lowe's chapter on starting an end-user service offers practical advice on issues that must be resolved beforehand. She is a strong advocate of using questionnaires and surveys to determine the level of interest and potential demand for the service. Another approach, not mentioned in the chapter, is to offer the service on a trial basis and see what the response is. When Telebase Systems (Easynet) wanted to determine the potential use of its Answer Machine in the academic market, the company did not ask students and faculty what their use would be but simply set up terminals in libraries and offered free searching for a month. From this sample the volume and pattern of use could be estimated. Libraries can imitate this approach. For example, laser disk producers will often permit libraries to test their products for a few weeks before requiring a commitment to buy.

End-User Searching is a useful handbook for librarians who are starting a service with little experience in online searching, but it is of limited interest to those who are operating end-user services or are experienced searchers. For the experienced, the work would have more appeal if it contained chapters on some of the topics I have listed below:

• Nontraditional databases and systems. The systems described in *End-User Searching* are, for the most part, widely available and familiar. I would have liked a discussion of databases and systems that are not library standards. In the area of business, for example, databases such as Compustat and systems such as ADP are most useful and generally ignored by library literature.

• The future of end-user searching. For example, although BRS Onsite is mentioned briefly in the book, there is no discussion of the potential for remote access to the library's time-sharing systems. Remote access brings with it a new set of problems and opportunities as the bulk of searching shifts outside the physical library. Gateways that allow the user remote access to the library's systems must be installed.

• Choosing among alternative technologies. With databases available on-disk, online, and on-site, how do we make appropriate choices concerning the format of machine-readable data?

Perhaps the editors will include these topics in their next volume.—*Michael Halperin, Lippincott Library, Wharton School, University of Pennsylvania, Philadelphia.* ■■

Webb, T. D. *The In-House Option: Professional Issues of Library Automation.* Library and Information Sciences Text Series, no.1. New York and London: Haworth, 1987. 166p. $34.95 (ISBN 0-86656-617-1).

Terry Webb studies the effect of automation on the library profession and the profession's response. The issue of host computer location serves as the vehicle for the central focus of the book.

He addresses the failure of the librarian to respond and interact with new technology. He feels this is the basis of the problem: that by surrendering knowledge of automation to non-librarians, the professional surrenders both control and authority over libraries.

To overcome this, librarians need to improve computer skills. Webb argues that traditionally, librarians have come from non-scientific backgrounds and, as such, have had little automation experience. They have exhibited responses ranging from technophobic avoidance to superstitious, counterproductive technophilia. Furthermore, computer education available to librarians has been inadequate. For example, he identifies two types of education: microcomputer operations, knowledge of which is useful for office management but not for mini and mainframe operation; and automation applications, which covers broad generalizations and is useful for management decision-making. Programming, systems design, and file structures are unfamiliar concepts to librarians. These are precisely the skills necessary to run even simple turnkey systems bought as packages from vendors.

Before discussing the case studies, Webb reviews the three major types of computers (micro, mini and mainframes) and their capabilities. The training and skill necessary to efficiently operate each of these types varies, and because hardware and software vendors offer limited training, it is up to the purchasing librarian to understand the nature of the tool bought.

Ultimately, the location of the computer decides the degree of authority library administrators will have. In-house (within the library) location ensures library administration autonomy. A remote location (outside the library, but not necessarily outside the institution) suggests a potential loss of authority. This can be avoided, Webb argues, by effective communication and even cross-training of librarians and computer experts. Webb's message is clear: install in-house and assure control and autonomy; locate outside and risk losing control. The choice therefore ought to be clear to any library administrator. But is it? The argument is uncomfortably seductive. The real issue is the need to educate the profession to use the new tools available to it.

Webb uses two illustrative case studies. The first, the Phoenix Public Library, chose a remote location to handle 1.5 million records. Implementation and maintenance operations, although difficult, were not unusual. What was unusual was the utter lack of communication between librarians and computer experts within the city's MIS department. Ultimately, the library was administratively swallowed by the Parks and Recreation Department. Remote location of computer resources, in this case, shattered what was perhaps an already tenuous administrative authority.

The Joseph F. Smith Library, on the other hand, located in-house and retained its autonomy despite the usual installation and operational horror stories. Admittedly, the collection was only one-tenth the size of the Phoenix Public Library, and the staff and resources available to the library were "automation friendly." Nonetheless, library administrators were able to retain authority at the cost of library space and increased staffing responsibility.

The case studies are allegorical. In either case, if administrative support had been weak and resources (budgetary and trained personnel) poor, the outcome would have been equally disasterous. The moral of both stories therefore is that in order to ensure a library's continued success, the library must have a clear and cohesive management as well as access to resources and trained staff.

Although Webb's point is perceptive, his arguments are sometimes faulty in his final analysis. For example, he is critical of the trend by vendors to develop ever-increasing levels of sophistication in response to market demand, but he maintains that librarians are "probably" satisfied, albeit overwhelmed by sophisticated systems. Nonetheless, he argues, the need to tailor systems to meet local needs precludes inhouse hybridization of vendor packages. He is even more critical of vendors for not offering enough training and support, yet the question remains: Is this really the responsibility of a contracted supplier? He writes that in an ideal world the automated system would be geared to the literacy level of the staff. But increasing sophistication of automation technology demands a positive response from the library profession. He emphasizes that only through the enhancement of computer expertise within the profession can it hope to retain traditional administrative responsibility.—*Mary Hemmings, Health Sciences Library, McGill University, Montreal, Canada.* ■■

Intellectual Property Rights in an Electronic Age: Proceedings of the Library of Congress Network Advisory Committee Meeting April 22–24, 1987. Network Planning Paper, no.16. Washington, D.C.: Network Development and MARC Standards Office, LC, 1987. 66p. paper, $7.50 (ISBN 0-8444-0592-2).

Intellectual property rights are among the formidable issues of the information field. Existing law is severely challenged by rapidly evolving development and use of newer technologies for the creation, compilation, production, storage, display, and performance of works. With as many views as players in the field, perhaps more as some players have more than a single role, and as resolution of the issues has far-reaching economic, educational, and social consequences, this publication is most welcome for the clarity of the papers and for setting the stage for further discussion. It is an excellent overview of several key issues and positions, often provided with feeling and humor. *Intellectual Property Rights in an Electronic Age* is the sixteenth in a laudable series of documents focused on network-related issues.

Consideration is given first to the April 1986 report from the Office of Technology Assessment on *Intellectual Property Rights in an Age of Electronics and Information.* D. Linda Garcia, project director of the OTA study, provides background on its conception and structure, and on the processes employed by the study for obtaining wide and diverging input. The history and purpose of copyright and its relation to printing technology is briefly traced and then contrasted with the burdens that digital technologies place on current copyright law—problems of identification of authorship, of infringements and of rights of enforcement, of the integrity of a work, of derivative use and private use. Distinctions in copyright and patent law separating works of authorship from processes or works of invention are blurred with functional works such as software. The OTA study calls attention to the long-term nature of the problems, to the need for significant changes in the construct of the intellectual property system. Piecemeal fixes are not likely to last beyond the short-term. Garcia notes that this "was not a popular message to the Congress."

Robert Kost, then the legal analyst for the OTA study, in a witty paper titled "The End of Copyright," concisely discusses information as property and the notion of copyright in relation to three historical periods of technology. These are the ages of static media (printing) and the control of copies, the age of dynamic media (film, TV) and the control of use of a work, and now the age of digital media. The digital information era poses practical questions of transaction costs or costs of enforcement and permissions, as well as theoretical questions about what is owned or controlled and about differentiating among works of function, fact, and art. If copyright is to be preserved in the digital era, he anticipates continuing development of case law regarding infringing conduct, and such technological stopgaps as disabling signals, as well as compulsory licensing and collection agencies.

Ralph Oman, register of copyrights, in "The Copyright Law: Can It Wrap Itself Around the New Technologies?" discusses the intent of Congress in passing the 1976 Copyright Act. He suggests that some truth exists in each of these statements: the 1976 law is not technology-independent; the law is technology-independent but courts have been reluctant to apply it expansively to new technology; and it is impossible to enact a truly technology-independent copyright law.

W. David Laird's pragmatic presentation partly is on copyright of bibliographic databases. He also suggests that a social theory concept of distributive justice rooted in fairness be applied to the distribution of intellectual property rights. If distributive justice cannot embrace print and non-print works in a law satisfactory to most stakeholders, he concludes that copyright law should be limited to print works and that contract law or regulation should protect other forms.

John Hearty and Barbara Polansky cover contracts and licenses in "ACS Chemical Journals Online: Is It Being Downloaded, Do We Care?" Both affirm occurrence of legal and illegal downloading with tales of how copyright violations come to light. Both care, but with differing priorities. Hearty presents the marketing perspective and suggests that unless violations involve reselling, then minor violations are less important than customer use of the product. Polansky, as a copyright administrator, takes a much stronger stand in addressing tensions between user needs and owner rights. She suggests that users want clearer guidelines about what is permissable, that contractual terms must be communicated to users, and that copyright law needs amending, not scrapping.

Although text is lacking for two other presentations by legal counsel to a House Judiciary Subcommittee and by an Information Industry Association representative, a nice introduction to the proceedings prepared by Carol Henderson briefly provides the gist of their remarks as well as that of audience discussion. As several of the papers mention transaction costs, the interested reader is referred to a recent, excellent overview by Roger A. McCain, "Information as Property and as a Public Good: Perspectives form the Economic Theory of Property Rights," *Library Quarterly*, v.58, no.3, July 1988, p.265–82.

A second NAC program meeting on intellectual property rights in a network context was held in March 1988 and its proceedings are forthcoming as number 17 in the LC Network Planning Paper series; a detailed summary appears in *Library of Congress Information Bulletin* 47, no.32 (Aug. 8, 1988), p.326–28.—*Alan R. Benenfeld, Northeastern University, Boston.* ■■

The Linked Systems Project: A Networking Tool for Libraries. Ed. by Judith G. Fenly and Beacher Wiggins. OCLC Library, Information, and Computer Science Series, no.6. Dublin, Ohio: OCLC, 1988. 138p. paper, $13.50 (ISBN 1-55653-039-0).

For several years the participants in the Linked Systems Project, the Library of Congress, the Research Libraries Group, and OCLC Online Computer Library Center (and from time to time the Western Library Network, which is not represented in this work) talked to any group that sat still long enough to listen to their vision of a bibliographic world in which the large bibliographic databases worked together in a functionally integrated fashion. This slim monograph, with some additions, is a collection of papers originally prepared for the San Francisco ALA meeting in 1987 and as such represents a status report of the Linked Systems Project.

Unfortunately, the proclivity to talk to anyone who would listen led to a failure to identify a target audience for the book. Consequently, it is a mix of introductory material useful for the LSP novice implementer, the middle manager, and the library administrator. While the two latter categories might well be served by a single monograph, trying to satisfy the technician as well leads to a hodge-podge product. It would have been better to invest more time and effort to produce a detailed presentation of the project for technical people and a similarly detailed, but very different presentation for potential users and managers of linked systems.

The editors had their hands full, given the politics of the institutions involved and the uneven nature of the contributions. It would have taken a far heavier editorial hand to produce a smooth-flowing book. Clearly, it is a collection of individual pieces, replete with repeated introductory information. The material by Sally McCallum and Linda Arret (Overview and Information Retrieval) is characteristically sound and well organized. David Bishop provides the sort of balanced discussion of the cost aspects of the project which all administrators will appreciate and applaud. Ray Denenberg's contribution on the protocols themselves, while a tough assignment to explicate, will be appreciated more by those immersed in the nuances of telecommunication protocols and familiar with the eleven new acronyms introduced in this chapter. Joe Matthews' contribution on the interest of the vendor community is one of the more hopeful future aspects of the LSP project that receives short shrift in the all-too-brief contribution on the future uses of the LSP link.

Readers will need some understanding of the national bibliographic scene, basic telecommunication strategies and concepts, and a notion of bibliographic system capabilities and the differences between the OCLC, RLIN, and the Library of Congress processing systems. This level of sophistication is assumed, and library and telecommunication jargon is rampant.

While the preface describes the work as "meant to be a full treatment of LSP," it is little more than a status report. There is minimal recognition of the financial support roots of the project. The Council on Library Resources and its Bibliographic Service Development Program are identified as "contributing to the increased networking activity at this time." CLR was one of the prime movers in very troubled waters when the project began, encouraging cooperation, cajoling reluctant institutions to stay involved, and ultimately investing well over a million dollars and thousands of staff hours in the series of projects that became LSP. No mention is made of the CLR-funded Battelle study of the alternatives and strategies for linking. One would expect that institutions with any prospect for future support from a faithful and supportive organization would recognize past roles. Recognition of the council's involvement is sparse: "Funding for the project was provided by CLR." CLR's contributions, its tenacity, its consistency, its integrative and encouraging support of the project deserve recognition, even in a status report of the project. How soon we forget or choose to ignore.

The editors chose to use a detailed table of contents as opposed to an index. This reviewer, familiar with LSP, could find his way around the book, but still would have preferred a good index. Tables of content do not replace good indexes, though they do supplement them.

As a status report on LSP both from a technical and management point of view, the book succeeds. It is not a history of where the project has been nor of its roots. Neither is it terribly predictive, except for implementations now in planning stages, of future uses of the link. If you know nothing about the LSP project and the protocols, you will come away from this book knowing more than when you picked it up.

As a final note, Hooray for OCLC for holding the price to a reasonable level.—*C. Lee Jones, Mid-Atlantic Preservation Service, Bethlehem, Pennsylvania.* ■■

Other Recent Receipts

Listed here are books and other publications received for review that are of potential interest to LITA members. Some of these materials may be reviewed in later issues of ITAL.

Advances in Library Automation and Networking: A Research Annual. Ed. by Joe A. Hewitt. Advances in Library Automation and Networking, v.1. Greenwich, Conn., and London: JAI, 1987. 232p. individuals, $28.25; institutions, $56.50 (ISBN 0-89232-385-X).

Bodian, Nat G. *Bodian's Publishing Desk Reference: A Comprehensive Dictio-*

nary of Practices and Techniques for Book and Journal Marketing and Bookselling. Phoenix, Ariz., and New York: Oryx, 1988. 439p. $49 (ISBN 0-89774-454-3).

Buckland, Michael K. *Library Services in Theory and Context.* 2d ed. Oxford and New York: Pergamon, 1988. 251p. paper, $45 (ISBN 0-08-035754-7).

Burger, Ralf. *Computer Viruses: A High-Tech Disease.* Grand Rapids, Mich.: Abacus, 1988. 282p. paper, $18.95 (ISBN 1-55755-043-3).

CD-ROM Applications and Markets. Ed. by Judith Paris Roth. Westport, Conn., and London: Meckler, 1988. 147p. $34.50 (ISBN 0-88736-332-6).

CD-ROMs in Print 1988–1989: An International Guide. Ed. by Jean-Paul Emard. Westport, Conn., and London: Meckler, 1988. 164p. paper, $37.50 (ISBN 0-88736-274-5).

Chan, Lois Mai, and Richard Pollard. *Thesauri Used in Online Databases: An Analytical Guide.* New York: Greenwood, 1988. 268p. $45 (ISBN 0-313-25788-4).

Computer-Readable Databases. 5th ed. Ed. by Kathleen Young Marcaccio and Janice A. De Maggio. Detroit, Mich.: Gale, 1989. 1,188p. paper, $160 (ISBN 0-8103-2775-9).

Connecting with Technology 1988: Microcomputers in Libraries. Ed. by Nancy Melin Nelson. Westport, Conn., and London: Meckler, 1988. 85p. $29.50 (ISBN 0-88736-330-X).

Corbin, John. *Implementing the Automated Library System.* Phoenix, Ariz.: Oryx, 1988. 153p. $30 (ISBN 0-89774-455-1).

Davis, Charles H., Gerald W. Lundeen, and Debora Shaw. *Pascal Programming for Libraries: Illustrative Examples for Information Specialists.* Contributions in Librarianship and Information Science, no.60. New York: Greenwood, 1988. 109p. $25 (ISBN 0-313-25259-9).

Desmarais, Norman. *The Librarian's CD-ROM Handbook.* Westport, Conn.: Meckler, 1989. 174p. $35 (ISBN 0-88736-331-8).

Eaton, Nancy L., Linda Brew McDonald, and Mara R. Saule. *CD-ROM and Other Optical Information Systems: Implementation Issues for Libraries.* Phoenix, Ariz.: Oyrx, 1989. 153p. paper, $29.50 (ISBN 0-89774-448-9).

Elshami, Ahmed M. *CD-ROM: An Annotated Bibliography.* Englewood, Colo.: Libraries Unlimited, 1988. 138p. paper, $24.50 (ISBN 0-87287-702-7).

Expert Systems: Concepts and Applications. Ed. by Charles Fenly and Howard Harris. Advances in Library Information Technology, no.1. Washington, D.C.: LC Cataloging Distribution Service, 1988. 37p. paper, $15 (ISBN 0-8444-0611-2).

Fensterer, Richard. *Essential Guide to Apple Computers in Libraries: Communications and Networking.* Essential Guide to Apple Computers in Libraries, v.3. Westport, Conn.: Meckler, 1988. 219p. spiralbound, $24.95 (ISBN 0-88736-076-9).

Hunter, Beverly, and Erica K. Lodish. *Online Searching in the Curriculum: A Teaching Guide for Library/Media Specialists and Teachers.* Santa Barbara, Calif., and Oxford: ABC-CLIO, 1989. 219p. spiralbound, $28.50 (ISBN 0-87436-516-3).

Information Horizons: The Long-Term Social Implications of New Information Technologies. Ed. by Ian Miles and others. Aldershot, Eng., and Brookfield, Vt.: Edward Elgar, 1988. 303p. $42.75 (ISBN 1-85278-041-X).

Informetrics 87/88: Select Proceedings of the First International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval, Diepenbeek, Belgium, 25–28 August 1987. Ed. by Leo Egghe and Ronald Rousseau. Amsterdam and New York: Elsevier, 1988. 329p. $92 (ISBN 0-444-70425-6).

Library and Information Science in China: An Annotated Bibliography. Comp. by Karen T. Wei. Bibliographies and Indexes in Library and Information Science, no.3. New York: Greenwood, 1988. 273p. $39.95 (ISBN 0-313-25548-2).

McCue, Janice Helen. *Online Searching in Public Libraries: A Comparative Study of Performance.* Metuchen, N.J., and London: Scarecrow, 1988. 272p. (ISBN 0-81808-2171-0).

Markey, Karen, and Diane Vizine-Goetz. *Characteristics of Subject Authority Records in the Machine-Readable Library of Congress Subject Headings.* Research

Report Series. Dublin, Ohio: OCLC, 1988. spiralbound, 164p. $14.75.

Olson, Nancy. *A Manual of AACR2 Examples for Microcomputer Software with MARC Tagging and Coding*. 3d ed. Lake Crystal, Minn.: Soldier Creek, 1988. 75p. spiralbound, $17.50 (ISBN 0-936996-34-X).

Planning in OCLC Member Libraries. Ed. by M. E. L. Jacob. OCLC Library, Information, and Computer Science Series, no.9. Dublin, Ohio: OCLC, 1988. 133p. paper, $16.50 (ISBN 1-55653-051-X).

Saffady, William. *Optical Disks vs. Micrographics As Document Storage and Retrieval Technologies*. Westport, Conn., and London: Meckler, 1988. 106p. paper, $27.50 (ISBN 0-88736-345-8).

Samuels, Alan R. *Essential Guide to the Library IBM PC: Shareware for Library Applications*. Essential Guide to the Library IBM PC, v.10. Westport, Conn.: Meckler, 1988. 234p. spiralbound, $24.95 (ISBN 0-88736-184-6).

Schuyler, Michael. *Now What? How to Get Your Computer Up and Keep It Running*. New York and London: Neal-Schuman, 1988. 184p. paper, $29.95 (ISBN 1-55570-022-5).

Schwartz, Ruth. *Multicampus Libraries: Organization and Administration Case Studies*. Metuchen, N.J., and London: Scarecrow, 1988. 262p. $27.50 (ISBN 0-8108-2171-0).

Swersey, Patricia Johnson. *Essential Guide to the Library IBM PC: Spreadsheets for the IBM: A Librarian's Guide.* Essential Guide to the Library IBM PC, v.6. Westport, Conn.: Meckler, 1987. 211p. spiralbound, $19.95 (ISBN 0-88736-047-5).

Tenopir, Carol and Gerald Lundeen. *Managing Your Information: How to Design and Create a Textual Database on Your Microcomputer*. Applications in Information Management and Technology Series. New York and London: Neal-Schuman, 1988. 226p. paper, $37.50 (ISBN 1-55570-023-3).

Weiskamp, Keith, and Namir Shammas. *Mastering HyperTalk*. New York: Wiley, 1988. 506p. paper, $24.95 (ISBN 0-471-61593-5).

Wheeler, Helen Rippier. *The Bibliographic Instruction-Course Handbook*. Metuchen, N.J., and London: Scarecrow, 1988. 626p. $59.50 (ISBN 0-8108-2131-1).

Winter, Michael F. *The Culture and Control of Expertise: Toward a Sociological Understanding of Librarianship*. Contributions in Librarianship and Information Science, no.61. New York: Greenwood, 1988. 154p. $37.95 (ISBN 0-313-25537-7).

Wygant, Alice Chambers, and O. W. Markley. *Information and the Future: A Handbook of Sources and Strategies*. New York: Greenwood, 1988. 189p. $37.95 (ISBN 0-313-24813-3). ■■

INDEX TO ADVERTISERS