Volume 12,
Number 3

ISSN 0730-9295

# Information Technology and Libraries

## September 1993

CONTENTS



12(3) 305–376 (1993); Inf Tech Libr

ISSN 0730-9295

# Information Technology and Libraries

Volume 12, Number 3: September 1993

CONTENTS

# 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 88 of the March 1992 issue. Manuscripts of articles, communications, and news items should be addressed to the editor: Thomas W. Leonhardt, *Information Technology and Libraries,* Bizzell Library, 401 W. Brooks, University of Oklahoma, Norman, OK 73019-0528; (405) 325-2611; Bitnet QC6305@UOKMVSA. Copies of books for review should be addressed to: Susan B. Harrison, *ITAL Book Reviews,* The New York Public Library, 455 Fifth Ave., New York, NY 10016. Copies of software for review should be addressed to: George S. Machovec, *ITAL Software Reviews,* Colorado Association of Research Libraries, 3801 E. Florida Ave., Ste. 370, Denver, CO 80210-2542. Advertising arrangements should be made with Stuart Foster, Advertising Sales Manager, and Dolores LaPointe, Advertising Coordinator, c/o *Choice,* 100 Riverview Center, Middletown, CT 06457; (203) 347-6933.

As a matter of policy, *Information Technology and Libraries,* as the scholarly organ of LITA, does not review LITA publications. Notice of new publications from LITA will generally be found in the "Other Recent Receipts" column following reviews.

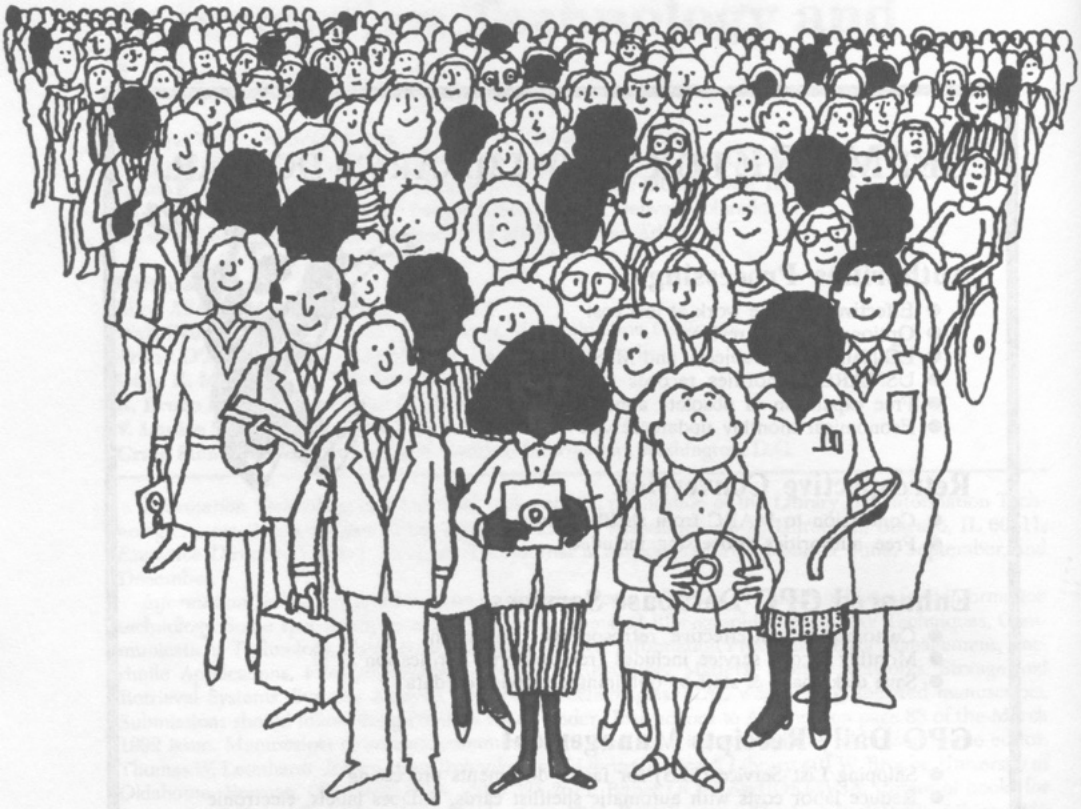
*Information Technology and Libraries* is a perquisite of membership in the Library and Information Technology Association. Subscription price, $22.50, is included in membership dues. Nonmembers may subscribe for $45 per year in the U.S.; $50 in Canada, Mexico, Spain, and other PUAS countries; $55 in other foreign countries. Single copies, $15. 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: ALA Production Services (Eileen Mahoney; Dianne Rooney; Bruce Frausto, Josephine Sharif, Dan Lewis, Suzanne Roe, and Donavan Vicha), American Library Association, 50 E. Huron St., Chicago, IL 60611.

Publication of material in *Information Technology and Libraries* does not constitute official endorsement by the Library and Information Technology Association or the American Library Association.

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, New Search,* 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 239.48-1984. ∞



# SETTING STANDARDS IS EVERYONE'S BUSINESS.

Establishing effective technical standards. It makes good business sense, no matter what business you're in. Because clear, cohesive practices and procedures help things run smoothly.

Which is why we need your help.

**Who are we?** We're NISO. The National Information Standards Organization. We develop technical standards for the entire range of information services and products: online information services, librarianship, bookselling, indexing, and library equipment and supplies. In fact, our standards are an integral part of many work environments.

NISO has developed over fifty information standards. Here are just a few of the standards that benefit you.

- Standard identifiers for books and serials
- Standard format and vocabulary for information sharing
- Standard computer-to-computer protocol linking local, regional, national, and international bibliographic systems
- Standards for permanent and durable book papers and bindings

**How can you help us?** By helping NISO set the standards in your profession. Because only with the input of successful organizations like yours can we develop proper standards. We also need your help with funding. You see, we depend on the support of our members. Members from all types of organizations including libraries, publishers, large and small companies, and government agencies. There are a range of membership options available; one is right for your organization.

**What's in it for you?** A chance to influence your profession's direction and growth by setting its standards. A chance to work closely with your peers. A chance to plan for the new standards that will directly affect you.

**What should you do now?** Call, write, or fax us. We'll tell you more about ourselves, how we can work for you, and how you can help us.

NISO

P.O. Box 1056 • Bethesda, MD 20827
Telephone: 301-975-2814 • Fax: 301-869-8071 • Bitnet: NISO@NBSENH

# A Better Way To Search Databases

We started in 1985, database searchers committed to better search software design. We became the premier vendor of Medline, then expanded our catalog to other databases. Last year we won Information World Review's PRODUCT OF THE YEAR for faster, easier search software. But a better way means meeting the evolving needs – individual and campus wide – of today's library users.

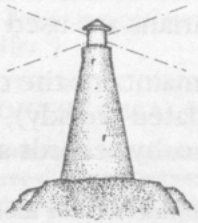
Announcing OVID: a database interface so flexible it molds itself to your search environment.

*With OVID you're free to move from one operating system to another without retraining. OVID's Common User Interface assures identical functionality in DOS, Windows and UNIX.*

*A haven for beginners, OVID's Easy Mode has on screen prompts. The more experienced can pull-down menus showing an array of search options. Experts will feel at home using online syntax.*

*Search with natural language if you like. OVID mapping cuts through the mystery of controlled vocabularies, homing in on precisely matching subject headings.*

*There's an unprecedented array of search tools – indexes, thesauri, limits and fields – many never before available in an interface. They're all standard OVID features.*



*HELP for every search function is context-sensitive and on screen, never more than a keystroke or mouse click away.*

**OVID. A better way to search ERIC, Current Contents, PsycINFO®, Medline, Readers' Guide Abstracts, EMBASE and more.**

**CD PLUS Technologies**

New York 800-950-2035/212-563-3006 London 44-(0)81-748-3777 Amsterdam 020-672-0242

# Filing, Filtering, and the First Few Found

**Michael K. Buckland, Barbara A. Norgard, and Christian Plaunt**

*Card catalogs are dominated by the alphabetic arrangement of the catalog records. Alphabetic arrangement has been carried over to online catalogs, with some unfortunate consequences. Alphabetic order is reconsidered in relation to the purposes of catalogs and differences in catalog technology. Document ranking, subset ranking, and adaptive filtering are examined as alternatives in online catalogs to displays of catalog records in alphabetic order of main entry.*

It has been customary to arrange catalog records alphabetically in book-form catalogs, card catalogs, and now, online catalogs. There has been some disagreement over details of filing concerning, for example, the treatment of numerals, modified and non-Latin letters, and the choice between "letter-by-letter" and "word-by-word" alphabetization. "Structured" or "categorical" arrangement, departing from strict alphabetization, has been used in some special cases, notably in the arrangements for prolific authors, of historical subdivisions, and in the past, for parts of the Bible. A "structured" approach arranges headings based on the categories to which they belong. For example, period subdivisions are arranged chronologically rather than alphabetically. Nevertheless, alphabetization has been the dominant and pervasive principle.

In the United States there has been a marked preference for providing subject access by alphabetized verbal subject headings rather than by a classified subject catalog. The use of verbal subject headings and the dictionary catalog that then becomes possible increases the prominence (and complexity) of alphabetical ordering.

With some exceptions, the alphabetization of catalog records has been carried over to the display of records in online catalogs. Alphabetic ordering by main entry has become standard, whereas dictionary catalog arrangement has not. The necessity and benefits of alphabetizing catalog records by main entry in online catalogs are less certain than in card files. Feasible and attractive alternatives will be discussed below, but first the issues underlying alphabetic ordering will be reviewed. (For a good, detailed discussion of alphabetical and structured arrangements of subject headings in online catalogs see *Headings for Tomorrow.*[1] For an older, more general discussion see *Beginning, of Course, with A*, by Robert Helfer.[2])

## THE FUNCTIONS OF THE CATALOG

The term *retrieval* is ambiguous in that it can subsume three distinct functions:

1. selecting (or identifying) documents (as in subject or author searching);
2. finding a document (or a record for a document) with known individual characteristics ("known item searching"), usually, but not necessarily, to ascertain the location of that document; and
3. fetching (delivering) documents.

Since the nineteenth century, library catalogs have been designed to support selecting in addition to the finding (locating) the function of a catalog.[3] Like bibliographies, modern catalogs are designed to enable the library user to identify or select items in the collections on a particular subject, by a particular

Michael K. Buckland is Professor and Barbara A. Norgard and Christian Plaunt are doctoral students at the School of Library and Information Studies, University of California, Berkeley.

author, and, to some extent, with other characteristics, such as form. However, the distinctive function of a library catalog is as a finding list that enables the library user to ascertain the location (call number) of specific items in the collection once the identity of the item is known. If it did not provide the shelf location of items, it would not be considered a catalog.

Alphabetic ordering is, in essence, a very effective device for locating words, usually names: names of people, names of places, names of things, and names of subjects. Alphabetic ordering is effective in the card catalog because as a "manual" system it enables the determination, both for filing and for finding, of the proper location of headings, subheadings, and individual records filed under headings and subheadings. It is as a result of this hierarchical process of locating that searching becomes feasible. Even with a subject catalog on cards arranged in classified order, an alphabetically ordered "relative index" of the names of subjects is needed for effective use. An obvious but significant qualification is that alphabetic ordering becomes useful only as the number of items in any set increases. Where there is one or very few items, alphabetic ordering is of little or no value.

Effective though alphabetic ordering may be for locating names, including the names of subjects, it is a relatively weak technique for the selecting function because attributes of records other than names are not well served by alphabetic order. Names are not the only significant attribute, and alphabetizing may yield an ordering of limited usefulness. Filing by date is better for currency and a classified arrangement, as on the shelves, is better for bringing related subject materials together. As Ranganathan states: "If the names of the classes, in a natural language, are used to arrange them, we do not get a helpful order. In fact, names scatter classes in a most unhelpful chaotic order. It will give us an order like algebra, anger, apple, arrogance, asphalt, and astronomy."[4]

Selection involves the principle of sorting (or partitioning) the file of catalog records into at least two sets: those selected and those not selected. The traditional approach to the evaluation of retrieval system performance, using recall and precision as measures, is based on how closely the partitioning achieved by the selection process matches an ideal partitioning based on the perceived relevance and nonrelevance of the individual documents.

## ONLINE SEARCHING

Although the objective of retrieval in the sense of selecting is supposed to be the identification of "relevant" items, relevance is a subjective and situational phenomenon. In practice, therefore, what all mechanical systems retrieve, to some measurable degree, are the records matching a specified objective attribute, such as subject heading or word occurrence. A virtue of online catalogs capable of searching keywords and performing Boolean (i.e., postcoordinate) searches is that they can be used easily and very flexibly to create partitions of the file of records in an astronomical number of combinations on an ad hoc basis. Indeed, one of the difficulties with current online catalogs is that the size of the retrieved set is difficult to predict: For 30 to 40 percent of searches zero records are retrieved, and in the other searches hundreds or even thousands of records are commonly retrieved.[5,6] It is within these retrieved sets that alphabetic ordering by heading, traditional in card catalogs, has been continued as the conventional ordering. The catalog may report, say, six hundred records found and is normally programmed to sort the records into alphabetic order of main entry in order to display the first screenful.

There is, however, no logical or technological necessity for the partitioning to be limited to only two sets. There could be any number of sets. One view, strongly held in some circles, is that partitioning should be taken to the limit, with as many subsets as there are records. This yields a ranking of each individual record in decreasing order of some computed measure taken to be an indicator of probable relevance ("document ranking"). Of course, in this case each set has a population of just one record, so there is no opportunity for alphabetic ordering. More generally the scope for alphabetic ordering arises to the extent that the number of records within sets is large enough to need ordering for locating specific records. Later we will show that alphabetic ordering is not necessarily to be preferred, even when feasible.

## CARD CATALOGS RECONSIDERED

The technology underlying online catalogs lends itself to the dynamic selection of sets of

records in many different ways. The computer could also be programmed to sort any set of retrieved records for display in any definable order. A characteristic of card, book-form, and microform catalogs is the stability of the ordering of the records. A strength of card catalogs compared with book-form and microform catalogs is that individual records can be continuously added and removed—and even rearranged. Nevertheless, although feasible in principle, one would not want to rearrange the order of the cards. To subdivide a dictionary catalog into separate name, title, and subject heading files or to rearrange cards under each subject heading from alphabetic to, say, reverse chronological or call number order would be possible but would be a tedious and expensive undertaking. One could not afford to switch lightly from one order to another, let alone on an ad hoc basis. Instead, holes in the cards and metal retaining rods are used to restrain the cards to the chosen ordering.

Although card catalogs do not permit the creation of arbitrary sets in the same way that online catalogs do, some of the same capability exists in ways that may not be obvious:

1. The set of records filed under any one name, title, or subject heading would be the same as the set retrieved using the same name, title, or subject heading used as an exact search in an online catalog;

2. Subdivisions of *Library of Congress Subject Headings (LCSH)* are used to create a multiplicity of smaller subsets. The practice of adding subdivisions to create complex *Library of Congress Subject Headings* has the same kind of effect, achieved prior to the search (precoordination), as postcoordinate searches, which use subject keywords using a Boolean AND;

3. Boolean AND searches can be performed in a card catalog, if one term is an established heading, by scanning the cards filed under that heading for the attributes specified in the rest of the AND statement;

4. Similarly, Boolean AND NOT searches can be handled, if the first term is an established heading, by scanning under the heading containing the first attribute and rejecting those records that contain the second; and

5. Boolean OR searches can be achieved in a card catalog by looking under two different headings, at two disjoint sets.

Nevertheless, for all that one may argue that card catalogs have, in principle, some of the functionality associated with online catalogs, the dominance of a single, linear alphabetic ordering remains a basic difference. Alphabetic ordering is needed not only to locate the records for known-item searching and to locate the headings and subheadings needed for selecting (e.g., subject searching), but also to determine the proper location in which to file each record in the first place. Since it is a human searcher who must locate each record, alphabetic ordering has emerged, with little challenge, as the preferred ordering not only for headings but also for the subset of records under each heading or subheading. Alternatives, such as arranging cards under each heading in chronological order,[7] reverse chronological order to present the searcher with the most recent material first, [8-10] or, in special cases, categorical arrangements, have merits, but alphabetic ordering is almost universally adopted. (For the slow development of alphabetizing, which was little used before printing, see Daly.[11])

## ORDERING RECORDS IN ONLINE CATALOGS

The prevalent alphabetic ordering of catalog records in card catalogs has been imposed on online catalogs, which typically store their records by an identification number based on the order in which records were received. The change of technology from card to electronic means that alphabetic ordering has now become a matter of choice rather than of practical necessity: The computer could be programmed to provide any one or combination of a variety of orderings: chronological, reverse chronological, by language, by proximity of holding library, by loan status, or by any attribute in or implicit in the data available to the computer. In brief, in online catalogs, unlike card catalogs, records do not have to be presented in alphabetic or, indeed, in any other order. Further, no one single consistent choice needs to be made because each set of records retrieved has to be newly ordered for display. The retrieved records could as well be arranged (and rearranged) in any of a variety of ways at choice.

More generally, and perhaps, more importantly, the flexibility of an online catalog to create retrieved sets at will also enables the online catalog to create *sub*sets at will from the retrieved set. The sorting can range from

the standard form of two sets—retrieved and not-retrieved—to the newer orthodoxy of strict document ranking. Better still, online systems can be made to subdivide a retrieved set into smaller subsets or create aggregations of closely ranked records. These subsets or aggregations, as well as the records within them, can be arranged at will in several different ways using whatever attributes are present in the records. A form of filtering is achieved. Choosing some other ordering first to create subsets prior to alphabetic ordering can make a substantial difference, as the examples below will show.

Meanwhile, it should be noted that partitioning, sorting, ordering, and filing are logically the same operation. Any sorting or ordering, as in alphabetic filing, is an example of partitioning. Since the locating and selecting functions of retrieval are achieved by partitioning, it follows that online catalogs (and online bibliographic systems in general) should be viewed not as performing a single retrieval process, but rather as performing a series of consecutive retrieval processes. In a typical online catalog, the first retrieval process establishes (partitions off) the set to be retrieved. This first retrieval process may well be iterated one or more times. What may be overlooked is that, after records have been selected, there is always another retrieval process that orders (partitions) the retrieved set into some specified arrangement for display, usually, but not necessarily, alphabetic order of main entry.

The significance, even the existence, of this secondary retrieval process is masked in two cases. Firstly, the primary retrieval process may be allowed to preempt the second, as when retrieval techniques based on strict document ranking are adopted. Even here it would be a feasible option to allow the aggregation and rearrangement of sets of closely ranked documents should one wish. After all, one's confidence in the ranking depends on how well the criteria used for ranking represent one's interests, and the initial search may not have reflected all of one's preferences fully. Secondly, the ordering of records for display is easily overlooked if it is in (or a practice simply carried over from) a card catalog, where it is, in effect, a necessity.

The choice, and flexibility in choice, of ordering of records becomes more important for an online system than in a card catalog because of the mechanics of displaying records online. Users generally follow the default display option: They normally start by inspecting the first screen of records as selected for presentation by the system. Typically, alphabetic order of main heading is the only option provided unless the user modifies the search itself. Users may or may not choose to page through successive screens of records. The consequence is that the first few records displayed can be expected to be the basis upon which the user will decide whether to display more, modify the search, attempt a new and different search, or abandon searching.[12,13] The basis provided for this decision by conventional online catalog design is, therefore, influenced by how main entries happen to be spelled.

## FILTERING

Bibliographic retrieval is normally a matter of searching. Rather unpredictable searches are made of a relatively stable set of records. Some other document retrieval systems, notably the screening of electronic mail and the analysis of electronic news services, are more a matter of filtering: Rather unpredictable flows of documents are compared with relatively stable search specifications. The secondary retrieval process of ordering retrieved sets of bibliographic records for display can be viewed as a form of filtering.

The following examples were created as a by-product of research on the design of online catalogs at the School of Library and Information Studies, University of California, Berkeley. A problem being addressed in a project titled "Prototype for an adaptive library catalog"[14] is the tendency of existing online bibliographic systems to yield too many records. The design ideal is that regardless of the query and regardless of the size and contents of the file, each search will ordinarily yield a small set of records, those most closely matching the searcher's preferences. This goal requires that the system know what preferences of the searcher should be used to reduce large sets. Implementation of the prototype (Otlet's Adaptive Search Information Service [OASIS]) is in the form of a workstation (DECstation 5000/200) acting as a front end to the University of California's MELVYL online catalog containing some seven million titles held in the libraries of the nine campuses of the university. The workstation ordinarily

*Table 1. Retrieved Set Analysis: Dresden.*

| | Location | | | | |
|---|---|---|---|---|---|
| | Berkeley | | Other Campuses | | |
| | English | Other | English | Other | Total |
| 1990–1991 | 0 | 5 | 2 | 3 | 10 |
| 1980–1989 | 9 | 37 | 6 | 48 | 100 |
| 1970–1979 | 3 | 25 | 12 | 54 | 94 |
| 1960–1969 | 5 | 26 | 12 | 42 | 85 |
| 1950–1959 | 0 | 15 | 0 | 14 | 29 |
| 1900–1949 | 6 | 28 | 8 | 39 | 81 |
| 1800–1899 | 1 | 9 | 4 | 18 | 32 |
| 1700–1799 | 0 | 4 | 0 | 3 | 7 |
| 1600–1699 | 0 | 0 | 0 | 1 | 1 |
| –1599 | 0 | 0 | 0 | 1 | 1 |
| Total | 24 | 149 | 44 | 223 | 440 |

serves transparently, like a dumb terminal connected to MELVYL, but through preprocessing and postprocessing, capabilities not available in MELVYL can be demonstrated. A new command FILTER downloads a set of records from MELVYL into the workstation and then sorts them according to the user's preferences. (This is an automated extension of what skillful use of search modifiers can achieve.[15])

In the first example, a subject keyword search for "Dresden" in the MELVYL catalog retrieves over 440 records (FIND SUBJECT DRESDEN, searched December 29, 1991). The 440 records are unlikely to be of equal interest to any catalog user. Browsing 440 records becomes tedious, and alphabetic ordering by main entry is very unlikely to present anyone's preferred records first. In postprocessing the workstation has been set to make three assumptions about the user's preferences: That, other things being equal, the user would tend to prefer recent material to older; English language works to foreign; and books held in the Berkeley campus to books held only on other campuses. These three preferences, operating on relevant fields in the downloaded records, are used to sort the records into a three-dimensional array of subsets. Table 1 is a tabular analysis of the retrieved set. The searcher may choose to "zoom in" on the last decade (see table 2).

Any subset (the contents of any cell) may be selected for inspection, but a default display could reasonably start, in either table, with the subset in the top left corner since that represents the preferred value for all three preferences. Thereafter, the searcher can inspect other subsets. The default ordering (ranking) of subsets would reflect the relative importance of the preferences. For example, if location mattered more than language and language more than recency of publication, extending the display would, in this case, proceed cell by cell from top to bottom in the left-most column, then down the each column, from left to right successively. Records are alphabetized within each subset containing two or more records. Even though the same pool of records is being drawn from, the effects of such filtering can be striking, especially if, as is common, only the first few records of any large retrieved set are inspected.

As an example of the effects of filtering, imagine that someone on the Berkeley campus wants a general work on chemistry. Consider and compare the following examples of the first few records found with and without filtering. A search for general works on chemistry in the MELVYL catalog would be correctly expressed as FIND XSU "CHEMISTRY" (i.e., an exact search for the Library of Congress Subject Heading "Chemistry," without any subheadings or truncation). MELVYL retrieves 1,494 records

*Table 2. Retrieved Set Analysis: Dresden, 1981 to Date.*

| | Location | | | | |
|---|---|---|---|---|---|
| | Berkeley | | Other Campuses Only | | |
| | English | Other | English | Other | Total |
| 1990–1991 | 0 | 5 | 2 | 3 | 10 |
| 1989 | 0 | 3 | 0 | 1 | 4 |
| 1988 | 1 | 7 | 0 | 3 | 11 |
| 1987 | 2 | 5 | 1 | 3 | 11 |
| 1986 | 2 | 4 | 0 | 7 | 13 |
| 1985 | 1 | 5 | 0 | 11 | 17 |
| 1984 | 0 | 4 | 1 | 2 | 7 |
| 1983 | 1 | 3 | 2 | 7 | 13 |
| 1982 | 2 | 3 | 0 | 4 | 9 |
| 1981 | 0 | 3 | 1 | 5 | 9 |
| Total | 9 | 42 | 7 | 46 | 104 |

and, following standardized alphabetic ordering, displays the following initial selection of records. It is a motley assortment ranging in publication date from 1807 to 1981. Three out of six are not held at Berkeley. Not one seems appropriate as a general introduction to present-day chemistry (searched December 29, 1991; records lightly edited for concise presentation):

1. 31st general assembly of the International Union of Pure and Applied Chemistry, Leuven, Belgium, 1981 : report of the delegation from the National Academy of Sciences. [Washington, D.C.] : National Academy Press, [1982?]. Two holdings, one at Berkeley.

2. Accademia del cimento, Florence.

Essayes of natural experiments made in the Academie del Cimento . . . [1684]. A facsimile . . . New York : Johnson Reprint Corp., 1964. Seven holdings, none at Berkeley.

3. Accum, Frederick, 1769–1838.

System of theoretical and practical chemistry / by Fredrick Accum. Philadelphia : Kimber & Conrad, 1808. [Microprint reproduction]. One holding, at Berkeley.

4. Accum, Friedrich Christian, 1769–1838.

System of theoretical and practical chemistry . . . [2nd American ed.]. Philadelphia : Kimber and Conrad, 1814 ([United States] : Merritt). [Microprint reproduction]. One holding, at Berkeley.

5. Accum, Friedrich Christian, 1769–1838.

System of theoretical and practical chemistry . . . 2nd ed. [London etc.] The author, 1807. One holding, not at Berkeley.

6. Acids and bases; a collection of papers by Norris F. Hall [and others] . . . [Easton, Pa.] The Journal of chemical education [c1941]. Two holdings, neither at Berkeley.

The same 1,494 records, if filtered on the basis of preferences for the English language, for recency of publication, and for items held on the Berkeley campus (as half of the preceding list is not) *before* alphabetic ordering by main entry, yield the following first few. All are held at Berkeley and all are modern, general introductions to chemistry in English published within the previous three years.

1. Chang, Raymond.

Chemistry. 4th ed. New York : McGraw-Hill, c1991.

2. Feigl, Dorothy M.

Foundations of life : an introduction to general, organic, and biological chemistry / Dorothy M. Feigl, John W. Hill, Erwin Boschmann. 3rd ed. New York : Macmillan Pub. Co., c1991.

3. McQuarrie, Donald A.

General chemistry / Donald A. McQuarrie, Peter A. Rock. 3rd ed. New York: W.H. Freeman, c1991.

4. Oxtoby, David W.

Principles of modern chemistry / David W. Oxtoby, Norman H. Nachtrieb. 2nd ed. Philadelphia: Saunders College Pub., c1990.

5. Bodner, George M.

Chemistry, an experimental science / George M. Bodner, Harry L. Pardue. New York : Wiley, c1989.

6. Sherman, Alan.

Chemistry and our changing world / Alan Sherman, Sharon J. Sherman. 2nd ed. Englewood Cliffs, N.J. : Prentice Hall, c1989.

The second set of titles promises a much greater likelihood of user satisfaction in normal library use as a selection of general works on chemistry. All of the titles on the second list could eventually have been found using the first list, but only after examination of well over a thousand records.

In practice, inexpert searchers are more likely to use the simpler search command FIND SUBJECT CHEMISTRY, which generates a subject keyword search and retrieves not only the 1,494 of the previous search but also another 21,000 records for works with any LSCH that includes the word *chemistry*. The first few records displayed for this search are:

1. 1,3-dipolar cycloaddition chemistry / edited by Albert Padwa. New York : Wiley, c1984. Nine holdings, one at Berkeley.

2. 20 Jahre Zentralinstitut für Physikalische Chemie : Festkolloquium . . . Berlin : Akademie-Verlag, 1978. One holding, not at Berkeley.

3. 25 years of structural chemistry in Turku University : from 1,3-dioxane to natural products . . . Helsinki : Suomalainen tiedeakatemia, 1990. One holding, not at Berkeley.

4. 25 years of the Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences. Budapest : [s.n.], 1974. One Berkeley holding, shelved off campus.

5. 31st general assembly of the International Union of Pure and Applied Chemistry, Leuven, Belgium, 1981 : report of the delegation from the National Academy of Sciences. [Washington, D.C.] : National Academy Press, [1982?]. Two holdings, one at Berkeley.

6. 50 [i.e. Piat'desiat] let: Sovetskaia khimicheskaia nauka i promyshlennost'. Moskva, Khimia, 1967. One holding, not at Berkeley.

This selection is more modern than the first list shown, but in this case all of the records are rather exotic works of narrowly specialized or historical interest. Again, half are not held at Berkeley. Not a single one seems to be a suitable find for anyone, especially a nonchemist, simply seeking a general book about chemistry. Yet these are the first few records found by a simple, typical search in an online catalog of standard design. The results are exacerbated by the large size of the MELVYL database, but the catalog of any large, long-established library will tend to yield such results of this kind if standard catalog practice is followed.

Filtering and sorting can, in principle, operate on any data in the catalog record and on anything implicit in the records. For example, a searcher may reasonably want an introductory college-level treatment of a subject. This intellectual level is not coded in catalog records, but it might be inferred. For example, the three undergraduate libraries in the University of California (Berkeley Moffitt Library, UCLA College Library, and the University of California San Diego Cluster Library) maintain collections generally at this level. A stated preference for titles held at one or more of these three libraries but not limited to the copies in those libraries would tend to bring forward general introductory works. Filtering as before but with this additional preference for works selected by one or more undergraduate library applied to the 22,341 records retrieved by FIND SUBJECT CHEMISTRY generates a display of records, all reflecting Berkeley holdings, which begins with:

1. Atkinson, Daniel E.

Dynamic models in chemistry : a workbook of computer simulations using electronic spreadsheets / Daniel E. Atkinson, Douglas C. Brower, Ronald W. McClard. Marina del Rey, Calif. : N. Simonson, c1990.

2. Chang, Raymond.

Chemistry. 4th ed. New York : McGraw-Hill, c1991.

3. Feigl, Dorothy M.

Foundations of life : an introduction to general, organic, and biological chemistry / Dorothy M. Feigl, John W. Hill, Erwin Boschmann. 3rd ed. New York : Macmillan Pub. Co., c1991.

4. Manahan, Stanley E.

Environmental chemistry / Stanley E.

Manahan. 4th ed. Chelsea, MI : Lewis Publishers, c1990.

5. Minkin, V. I. (Vladimir Isaakovich). Quantum chemistry of organic compounds : mechanisms of reactions / V. I. Minkin, B. Ya. Simkin, R. M. Minyaev. Berlin ; New York: Springer-Verlag, c1990.

6. Practicing to take the GRE chemistry test. 2nd ed. Princeton, N.J. : Educational Testing Service for the Graduate Record Examinations Board, c1990.

Again, with this second pair of lists, the difference between the first, conventional ordering and the second, filtered ordering of the same results of the same subject search is remarkable. The contrast indicates that the routine adherence to alphabetic ordering is a serious disservice. Online catalog displays can be customized to provide differently tailored service to different terminals. Our experience indicates that filtering and ordering options can be very useful, especially when they may be chosen dynamically by the individual searcher.

## SUMMARY

Alphabetic ordering is necessary for locating headings, subheadings, and individual records filed in a card catalog. Since only one, fixed ordering arrangement can be supported, alphabetic order has dominated, despite the weakness of this arrangement, for the selecting (e.g. subject searching) role of the catalog.

Alphabetic ordering has been carried over into online catalogs even though the technological constraints are different. Searching can be seen as implying multiple retrieval processes: Partitioning to retrieve and ordering for display. A choice of alternative orderings could be supported, including orderings based on multiple criteria such as date, language, and convenient availability. Subordinating the traditional alphabetic ordering by main entry to other forms of filtering using non-topical attributes can yield useful results. The carrying over of the alphabetic ordering of the card catalog to the online catalog appears to have been done at a high price to users. Future online catalogs and online bibliographies should allow users to specify their own personal preferences as routine filtering devices for ordering the display of any or all of their search results. The result is not only a significant enhancement in service but also an indictment of the traditional dominance of alphabetic ordering.

## ACKNOWLEDGMENTS

This work derives from a project entitled "Prototype for an adaptive library catalog" and was partially supported by the Digital Equipment Corporation, Inc., and by the U.S. Department of Education with HEA IID grant no. R197D00017. The comments and insights of Michael G. Berger and Michael Gorman and the assistance of Ziming Liu are gratefully acknowledged.

## REFERENCES

1. American Library Association, Subcommittee on the Display of Subject Headings in Subject Indexes in Online Public Access Catalogs, *Headings for Tomorrow: Public Access Display of Subject Headings* (Chicago: American Library Assn., 1992).
2. Robert S. Helfer, "Beginning, of Course, with A: A Study of Library Filing and the 1960 Filing Codes" (master's diss., University of Chicago, 1982).
3. Michael K. Buckland, "Bibliography, Library Records, and the Redefinition of the Library Catalog," *Library Resources & Technical Services* 33 (Oct. 1988): 299–311.
4. Shigali R. Ranganathan, *Classification and Communication* (Delhi: University of Delhi, 1951), 34.
5. Karen Markey, *Subject Searching in Library Catalogs: Before and After the Introduction of Online Catalogs* (OCLC Library Information and Computer Sciences Series) (Dublin, Ohio: OCLC, 1984).
6. Joseph R. Matthews and Gary S. Lawrence, "Further Analysis of the CLR Online Catalog Project," *Information Technology and Libraries* 3 (Sept. 1984): 345–76.
7. Karl Uhlirz, "Ueber die Ordnung der Buechertitel im Systematischen Kataloge," *Centralblatt für Bibliothekswesen* 1 (Dec. 1884): 461–67.
8. William S. Merrill, "Order of Books by Date under Subjects," *Library Quarterly* 4 (Apr. 1934): 282–84.
9. Herbert Kleist, "Inverse Time Order and Sub-

ject Filing," *College & Research Libraries* 6 (June 1945): 228–31.

10. Richard H. Werking, Ruby E. Miller, and Jay Whaley, "Rearranging the Subject Catalog at Trinity University," *College & Research Library News* 47 (Jan. 1986): 7–9.
11. Lloyd W. Daly, *Contributions to a History of Alphabetization in Antiquity and the Middle Ages* (Collection Latomus, 90) (Brussels: Latomus, 1967). Also Lynn White, "New Horizons for the Social Sciences," in Princeton University, Graduate College, The Association of the Princeton Graduate Alumni, "Report of the Sixth Conference held at the Graduate College of Princeton University on June 19–20, 1957" (np: nd, pp. 106–15); Robert K. Logan, *The Alphabet Effect: The Impact of the Phonetic Alphabet on the Development of Western Civilization* (New York: Morrow, 1986); Tom McArthur, *Worlds of Reference: Lexicography, Learning, and Language from the Clay Tablet to the Computer* (Cambridge: Cambridge Univ. Pr., 1986).
12. Marcia J. Bates, "The Fallacy of the Perfect 30-Item Online Search," *RQ* 24 (Fall 1984): 43–50.
13. Michael G. Berger, "The MELVYL System: The Next Five Years and Beyond," *Information Technology and Libraries* 11 (June 1992): 146–57.
14. Michael K. Buckland and others, "OASIS: A Front-End for Prototyping Catalog Enhancements," *Library Hi Tech* 40 (1992): 7–22.
15. Michael K. Buckland and Doris Florian, "Expertise, Task Complexity, and the Role of Intelligent Information Systems," *Journal of the American Society for Information Science* 42 (Oct. 1991): 635–43. ■■

# Search Patterns of Remote Users: An Analysis of OPAC Transaction Logs

**Larry Millsap and Terry Ellen Ferl**

*The focus of this study is the search behavior of remote users of the University of California MELVYL Library System, an online public access catalog (OPAC). Transaction logs from randomly selected remote user search sessions are analyzed. Descriptive data on the number and type of searches, choice of search mode and database, number of retrievals, number and type of errors, and use of system HELP facilities are presented. The search data have been cross-tabulated with demographic data on the same group of remote users, collected through an online survey conducted by the authors. Effectiveness of system usage is discussed. A case is made for the desirability of additional heuristics in the catalog portion of the system.*

The MELVYL Library System of the University of California (UC) first became accessible outside the library setting in the mid-1980s. Remote usage has risen steadily since that time and typically accounts for more than one-third of the half-million queries entered in the system each week during peak usage periods.[1] In an effort to understand more fully this growing user population, the investigators undertook a two-part study. The findings of the first part of the study, an online survey of users who accessed the MELVYL system from outside the library setting, were reported previously.[2] The present report contains the results of the second phase of the study. In this phase, the investigators coded selected data from the transaction logs of the surveyed group, used microcomputer programs to compare those data by user status and other user characteristics, then visually reexamined the user command portions of many of the individual logs to gain further insight into user search behavior.

## THE MELVYL LIBRARY SYSTEM

The MELVYL system provides access to nearly eight million monograph and periodical titles held principally by libraries of the University of California.[3] In addition, the system offers its users access to several periodical index databases and serves as a gateway to many other specialized databases and library catalogs. Users may access this rich array of resources directly from their homes, offices, or other sites, through dial-up or networked connections.

The MELVYL system began as a prototype online catalog for the University of California, a nine-campus, doctorate-granting institution, which currently supports a main library on each campus, nearly one hundred branch and specialized libraries across the system, and an enrollment of more than 166,000 students. The system serves as a union catalog, to which the campus cataloging agencies contribute their records. Most of the campus

Larry Millsap is Head, Bibliographic Records Section, and Terry Ellen Ferl is Principal Cataloger, McHenry Library, University of California, Santa Cruz. MELVYL is the registered trademark of The Regents of the University of California.

libraries implemented local online catalogs during the 1980s; these serve as their primary catalogs and as a gateway to the UC union catalog residing within the MELVYL system.

After a decade of development, the catalog portion of the MELVYL system has achieved the status of a second-generation OPAC. It generally reflects Charles Hildreth's hypothetical construct of features that constitute "a qualitative leap of progress over first-generation online catalogs."[4] For example, the system supports keyword access to a variety of fields, explicit Boolean search logic, limiting capabilities, optional and automatic truncation of search terms in some kinds of queries, extensive help facilities (including contextual help screens), and multiple display formats. Some examples of special processing introduced to improve retrieval are the "normalization" of search terms in several fields and the treatment of title words as "exact" titles under certain conditions.[5] Appendix A contains a summary description of system commands and indexes.[6]

Hildreth states that researchers involved in information retrieval generally acknowledge that "today's conventional keyword-indexed, inverted file, Boolean logic search and retrieval systems like BRS, DIALOG, . . . LEXIS-NEXIS (and all second-generation OPACs) are powerful and efficient but are dumb, passive systems which require resourceful, active, intelligent human searchers to produce acceptable results."[7] Indeed, retrieval systems like the MELVYL system are the focus of continued study designed to solve the problems they still pose to the user.[8]

Clifford Lynch, director of the UC Division of Library Automation, has predicted that "the evolutionary descendants of the MELVYL system (and other systems of its generation) will differ in many ways from today's online catalogs. . . . Not only will the user interface and searching algorithms change, but the contents and scope of the information bases to which the system provides access will change also. These future systems will be more heuristic, and will evaluate information and guide the naive user, while still permitting the 'expert' user total and direct control."[9]

The advent of next-generation OPAC development is an exciting prospect, but it is not yet clear how some of the major design challenges will be addressed. Hildreth does not expect a "giant, discontinuous leap forward to the next generation of online catalogs" but, rather, that "progress is likely to be made in small, incremental steps."[10] Lynch and his associates suggest that the evolution of the existing MELVYL system interface and other elements of its functionality will continue to evolve over the next decade.

Because the investigators had a unique opportunity to include demographic data in their analysis of transaction logs, the principal aim of the present study has been to compare the remote use of the MELVYL system by various characteristics of remote users. The factors expected to make the most difference were status and frequency of use of the system. The data support that hypothesis, as reported below in the Analysis section under subject searching and in the Effectiveness of Searches section, where those factors were particularly apparent. Also investigated were the types of problems remote users encountered in trying to make effective use of the system and the degree to which the system helped or failed to help users overcome those problems.

## COMMENTS ON TRANSACTION LOG ANALYSIS

Recently published reports of OPAC transaction log studies by Rhonda Hunter, Steven Zink, Thomas Peters, and Sally Kalin were examined for methodological relevance to the present study.[11] All of these interesting studies, which analyzed actual user OPAC searches, were conducted by public service librarians in academic libraries. Their methodologies varied considerably, but their aims were essentially the same: to collect data relevant to instruction in catalog use or to determine whether changes in on-screen user instructions might improve user success with their respective OPACs. In most of the studies, sampling was employed, printouts of logged data were visually examined and coded, and tabulations of selected data were reported. Only Kalin's study focused directly on remote users.[12] None of the studies incorporated demographic data about the users.

A decade ago, John Tolle noted the difficulty in comparing OPACs through analysis of transaction logs because the logging systems vary as to the information they provide.[13] Even if formats for transaction files were standardized, a suggestion made several years

ago by Lois Ann Colaianni,[14] the considerable design variations among OPACs would continue to pose difficulties to investigators trying to compare user performance across systems. The present study does not escape this problem of comparability, but special effort has been made to focus on data that would tend to be of interest beyond the system being studied.

A principal goal has been to provide basic demographic and search behavior data about patrons who use an OPAC outside the academic library setting and who, therefore, must typically depend on the system itself to solve the problems they encounter. As reported previously, just over half (50.9%) of remote users rarely or never use the MELVYL system at public terminals in a UC library. Even among UC-affiliated remote users, that figure is about 33%. Interestingly, 32.8% of the remote users are either first-time, rare, or monthly users of the system.[15] Since at least one-third of all searches in the system are now conducted by remote users (a phenomenon only a few years in the making), the experiences of this population merit close attention.

Transaction logging data have been collected by the developers of the MELVYL system since its inception in order to monitor system performance, provide information on system usage, and support special analyses like the present study.[16] Statistics extracted from the logging data tell a great deal about system usage but do not prepare one for the sometimes startling observations gained through reading the user command portions of the transaction logs.

Several problems arise in the design and execution of studies of the type reported here. For example, when logs of user commands are to be examined by investigators, the scale of the study is a leading determinant in the design. The study may be confined to a brief, representative usage period, to particular types of transactions, to particular user populations, or to a combination of any of these or other elements. Methodologies that include the review of the actual displays resulting from a user's search typically require the investigator to replicate the search shortly after the logging data are captured.

Some problems arise in the course of visual analysis and interpretation. A user's intent may be clearly reflected in some logs but elusive in others. In the absence of specific information on the user's intent gained through, say, interviews or surveys, judgments about some transactions will necessarily be guarded or absent altogether. The relationship between the number of search retrievals, including zero retrievals, and the associated failure or success of those searches poses yet another methodological problem.[17] The characterization of a specific search session in its entirety as either effective or ineffective can also be problematic.

In the present study, the investigators encountered and dealt with each of the aforementioned design problems. The focus on *remote* users has provided the investigators with two distinct advantages. First, a remote user session is virtually always individual: a single user logs on, conducts searches, then logs off. By contrast, a search "session" at a public access terminal in a library frequently represents the searches of multiple users and may, therefore, be more difficult to analyze.[18] Secondly, remote users are not likely to have received help with their search sessions from library staff, so their individual effectiveness may be more reliably judged.[19]

## METHODOLOGY

As reported in detail elsewhere, a survey instrument in the form of an online questionnaire was presented to every third user who logged on remotely to the MELVYL system through the UC telecommunications network from May 7 to 13, 1991.[20] All transaction logs of those who responded to the online survey were preserved. The investigators reviewed printouts of these logs to determine which data to select for coding and analysis. After several practice coding efforts, a set of coding rules was established, the data coding form was finalized (see appendix B), the search logs of each respondent were coded and entered in a computerized file, and the Statistical Package for the Social Sciences (SPSS-PC+ version 4.0) was used to tabulate the data. These results provided basic descriptive data on the number and type of searches remote users performed (e.g., author, title, subject, zero-retrieval, Boolean, and so forth).

In the initial examination of the logs, the investigators noted that most of the categories they planned to study (for example, number of subject searches) had low numbers of occurrences per session, typically five or fewer. Therefore, ranges of numbers (e.g., 6–10 and

*Table 1. Status of Remote Users**

| Status | UC No. | UC % | Non-UC No. | Non-UC % | Total No. | Total % |
|---|---|---|---|---|---|---|
| Graduate student | 225 | 22.0 | 41 | 4.0 | 266 | 26.0 |
| Faculty | 153 | 15.0 | 51 | 5.0 | 204 | 20.0 |
| Library staff | 44 | 4.3 | 112 | 11.0 | 156 | 15.3 |
| Junior or senior | 71 | 6.9 | 16 | 1.6 | 87 | 8.5 |
| Staff | 55 | 5.4 | 10 | 1.0 | 65 | 6.4 |
| Postdoctoral | 51 | 5.0 | 11 | 1.1 | 62 | 6.1 |
| Other | 17 | 1.7 | 39 | 3.8 | 56 | 5.5 |
| General public | 3 | 0.3 | 44 | 4.3 | 47 | 4.6 |
| Research assistant | 22 | 2.2 | 9 | 0.9 | 31 | 3.0 |
| Freshman or sophomore | 24 | 2.3 | 4 | 0.4 | 28 | 2.7 |
| Programmer | 10 | 1.0 | 10 | 1.0 | 20 | 2.0 |
| Total | 675 | 66.0 | 347 | 34.0 | 1,022 | 100.0 |

*23 surveyed users declined to state either their status or affiliation.

11+) were established for coding higher numbers of occurrences. This facilitated the tabulation and presentation of data and provided more meaningful groupings of data.[21]

The responses to the online survey, which were captured with the logs, provided information on location, affiliation, status, subject interest, frequency of system usage, connect and search help received, and the form of help desired in the future. Table 1, based on data gathered in the first phase of the project, contains status information on the survey respondents.

The file of questionnaire responses was joined with the file of selective data coded from the transaction logs and subjected to various cross-tabulations to test hypotheses about user search patterns. For example, a particular group of users (faculty, graduate students, or first-time users) might be compared by several measures from the coded data (validity of subject search terms, numbers of typographical errors, or use of combined indexes).

It was not feasible to preserve the actual bibliographic displays resulting from the searches. Occasionally the investigators replicated searches to observe the type of retrieval, HELP screens, or other information the user might have gotten in response to a search command.

Nearly 5,000 remote users were offered the online questionnaire during their sessions. About 28% (1,417) of them completed it. Eighty of these cases were eliminated from the study when they were discovered to be in-library rather than remote users. Among the 1,337 remaining users whose logs were eligible for analysis, 292 simply answered the questionnaire without making any further use of the system and, thus, have been excluded from the present study.[22] The number of remote user session logs linked to the questionnaire responses and subsequently analyzed was, finally, 1,045.

During the log review and coding process described above, the investigators identified and later studied more closely many user sessions that illustrated difficulties or successes in using the system. Some of these sessions have been reproduced below to illuminate significant findings.

## FINDINGS

### Number of Searches

The FIND command is the basic search command in the system. Three-fourths of the users issued five or fewer FIND commands per session, with nearly one-third entering only one FIND command (see table 2). Longer sessions of from twenty-six to eighty FIND commands were conducted by only 3% of the users. Twenty-seven users issued no FIND commands in the MELVYL system but

*Table 2. Number of Find Commands*

| | No. | % |
|---|---|---|
| 0 | 27 | 2.6 |
| 1 | 316 | 30.2 |
| 2 | 190 | 18.2 |
| 3 | 96 | 9.6 |
| 4 | 81 | 7.8 |
| 5 | 65 | 6.2 |
| 6 | 38 | 3.6 |
| 7 | 40 | 3.8 |
| 8 | 33 | 3.2 |
| 9 | 26 | 2.5 |
| 10–15 | 73 | 7.0 |
| 16–20 | 30 | 2.9 |
| 21–25 | 11 | 1.1 |
| 26–30 | 5 | 0.5 |
| 31–35 | 2 | 0.2 |
| 36–40 | 4 | 0.4 |
| 41–45 | 2 | 0.2 |
| 46–50 | 0 | 0.0 |
| 51–60 | 3 | 0.3 |
| 61–70 | 2 | 0.2 |
| 71–80 | 1 | 0.1 |
| Total | 1,045 | 100.0 |

employed the system as a gateway to other online library catalogs and specialized databases.

### Databases Searched

The search sessions of 67% of the users contained commands that were executed in the catalog portion of the MELVYL system. A subset of the catalog containing materials published in the last ten years (the Ten-Year database) was searched in 7.9% of the sessions. Searches of the MEDLINE and the CURRENT CONTENTS journal article citation databases—the only databases of that type available at the time of the survey—occurred in 12.4% and 11.2% of the sessions, respectively. The periodicals database (UPE), a union list of holdings for California academic libraries, was searched in 16.2% of the sessions. The CCT database, which contains the table of contents pages for journals indexed in CURRENT CONTENTS, was accessed in only 1.1% of the sessions.

Fewer than 10% of the users issued FIND commands in more than one database, and fewer than 0.1% used more than two databases. Among those who used the MELVYL catalog, 85% used it exclusively.

### Indexes Searched

Title searches appeared in 62.2% of the sessions, author searches (overwhelmingly for personal names) in 38.1%, and subject searches in 23.9%. Only 9.2% of the sessions contained combined index (Boolean) searches explicitly entered by the user. The title searches conducted were principally keyword title searches, but some sessions included "exact" title searches. Table 3 presents index usage by database.

Especially noteworthy is the occurrence of one or more explicit subject searches in 48.2% of the sessions that included searches in the Ten-Year database. This is considerably higher than the 30% rate for sessions conducted in the entire MELVYL catalog. Also notable is the use of explicit combined indexes in twice as many sessions conducted in MEDLINE as those conducted in the MELVYL catalog (19.2% versus 10.1%). Many of the MEDLINE searches were found to be combined author/title searches, a type of search that tends to be burdensome on response time in the very large MELVYL catalog and prohibited there under certain conditions.

### Number of Retrievals

The investigators' principal interest regarding number of retrievals was with data for the catalog portion of the MELVYL system. Among the 616 users who searched the MELVYL catalog, 82% conducted one or more searches that retrieved 1 to 50 citations, 41% performed searches that retrieved 51–200 citations, and 10% got 201–500 retrievals in some of their searches. Only 7% of users conducted searches that retrieved 500–2,000 citations. Less than 1% entered searches that resulted in over 2,000 citations (see table 4).

### Zero Retrievals

Over half of the MELVYL catalog users (57%) conducted one or more searches that resulted in zero retrievals. Those who entered more searches increased their risk of getting zero retrievals. Among 203 users who entered only one search in the catalog, 87% managed to avoid zero retrievals. Among those who

*Table 3. Index Use by Database*

| | CAT | | TEN | | MED | | CC | | CCT | | UPE | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total Number of Users | 616 | | 83 | | 130 | | 117 | | 11 | | 169 | |
| | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % |
| Title Index | | | | | | | | | | | | |
| Used | 369 | 59.9 | 46 | 55.4 | 84 | 64.6 | 79 | 67.5 | 10 | 90.1 | 153 | 90.5 |
| Did not use | 247 | 40.1 | 37 | 44.6 | 46 | 35.4 | 38 | 32.5 | 1 | 9.9 | 16 | 9.5 |
| Author Index | | | | | | | | | | | | |
| Used | 281 | 45.6 | 35 | 42.2 | 49 | 37.7 | 51 | 43.6 | 0 | 0.0 | 24 | 14.2 |
| Did not use | 335 | 54.4 | 48 | 57.8 | 81 | 62.3 | 66 | 56.4 | 11 | 100.0 | 145 | 85.8 |
| Subject Index | | | | | | | | | | | | |
| Used | 185 | 30.0 | 40 | 48.2 | 21 | 16.2 | 14 | 12.0 | 3 | 27.2 | 25 | 14.8 |
| Did not use | 431 | 70.0 | 43 | 51.8 | 109 | 83.8 | 103 | 88.0 | 8 | 72.8 | 144 | 85.2 |
| Combined Indexes | | | | | | | | | | | | |
| Used | 62 | 10.1 | 8 | 9.6 | 25 | 19.2 | 14 | 12.0 | 1 | 9.9 | 8 | 4.7 |
| Did not use | 554 | 89.9 | 75 | 90.4 | 105 | 80.8 | 103 | 88.0 | 10 | 90.1 | 161 | 95.3 |

*Note:* CAT = entire catalog, TEN = latest ten years catalog, MED = MEDLINE, CC = CURRENT CONTENTS, CCT = CC tables of contents, UPE = Union Periodicals Database.

entered two searches, that percentage dropped to 42; of those who did three searches, 30% avoided zero retrievals; and of those who performed four searches, only 12% avoided zero retrievals. One-quarter of those who entered eleven or more searches got eleven or more zero retrievals. Less than one-half of 1% of those with the lengthiest sessions managed to avoid zero-retrieval searches (see table 5).

*Table 4. Number of Retrievals in the MELVYL Catalog*

| | Sessions | |
|---|---|---|
| | No. | % |
| One or more searches with zero retrievals | 356 | 57.7 |
| One or more searches with 1–50 retrievals | 503 | 81.7 |
| One or more searches with 51–200 retrievals | 181 | 29.4 |
| One or more searches with 201–500 retrievals | 65 | 10.6 |
| One or more searches with 501–1,000 retreivals | 26 | 4.2 |
| One or more searches with 1,001–2,000 retrievals | 19 | 3.1 |
| One or more searches with 2,001 or more retrievals | 5 | 0.2 |
| Total Sessions | 616 | |

## ANALYSIS

The kinds of problems remote users had with the MELVYL system can be divided into four main categories. Users had difficulty (1) constructing commands that the system would accept, (2) typing commands correctly, (3) using search terms that resulted in some retrievals but not too many, and (4) choosing the appropriate database. Each of these categories will be discussed below, with examples from the logs that contain the user's search statement and the associated number of retrievals.

### Constructing Commands

To obtain results in the MELVYL system, users must be able to construct valid search statements that include a FIND command, an index name, and search terms.[23] Users omitted FIND commands in only 1.8% of the sessions. System prompts or HELP screens seem to have ameliorated this problem because only three users omitted the command more than once (see table 6). A larger group (10%) used the wrong index name or omitted the index name in their search statements, but most (8.9%) did not do this more than twice (see table 7).

*Table 5. Number of Searches in the MELVYL Catalog by the Number of Sessions with Zero Retrievals*

| | Sessions with Zero Retrievals | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Searches | 0 | 1 | 2 | 3 | 4 | 5 | –10 | 11 or more | Total Sessions |
| 1 | 176 | 23 | 1 | 2 | — | 1 | — | — | 203 |
| 2 | 51 | 51 | 15 | 1 | — | — | 2 | 1 | 121 |
| 3 | 17 | 20 | 12 | 5 | — | 1 | 1 | — | 56 |
| 4 | 7 | 9 | 17 | 13 | 5 | 1 | 4 | — | 56 |
| 5 | 3 | 9 | 9 | 7 | 8 | 1 | — | — | 37 |
| 6–10 | 4 | 9 | 13 | 17 | 17 | 5 | 17 | 1 | 83 |
| 11 or more | 2 | 1 | 2 | 4 | 6 | 4 | 26 | 15 | 60 |
| Total | 260 | 122 | 69 | 49 | 36 | 13 | 50 | 17 | 616 |

*Table 6. Omitted Find Command*

| | No. | % |
|---|---|---|
| 0 | 1,026 | 98.2 |
| 1 | 16 | 1.5 |
| 2 | 3 | 0.3 |
| Total | 1,045 | 100.0 |

*Table. 7. Used Wrong Index Name/Omitted Index*

| | No. | % |
|---|---|---|
| 0 | 941 | 90.0 |
| 1 | 70 | 6.7 |
| 2 | 23 | 2.2 |
| 3 | 5 | 0.5 |
| 4 | 3 | 0.3 |
| 5 | 2 | 0.2 |
| 6–10 | 1 | 0.1 |
| Total | 1,045 | 100.0 |

| Hits | Command |
|---|---|
| 0 | Find Board of Education |
| | Help |
| 0 | SU |
| | Help |

*Figure 1. No Valid Search Statements.*

Figure 1 is the search session of a first-time user at a public library outside California who failed to construct any valid search statements. In response to his first search statement, the user received an error message and instructions to type HELP. When he typed HELP, the system responded that a basic search consists of three parts in the following order:

FIND [index] [search terms]

The system then advised the user that the missing part in his search statement was the index name and that he should be sure that all three parts are included when he retyped his command. A list of possible index names was displayed. The user next supplied only the index name (SU, for subject) instead of the required three-part statement, got zero retrievals, keyed HELP again, and received an appropriate instructive message, after which he ended his session. In his responses to the online questionnaire, this user, a member of the general public, stated he did not expect to need any additional help in the future.

Some users keyed inappropriate index names or invalid combinations of index names. The latter occurred infrequently since fewer than 10% of the sessions included combined index searches.

Figure 2 is a search session conducted in the CURRENT CONTENTS (CC) database by a UC staff member. After issuing four HELP or EXPLAIN commands, the staff member, who rarely uses MELVYL, omitted

| Hits | Command |
|---|---|
| | Start CC |
| | E guide |
| | Explain commands |
| | Set database CC |
| | Help |
| | Explain commands |
| 0 | Find antique recording |
| | Help |
| 0 | Find TW 78 recordings |
| 0 | Find XT 78 rpm recordings |
| 0 | Find JO audio TW 78 recordings |
| 0 | Find SC antique recordings |
| 0 | Find SC sound recordings, TW antique recordings |
| 0 | SC Audio, TW antique recordings |
| 0 | Find XT 78 rpm |
| 1 | Find TW old recordings |
| 0 | Find JO audio engineering society, TW 78 rpm recordings |
| 0 | Find JO audio, TW restoring |
| 0 | Find JO journal of the audio engineering society, TW restoring |

*Figure 2. Index Errors.*

| Hits | Command |
|---|---|
| 0 | F A Allen, Paula |
| | Help |
| 0 | Find PA Allen, Pau |
| 0 | Find PA Allen |
| 130 | Find PA Allen, P |
| 2 | Find PA Allen, Pa |
| | D |
| 22 | Find PA Allen Paula |
| | D |
| | D 18 long |

*Figure 3. Illegal Truncation.*

the index name in his first FIND command. Then he used almost every index that was suggested on the screen that he saw after typing HELP, and he combined journals (JO) and title-word (TW) indexes in CURRENT CONTENTS incorrectly. Had he keyed "Find JO audio engineering society," he would have retrieved more than two hundred citations. Use of TW rather than XT (exact title) in the search statement "Find XT 78 rpm" would have retrieved two instead of zero citations.

While the MELVYL system supplies truncation at the end of a search statement for some indexes (for example, keyword searches of the title index), some users apparently expect the system to apply truncation to individual search terms as some other systems do. The search session in figure 3 illustrates this mistaken assumption. Ultimately this user retrieved the desired record, but the initial search statement contained an invalid index name (A instead of AU, for author), which caused the user to adopt a more tentative approach leading to several zero retrievals. On the way to the desired hit, the user employed "illegal" truncation, keyed a single common word leading to a forbidden long search, then got other large and small retrievals that were not satisfactory. This user, a member of the general public in California who uses the system monthly, expressed the desire for more on-screen help in the future.

| Hits | Command |
|---|---|
| 0 | F PA Navikoff |
| 0 | F PA Navacoff |
| 0 | F PA Navikof |
| 0 | F PA Hanavacoff |
| 0 | F PA Navicoff |
| 0 | F PA Navicough |
| 1742 | F SU Russian authors |
| 0 | F SAT UCD |
| 785 | AT UCD |
| | D short |
| | Show 80 |
| | Display 80 |
| | Display 100 |
| | Display 300 |
| | Display 400 [+ 11 more display commands] |

*Figure 4. Misspellings.*

## Typographical Errors

The investigators readily determined from examining the logs that the most common categories of system-identified errors (first word of command is invalid and FIND command contains invalid index name) were more often due to typographical errors than to a lack of understanding about how the system works. The investigators examined the context of such errors and coded a command or index name error as such only when it was clear that it was not a typographical error.

While misspellings were extremely rare among the surveyed population of remote users, a UC junior or senior in business or management did have a critical spelling problem while searching for the Russian-born author Vladimir Nabokov (see figure 4). After

*Table. 8. Long Searches*

| | No. | % |
|---|---|---|
| 0 | 979 | 93.7 |
| 1 | 47 | 4.5 |
| 2 | 8 | 0.8 |
| 3 | 3 | 0.3 |
| 4 | 5 | 0.5 |
| 5 | 2 | 0.2 |
| 6–10 | 1 | 0.1 |
| Total | 1,045 | 100.0 |

many unsuccessful attempts, the user changed his strategy and tried to catch the elusive Nabokov through the search "F SU Russian authors." But this search retrieved a large number of collections of literary criticism about Russian authors rather than works by Nabokov. It does not seem likely that this student found something useful in this session. He stated that he would like additional help over the telephone in the future.

Commands cannot readily be edited once the user has sent his query to the system.[24] The query is typically reentered, which appears to cause some people to make a series of typographical errors. Figure 5, a session conducted by a UC graduate student working at home, illustrates this problem.

The student also made a display command error that commonly occurs in use of the union periodical file. The default display in that database is a very brief display if a search results in more than one retrieval. When the student saw a title for which he apparently wanted a longer display, he simply typed "50" (the "line" number of that title on the screen) rather than typing the appropriate display command to receive a longer display of item 50. The system does not recognize this as an attempt to select from a list and responds, "A periodical command cannot begin with 50. Please type help for a list of command words." This student reported that he uses the system monthly and does not expect to need additional help in the future.

```
Hits  Command
3239  Find PE stat# scie#
   0  Find PE statitical science
   0  Find PE statuistical science
  53  Find PE statistical science
      D
      50
      D short 50
```

*Figure 5. Typographical Errors.*

Figure 6 is the search session of a staff member at a public library. The user did not seem to realize that he was repeatedly mistyping the word "action." In the user's fourth effort, "words" was processed as a title term, not as part of the index name. This library staff member uses the system weekly and does not expect to need additional help in the future.

### Problems with Search Terms

Constructing search statements that are completely valid does not guarantee desired results. Users of the MELVYL catalog can search with subject terms that do not match the authorized terms in *Library of Congress Subject Headings* and retrieve no citations, or can use terms that are so common that they retrieve more citations than they wish to display. Use of very common subject search terms can also result in searches that the system will not process during peak usage periods.

Within the catalog portion of the system, restrictions are placed on searches that would adversely affect response time. These are typically search statements that consist of one or more search words that appear with great frequency in the database (e.g., "history," "American," "economic") and would require lengthy processing to complete. A user whose search is detected as a potential long search will receive a caution message, may be offered partial results, or may, in peak-load hours, be denied completion of the search and encouraged to reformulate the search statement.

Sixty-six users of the catalog (6.3%) conducted one or more searches that the system detected as long searches (see table 8). In the logging records, the retrieval count for such searches is ordinarily zero, though the user may have entered a fully legitimate search statement (e.g., FIND SUBJECT Art History)

```
Hits  Command
   0  Find title kids guide to social actin
   0  Find title kids guide to social actin
   0  Find title kids guide social actin
   0  Find title words kids guide social action
   8  Find PN Lewis, Barbara
      D
```

*Figure 6. Typographical Errors .*

*Table 9. Validity of Subject Search Terms*

| | No. | % |
|---|---|---|
| Not applicable | 837 | 80.1 |
| No terms were valid | 44 | 4.2 |
| Some terms were valid | 58 | 5.6 |
| All terms were valid | 104 | 10.0 |
| Can't tell | 2 | 0.2 |
| Total | 1,045 | 100.0 |

which, if processed, would have resulted in a very large number of retrievals.

Figure 7 represents the search session of a user who entered legitimate Library of Congress (LC) subject terms that occur very frequently in the catalog database. The result was several long searches the system would not process. Even a BROWSE command using one of the common terms as an exact subject (BROWSE XS TAXATION) led to a long search, but one in which the user was finally rewarded with a partial display of two hundred records.[25] Ultimately overcoming the effects of the system's restrictions and his own typographical errors, this first-time user, a library staff member in a foreign country, expressed the desire for more on-screen help in the future.

## Numbers of Retrievals

The investigators looked more closely at the problem of retrieval size in respect to the MELVYL catalog database. Chandra Prabha has stated that keyword searching and Boolean operations that worked well for small or medium-sized OPACs are now creating a growing problem of large retrievals in very large OPACs. Prabha's exploration of the literature on retrieval size concludes that there is no consensus on what constitutes a large retrieval. Some researchers have suggested that 20 to 30 records would be optimum and any number beyond that would be considered large, while others who have investigated the matter find 100 or more hits to be large or "unmanageable."[26] The *average* number of retrievals per search in the MELVYL catalog is about 100 records, but users display only about fifteen records per search.[27]

As noted above, among the remote users

| Hits | Command |
|---|---|
| 409 | Find SU pollution—economic aspects |
| 0 | Find SU pollution—economic aspects and recent |
| 105 | Find SU pollution—economic aspects and date recent |
| 0 | Find SU taxation and SU pollution and date recent |
| | set database ten-year |
| 0 | Find SU pollution and SU taxation |
| 0 | Find SU pollution—economic aspects and taxation |
| 0 | Find SU pollution—economic aspects and SU taxation |
| 0 | Find SU pollution—economic aspects and XS taxation |
| 0 | Find SU pollution—economic aspects and XS taxation # |
| 0 | Find SU pollution—economic aspects and XS taxation and date current |
| 200 | Browse XS taxation |
| | D counts 27 |
| 9 | Find SU taxation—compliance cost# |
| | D |
| 0 | SU environmental policy—cost effectiveness and date current |
| | Help |
| 0 | Find SU environmental policy—cost effectiveness and date current |
| 2 | Find SU environmental policy—cost effectiveness and date current |
| | D |
| 26 | Find SU environmental policy—cost effectiveness and date recent |
| | Display long |

*Figure 7. Large or Zero Retrievals.*

of the MELVYL catalog surveyed, 82% had one or more searches that retrieved 1 to 50 citations. If 30 citations is the desired threshold, some of these results may be too large. There were 181 sessions (30%) with one or more searches that retrieved 51–200 citations. Clearly these either approached or were well over the threshold that some consider a manageable retrieval size. Nearly one-fifth of the sessions included searches whose results ranged between 201 and 2,001 or more citations. One needs to remember, too, that some of the zero-retrieval searches in the MELVYL catalog are actually searches that would retrieve so many records the system will not process them.

| Hits | Commands |
|---|---|
| 0 | Fi SU grading in classes |
| 663 | Fi SU grading |
| | D |
| 0 | Fi SU grades in education |
| 24 | fi TW grades and classes |
| | D |
| 0 | Fi SU grading and schools |
| 11 | fi TW grading and schools |
| | D |
| 0 | Fi SU grades and education |
| 0 | Fi SU assessment in classrooms |
| 20 | fi TW assessment and higher education |
| | D |

*Figure 8. Subject Term Problems.*

## Subject Searching

This study examined subject term validity in remote user searches conducted in the catalog portion of the MELVYL system. There are several problems with subject searching in the MELVYL catalog. One fundamental problem is the lack of a genuine authority file for topical subjects that appear in the catalog. Consequently, there are no references from unused headings to ones that are used. Responsibility for subject authority control resides with the individual campus cataloging agencies and is not formally coordinated within the MELVYL system. It is therefore possible to obtain search results from use of invalid terms. For example, one user surveyed keyed "FI SU Gyroscope" and retrieved one record. Had he used the authorized term "Gyroscopes," he would have retrieved about 100 records.

About 20% of the sessions contained explicit subject searches that fell within the scope of the study (see table 9). Among those 208 users, 22% used no valid terms, 28% used some valid terms, and 50% conducted sessions in which all the terms they used were valid. The investigators counted as a valid term any subject search word that matched an authoritative word or formed part of an authoritative phrase in the current edition of *Library of Congress Subject Headings*. If a personal or corporate name was searched as a subject and a record was retrieved, the term used was counted as a valid subject search term.

The nationwide survey of online catalog usage conducted in 1982 under the auspices of the Council on Library Resources suggested that searching by subject could be improved if keyword searching were implemented and if users were permitted to browse the authoritative subject index or thesaurus.[28] Keyword searching of subject headings is available in the MELVYL system. It is not possible, though, to access online a full listing of *authoritative* topical subject terms and their associated cross-references. The user may perform a keyword browse of listings of subject terms in the catalog to which bibliographic records are attached, but it is rare for a user to do so. During the survey period, only 0.4% of the commands issued in the catalog by *all* remote users were BROWSE commands. (For in-library users the figure was 0.5%.) Those few remote users who did browse subjects still had problems: 27.2% of those searches had zero retrievals. When EXACT SUBJECT was browsed, the zero retrievals count rose to 54.8%.

The search session in figure 8 was conducted by a UC faculty member who uses the system daily. This user pursues the appropriate subject term to no avail: each of the subject searches results in zero retrievals, with the exception of the second one. The subject term "grading" retrieved a very large number of records, many of which did not pertain to the desired subject. The results of the various title word searches, while yielding smaller numbers of records, did not appear to achieve the desired precision. The title words "grading and schools" retrieved several works on the evaluation of nursing schools and the topic of grades in elementary schools.

The title words "assessment and higher education" retrieved several works on higher education that had nothing to do with grades in academia, the probable subject of interest to this user. Among the retrieved works were, for example, *Palestinian Higher Education in the West Bank and Gaza: A Critical Assessment* and *The Price of Admission: An Assessment of the Impact of Student Charges on Enrollments and Revenues in California Public Higher Education*.[29] It is not clear whether this user ultimately retrieved any useful records. He did express the wish for more on-screen help in the future.

When the use of valid subject search terms was compared by status of user, library staff (not surprisingly) had the most apparent success (see table 10). In 87% of their sessions, some or all of the subject terms

*Table 10. Validity of Subject Search Terms by Status*

| | Not Applicable | | No terms were valid | | Some terms were valid | | All terms were valid | | Can't tell | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Status | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % |
| Junior or senior | 4 | 1.9 | 10 | 4.7 | 9 | 4.2 | 19 | 8.8 | | | 42 | 19.5 |
| Graduate student | 3 | 1.4 | 8 | 3.7 | 9 | 4.2 | 19 | 8.8 | 1 | 0.5 | 40 | 18.6 |
| Library staff | 5 | 2.3 | 3 | 1.4 | 2 | 0.9 | 19 | 8.8 | 1 | 0.5 | 30 | 14.0 |
| Faculty | 7 | 3.3 | 3 | 1.4 | 9 | 4.2 | 7 | 3.3 | | | 26 | 12.1 |
| Other | 4 | 1.9 | 4 | 1.9 | 5 | 2.3 | 5 | 2.3 | | | 18 | 8.4 |
| General public | 3 | 1.4 | 4 | 1.9 | 5 | 2.3 | 5 | 2.3 | | | 17 | 7.9 |
| Freshman or sophomore | | | 1 | 0.5 | 6 | 2.8 | 7 | 3.3 | | | 14 | 6.5 |
| Staff | 3 | 1.4 | 5 | 2.3 | 3 | 1.4 | 3 | 1.4 | | | 14 | 6.5 |
| Programmer | | | 1 | 0.5 | 1 | 0.5 | 5 | 2.3 | | | 7 | 3.3 |
| Research assistant | 1 | 0.5 | 2 | 0.9 | 1 | 0.5 | 1 | 0.5 | | | 5 | 2.3 |
| Postdoctoral | | | | | 1 | 0.5 | 1 | 0.5 | | | 2 | 0.9 |
| Total | 30 | 14.0 | 41 | 19.1 | 51 | 23.7 | 91 | 42.3 | | | 215 | 100.0 |

searched by library staff were valid. The figure for sessions of faculty members was 84%; for graduate students, 77%; and for juniors and seniors, 75%.

Ray Larson's landmark study of the decline of subject searching in an online catalog has led him to speculate that those who might be considered more experienced users (such as library staff, faculty, and graduate students) are less likely to do subject index searches than are undergraduates.[30] This hypothesis is supported by data in the present study. Only 14.6% of the faculty, 17% of the graduate and postgraduate students, and 19.7% of library staff conducted subject searches, while 51.3% of the undergraduates did so. Among daily or weekly users, 17.6% performed subject searches, while 47.9% of first-time users performed subject searches. The figure for first-time and rare users combined was 43.9%.

The investigators' close examination of the logs of all first-time users who performed subject searches revealed little evidence of effectiveness. There were thirty-two relevant sessions in which eighty-eight subject searches were undertaken. Exactly one-half of the searches (forty-four) resulted in zero retrievals; 9 of the sessions had zero retrievals for all of the subject searches. Twenty-one of the eighty-eight searches had 1–50 retrievals (4 of the searches were for personal names, not topical headings), and twenty-three had more than 50 retrievals. The highest number of retrievals was 1,607.

```
Hits  Command
      Set database PE
   0  Find SU attraction
   0  Find SU facial hair
   0  Find social psychology
 133  Find SU social psychology
   0  Find SU social psychology attraction
      D
 133  Find SU social psychology
      Help
      Show search
      D
      36
      Help
   0  Find 36
```

*Figure 9. Incorrect Database Searched.*

### Wrong Database

A relatively small percentage of users (1.2%) appeared to be searching in the wrong database. Typically, they were searching for journal articles in the union list of periodicals

*Table 11. Users of LOOKUP Mode by Status and UC Affiliation**

| | UC | | Non-UC | | Total | |
|---|---|---|---|---|---|---|
| Status | No. | % | No. | % | No. | % |
| Graduate student | 10 | 8.3 | 10 | 8.3 | 20 | 16.7 |
| Junior or senior | 12 | 10.0 | 6 | 5.0 | 18 | 15.0 |
| Library staff | 1 | 0.8 | 15 | 12.5 | 16 | 13.3 |
| Other | 2 | 1.7 | 13 | 10.8 | 15 | 12.5 |
| Faculty | 5 | 4.2 | 9 | 7.5 | 14 | 11.7 |
| Staff | 8 | 6.7 | 4 | 3.3 | 12 | 10.0 |
| General public | | | 10 | 8.3 | 10 | 8.3 |
| Freshman or sophomore | 6 | 5.0 | | | 6 | 5.0 |
| Programmer | 1 | 0.8 | 3 | 2.5 | 4 | 3.3 |
| Research assistant | 2 | 1.7 | 1 | 0.8 | 3 | 2.5 |
| Postdoctoral | 1 | 0.8 | 1 | 0.8 | 2 | 1.7 |
| Total | 48 | 40.0 | 72 | 60.0 | 120 | 100.0 |

* Two surveyed users of LOOKUP declined to state either their status or affiliation.

(UPE) database. Figure 9 contains part of a search session conducted by a UC freshman or sophomore in the social sciences, accessing the MELVYL system from a dormitory. This student, who rarely uses the system, should have searched in the CURRENT CONTENTS journal citation database rather than the UPE database. The student also made the display command error that appears above in figure 5. He did not pursue his effort to obtain a longer display and instead switched to performing subject searches in the Ten-Year database. He stated that he would like more on-screen help in the future.

| Hits | Command |
|---|---|
| 0 | Find KW desensitization, ach, voltage clamp |
| 0 | Find KW desensitization, receptors, ach |
| 0 | Find KW desensitization, phosphorylation, receptors |
| 0 | Find KW desensitization, receptors |
| 0 | Explain Find KW desensitization, receptors, ach |
| 0 | Find desensitization |
| | Help |
| | SC |
| 0 | Find SC desensitization, receptors |
| | Help |
| 0 | Find JO desensitization, receptors |
| 0 | Find JO desensitization |
| | Help |

*Figure 10. Incorrect Database Searched.*

Figure 10 is the search session of a postdoctoral student in the physical or biological sciences, who searched the CCT database from an office on a UC campus. This student, a weekly remote user of the MELVYL system, did get some help on use of the indexes in CCT, but it is clear that he did not understand the scope of this database. His search statements suggest that he should have been searching CURRENT CONTENTS itself, instead of the associated database containing journal contents pages. He made two typographical errors, one in typing "receptors" and one in omitting an index name. However, these errors are immaterial since he could not overcome the problem of searching in the wrong database. He expressed the wish for more printed instructions in the future.

### Use of Menu Mode (LOOKUP)

During the survey period, two search modes were available to users of the MELVYL catalog, COMMAND mode and the menu-driven mode called LOOKUP. LOOKUP mode was selected by 122 (17.8%) of the 685 users who searched either the entire catalog database or the Ten-Year database. Non-UC users accounted for 60% of LOOKUP mode usage from remote sites (see table 11). Among the UC-affiliated users of LOOKUP, 25% were

*Table 12. Effectiveness of Searches*

| | No. | % |
|---|---|---|
| Zero retrievals for all searches | 88 | 8.4 |
| Some retrievals, no displays | 9 | 0.9 |
| Displayed some records | 575 | 55.0 |
| Displayed records, refined searches | 373 | 35.7 |
| Total | 1,045 | 100.0 |

juniors and seniors, 20.8% were graduate students, 16.6% were staff, 12.5% were freshmen and sophomores, and 10.4% were faculty. Among the ninety-five UC undergraduates who participated in the survey, close to one-fifth (18.9%) chose LOOKUP mode to access the catalog.

As expected, those who used LOOKUP mode made significantly fewer mistakes in formulating FIND commands than those using COMMAND mode. LOOKUP users, of course, did not omit the FIND command or combine index names illegally since such errors cannot be committed in LOOKUP. Only 5.7% of the LOOKUP users omitted the index name or used the wrong index name; within the entire survey group, 10% made such errors. LOOKUP users had a slightly lower rate of typographical errors, 13.8% in LOOKUP sessions compared to 16.5% in all sessions.

A much higher proportion of the 122 LOOKUP sessions included subject searches than did the survey population as a whole, 43.1% compared to 19.9%. Among those sessions, 28.3% had no valid terms, 34% had some valid terms, and 37.7% had all valid terms. These percentages represent a lower rate of success in choosing terms than is found in the entire group. LOOKUP mode users also had a higher percentage of long searches: 10.6% compared to 6.3%. However, the general effectiveness of LOOKUP sessions was slightly higher: 5.7% had no retrievals, 59.3% had some retrievals and displays, and 35% had refined searches. Two LOOKUP mode users (1.6%) were in the wrong database. Only one person in this group (0.8%) used HELP and that use was effective.

LOOKUP mode ceased to be an advertised search option on August 15, 1992, and was discontinued altogether on March 31, 1993. Because various limiting options, customized displays for printing and downloading, and a growing number of other information resources on the MELVYL system were not directly available within LOOKUP mode, the UC User Services Group of librarians recommended that the mode be discontinued. Interestingly, during the period surveyed, LOOKUP mode was used in 11.2% of all remote sessions in the catalog and in a remarkable 38.3% of remote sessions in the Ten-Year database.

### Use of Help Facilities

The original specifications for the MELVYL system proposed a guided mode (which later became LOOKUP mode), a tutorial mode, and a command mode. The tutorial mode, which was never developed, was to provide a transition for users who wanted to use COMMAND mode but needed additional instruction. Instead, extensive help facilities were developed over the past decade in an effort to meet the online "tutorial" needs of users. These facilities now consist of over seven hundred messages and contextual HELP screens, plus an extensive EXPLAIN glossary.[31]

The HELP and EXPLAIN screens must be invoked by the user, but the system also displays messages that prompt the user to ask for help under various conditions. When the system detects a command or index name error or a potential long search, a message alerts the user to the situation and suggests use of HELP for more information. HELP is also recommended when a search results in zero retrievals.

The investigators found that this complex of error messages, reminders, and HELP screens, while useful to some of the remote users surveyed, did not adequately rescue users from all of their search difficulties. Slightly more than 80% of the users did not invoke either HELP or EXPLAIN screens during their sessions, though examination of the logs suggests that some kind of help was often needed. Among the 20% who did use HELP, about three-quarters were actually helped. The investigators were generous in assigning a positive value to the use of the HELP and EXPLAIN features. For example, if a user invoked HELP more than once but was not demonstrably helped in each instance, the facility was still credited with being helpful.

## EFFECTIVENESS OF SEARCHES

For coding purposes, the investigators based their judgment about the effectiveness of each session on a single measure that divided the survey group into four mutually exclusive subgroups. Each session log was examined for evidence that the user either (1) had zero retrievals for all searches, (2) had some retrievals but displayed no records, (3) displayed some records, or (4) displayed some records and refined some searches (see table 12). This measure, however rough, was useful in sorting cases for a variety of cross tabulations and further visual examination.[32]

About 90% of the users surveyed retrieved and displayed some records. About 40% within this group refined their searches, typically by using commands that confined searches in the MELVYL catalog to a particular campus library. Other limiting features or search refinements available in the MELVYL catalog were rarely evident. While the MELVYL system automatically applies keyword Boolean search logic for several indexes, it is also possible for users to search with explicit Boolean operands (AND, OR, AND NOT) in COMMAND mode. However, fewer than 10% of the sessions included this type of refinement. The highest incidence of combined index use occurred in MEDLINE searching, where it appeared in about 20% of the sessions (see table 3).[33]

Significant differences in effectiveness of MELVYL system use are evident when one compares the sessions of frequent (daily and weekly) users against those of infrequent (rare and first-time) users. In the survey population there were 714 daily and weekly users and 147 rare or first-time users. Among the frequent users, 50.3% had one or more searches with zero retrievals; 57.8% of the infrequent or first-time users had one or more searches with zero retrievals. Use of combined indexes (a "refinement") appeared in 11.3% of frequent-user sessions, but in only 2% of infrequent-user sessions. Rates for no retrievals and no displays were similar, but 52.7% of frequent users displayed some records, and 38.1% refined searches. For infrequent users, the rates were 61.9% and 27.2%, respectively.

Among frequent users who performed subject searches, 15.4% had no valid terms, 26.9% had some valid terms, and 56.7% had all valid terms. For infrequent users, 22% had no valid terms, 29.6% had some valid terms, and 46.3% had all valid terms. (One session in each group was coded "can't tell.") Only 6.2% of the frequent users used LOOKUP mode; 32.3% of the infrequent users did. The frequent users used the HELP or EXPLAIN features less often but more effectively than infrequent users. Only 14.4% of frequent users used HELP; 73.3% of those who did were helped. Within the infrequent or first-time user group, 36.7% used HELP, but only 62.9% were actually helped.

## CONCLUSIONS

A principal finding in this study is that 40% of the users surveyed conducted sessions that were very brief, typically for known items, with retrievals in each of their searches. This type of searching, which Hildreth calls *query searching,* appears to be well served by the system. The remaining 60% of the users conducted longer sessions. Examination of their logs reveals that their searches were generally less effective and that they sometimes got zero retrievals or high numbers of retrievals, tended to make various types of errors, and encountered system restrictions due to the size of the catalog database. They also encountered what one may call "navigational" problems. One reads the logs and can imagine the user asking himself, where am I? or how did I get this? or what should I do now?

The logs for the longer sessions often reflected the type of searching that Hildreth describes as *browse searching.* With this type of searching, "the search aim is not specific (regarding, for example, discipline or topic, type of publication, level of treatment, perspective, etc.), the desired results are not precisely known in advance, or the correct terms for representing the user's query (which may be vague) are not known at the outset."[34]

A user with diffuse search aims has difficulty with second-generation catalogs because he is repeatedly forced to reformulate and reenter searches until he obtains satisfactory results. Hildreth notes that this sort of OPAC design assumes "the user knows what he wants and can describe it in the language of the catalog database being searched. [Roland] Hjerppe quite correctly rephrases this problem as the fundamental paradox of information retrieval: 'the need to describe that which you do not know in order to find it.'"[35]

One of the original design principles of the

MELVYL system was that the user should remain in control of interactions with the system and understand what the system is doing.[36] This principle, according to Lynch, probably embodies "a bit of wishful thinking," but it has not been abandoned. Nevertheless, Lynch reports, MELVYL system performance problems and user difficulties with very large retrievals eventually led to several changes designed to increase precision in query processing. These measures have helped, but it seems fair to say that even sophisticated clientele (librarians, computer programmers) have difficulty explaining how the system is processing commands.

The findings in the present study support the judgment that a majority of all remote users have difficulty reformulating searches effectively when they do not get what they want. At any one time, a considerable number of these users are accessing the system for the first time or as infrequent users. Interestingly, those who do so have profited from use of the menu-based LOOKUP mode. Discontinuance of this mode seems premature and is very likely to place an even greater burden on the sometimes ineffective HELP and prompt facilities. Some alternative to COMMAND mode seems to be essential for a significant and growing population of users.

The need for more heuristics—design elements that guide or lead the user toward the discovery of desired information—seems obvious. Especially needed are better methods to help the user (1) obtain results following a zero retrieval and (2) review and reduce retrieval sets that are large.[37] While a few heuristics have been introduced in the MELVYL system, largely in response to extensive analysis of transaction logs and other user studies, Lynch notes that their introduction has been attended by much controversy.[38] Some librarians involved in bibliographic instruction on the campuses have tended to be distrustful of heuristics.

Lynch offers a very illuminating description of the ambiguous attitude toward instruction in using the catalog. There appears to be a collective wish to have a system that can be used without instruction, but one that will also reward those who do invest in instruction. Lynch believes that instructors have "a natural tendency to focus on the relatively sophisticated and serious user, perhaps sometimes to the detriment of a user who might feel well served by a more heuristic, actively helpful catalog."[39] One of their counterarguments to implementing more heuristics is that the OPAC user might go away "satisfied" but ignorant of information about additional resources that the instructor can provide.

Virtually all of the studies of transaction logs encountered in the course of this project point to the critical need OPAC users have for more help. Hunter's findings also confirm that "people do not want to take the time to learn to use the online catalog."[40] The retirement of the LOOKUP mode in the MELVYL system is supposed to be compensated for eventually by the introduction of an online tutorial. It remains to be seen whether such a tool would be useful, given the impatience of users to get on with their information seeking without having to learn about complex system commands and other advanced features. In the conclusion to her study, Hunter suggests that the best answer to this problem may be for librarians to abandon trying to develop an overall instruction method and service strategy and instead use their knowledge of searching and learning behavior to help information scientists develop better systems and user interfaces. Hunter's conclusion is particularly pertinent to the remote users in the present study.

The developers of the rather primitive first-generation OPACs were advised by researchers to improve them by adding more advanced capabilities. As a consequence, we find that second-generation OPACs generally provide features such as keyword access to titles and subjects, Boolean logic, full records, multiple display formats, and online assistance. The MELVYL system provides (arguably) the greatest array of advanced features for a system of its scale. However, the transaction logs of remote users and the general statistics for system usage reveal that users overwhelmingly employ basic commands and single indexes. Indeed, the logs for the catalog database demonstrate the need for the machine, not the user, to do the complex work in search formulation and retrieval ranking. The sophisticated features of this advanced second-generation OPAC seem destined to remain the province of librarians and perhaps some users of specialized databases on the system, such as MEDLINE.

It is therefore time to revisit the uneven balance between offline instruction and online heuristics. Evidence strongly suggests that more and better heuristics are needed to serve the greater mass of OPAC users.

## ACKNOWLEDGMENTS

This study was supported by a grant from the Council on Library Resources. The authors thank Lynne Grigsby-Standfill, programmer/analyst for the University of California Division of Library Automation, for her technical assistance in capturing the transaction logs.

## REFERENCES AND NOTES

1. Remote users issued 25% of the FIND commands during the period studied by the investigators, May 7–13, 1991, and 33% of the FIND commands during the week of October 17–23, 1992. During academic breaks and other periods of low instructional activity, remote use frequently rises to 50%.
2. Terry Ellen Ferl and Larry Millsap, "Remote Use of the University of California MELVYL Library System: An Online Survey," *Information Technology and Libraries* 11 (Sept. 1992): 285–303.
3. Statistics published on the system in November 1992 were 7,022,367 book titles and 792,980 periodical titles. Total volume holdings reported at that time were about thirteen million. In addition to holdings of the UC libraries, the MELVYL system also reflects holdings of the California State Library, the California Academy of Sciences, and the Center for Research Libraries (Chicago).
4. Charles R. Hildreth, *Intelligent Interfaces and Retrieval Methods for Subject Searching in Bibliographic Retrieval Systems* (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1989), 6.
5. *The MELVYL System Reference Manual* (Oakland, Calif.: Division of Library Automation, Univ. of California, c1990), which is updated periodically, provides detailed information about the system, its databases, and patron interface.
6. Reproduced, with permission, from *Search Sheets for OPACs on the Internet* (Westport, Conn.: Meckler, 1991), 33–34.
7. Hildreth, *Intelligent Interfaces*, 2.
8. For an extensive review of current research on OPAC user interfaces, see Martha M. Yee, "System Design and Cataloging Meet the User: User Interfaces to Online Public Access Catalogs," *Journal of the American Society for Information Science* 42 (Mar. 1991): 78-98.
9. Clifford A. Lynch, "The Next Generation of Public Access Information Retrieval Systems for Research Libraries: Lessons from Ten Years of the MELVYL System," *Information Technology and Libraries* 11 (Dec. 1992): 414.
10. Charles R. Hildreth, "Beyond Boolean: Designing the Next Generation of Online Catalogs," *Library Trends* 35 (Spring 1987): 652.
11. Rhonda N. Hunter, "Successes and Failures of Patrons Searching the Online Catalog at a Large Academic Library: A Transaction Log Analysis," *RQ* 30 (Spring 1991): 395–402; Steven D. Zink, "Monitoring User Search Success through Transaction Log Analysis: The WolfPAC Example," *Reference Services Review* 19 (Spring 1991): 49–56; Thomas A. Peters, "When Smart People Fail: An Analysis of the Transaction Log of an Online Public Access Catalog," *Journal of Academic Librarianship* 15 (Nov. 1989): 267–73; and Sally W. Kalin, "The Searching Behavior of Remote Users: A Study of One Online Public Access Catalog (OPAC)" in *Proceedings of the 54th Annual Meeting of the American Society for Information Science* 28 (1991): 178–85.
12. Kalin's project aim was "to determine if remote users of LIAS [Pennsylvania State University's online catalog] search any differently than those who search it in-house. Analyzing the searches would identify areas that might be problematic for remote users, and assist the Libraries in developing appropriate support services" (p. 178).
13. John E. Tolle, *Current Utilization of Online Catalogs: A Transaction Log Analysis*, vol. 1 (Dublin, Ohio: OCLC Office of Research, 1983), 107–10.
14. Lois Ann Colaianni, "Evaluating Online Catalogs: The Need for Data," in *The Impact of Online Catalogs* (New York: Neal-Schuman, 1986), 81–87.
15. Ferl and Millsap, "Remote Use," 288–89.
16. For a description of the MELVYL transaction logging system, see Ray R. Larson's, "Transaction Logs: Gathering Information about MELVYL," *DLA Bulletin* 7 (Oct. 1984): 17–18; and Ray R. Larson and Vicki Graham, "Monitoring and Evaluating MELVYL," *In-*

*formation Technology and Libraries* 2 (Mar. 1983): 93–104. For the present study, the UC Division of Library Automation provided the investigators with detailed documents that described the transaction monitoring system and defined the various elements captured in the logs.

17. A zero retrieval in itself is neutral. It must always be viewed in some context to gain significance. For example, a knowledgeable observer may detect that a search word is misspelled, a command word is invalid, a search term does not match the controlled vocabulary of the index being searched, and so forth. Or the observer may note that a succession of searches strongly suggests the user is in the wrong database. Nevertheless, OPAC system designers and researchers remain concerned about large numbers of searches with zero retrievals. In 1988, Clifford Lynch stated that the statistics for zero retrievals in the MELVYL catalog were "alarming." At that time, about 31.5% of MELVYL searches in COMMAND mode resulted in zero retrievals. The figure still remains at that level. For Lynch's concerns on this and related matters, see his "Large Database and Multiple Database Problems in Online Catalogs," in *OPACs and Beyond* (Dublin, Ohio: OCLC, 1989).
18. A MELVYL catalog session will "timeout" (end automatically) after ten to twelve minutes with no commands or actions from a user. This default setting may be changed by a user for any amount of time between five and thirty minutes.
19. Only 1% of the surveyed users responded that they had received help from library staff in conducting their searches.
20. Ferl and Millsap, "Remote Use," 286.
21. The system logging programs, of course, capture total numbers in a myriad of categories. The Division of Library Automation provided the investigators with tables of remote user statistics captured during the survey period, some of which were more detailed than those routinely published on the system. These data occasionally helped inform the conclusions of the investigators.
22. The status profiles of the 292 users in the excluded group mirror those of the group whose logs were studied.
23. Users who chose the menu mode, called LOOKUP, rather than the COMMAND mode, did not have to type the command word FIND, but did have to select and type an index name followed by a search term or terms.
24. The user may modify, replace, or delete characters in a command that was just executed. This capability, however, involves three editing steps, so it is essentially unused by the system's clientele. For a description of the feature, log on to the system and type EXPLAIN MODIFY.
25. The MELVYL catalog permits a BROWSE command in several indexes. Terms are processed as keywords, and the feature may help the user discover authoritative forms of names and subjects.
26. Chandra Prabha, "Managing Large Retrievals: A Problem of the 1990s?" in *OPACs and Beyond* (Dublin, Ohio: OCLC, 1989), 33.
27. Michael S. Berger, "The MELVYL System: The Next Five Years and Beyond," *Information Technology and Libraries* 11 (June 1992): 155.
28. Joseph R. Matthews, Gary S. Lawrence, and Douglas K. Ferguson, eds., *Using Online Catalogs: A Nationwide Survey* (New York: Neal-Schuman, 1983), 177.
29. When the investigator tried searching in several different ways by authorized LC subject terms—"Grading and marking (Students)," "Higher education," "College students, Rating of"—she encountered similar problems of irrelevant retrievals, too many retrievals, zero retrievals, or long searches.
30. Ray R. Larson, "The Decline of Subject Searching: Long-Term Trends and Patterns of Index Use in an Online Catalog," *Journal of the American Society for Information Science* 42 (Apr. 1991): 197–215.
31. Mike Berger, "The Patron Meets the MELVYL Catalog: A Short History of the MELVYL/Patron Interface," *DLA Bulletin* 26 (Spring 1992): 7.
32. In the online questionnaire, the users were not asked whether they had "found what they wanted" or were "satisfied with their session." A yes or no answer to such questions would have added little to the analysis of the logs themselves. Instead, survey questions about help received and anticipated need for future help were asked in order to elicit from the user a general impression about ease of system use.
33. During the survey period, 1.9% of all remote user FIND commands in the MELVYL catalog contained the AT limiter to confine the search to a particular library, and 1.0% contained Boolean operands (AND, OR, AND NOT). In the same period, the figures for in-library users were 3.8% and 0.7%, respectively.
34. Hildreth, *Intelligent Interfaces*, 10–11.
35. Hildreth, "Beyond Boolean," 653. As the source for Hjerppe's comment, Hildreth cites Hjerppe's "Project HYPERCATalog: Visions and Preliminary Conceptions of an Extended and Enhanced Catalog" (Paper presented at IRFIS 6 Conference [International Research

Forum in Information Science] in Frascati, Italy, 15–18 Sept. 1985).

36. Lynch, "The Next Generation," 409–11.
37. Michael Buckland has described several experimental "amenities" that would give the OPAC user a summary analysis of the composition of the retrieved set to help him decide what to do next. Buckland's contention that such amenities are needed and desirable is based in part on unpublished analyses of transaction logs that reveal "unexpectedly low levels of effectiveness in use of the MELVYL system (and probably of online catalogs generally)." See his "Agenda for Online Catalog Designers," *Information Technology and Libraries* 11 (June 1992): 157–63.
38. One example of a heuristic feature in the MELVYL catalog is the following: If the user has zero results from a subject search, an on-screen message suggests that he may get results by reentering his search as a title word search. If the user also invokes the HELP screen offered at that point, additional strategies are suggested.
39. Lynch, "The Next Generation," 410. Findings in a recent study by Bryce Allen and Gillian Allen suggest "that librarians should be careful to distinguish between their own abilities and those of their patrons when designing information systems, bibliographic instruction programs, and similar library services. The abilities of professionals may lead to preferences for sophisticated interfaces, complex instructional programs, or services that assume high levels of cognitive abilities, while their library patrons may find browse interfaces, simple point-of-need instruction, and basic services to be more appropriate to their levels of cognitive abilities." See their "Cognitive Abilities of Academic Librarians and Their Patrons," *College & Research Libraries* 54 (Jan. 1993): 67–72.
40. Hunter, "Successes and Failures of Patrons," 402. ■■

**APPENDIX A. MELVYL SYSTEM COMMANDS AND INDEXES**

| Command Type | Command | Label (Index) | Index Description | Example |
|---|---|---|---|---|
| Search | **f, find** | | | |
| | | **pa** | personal author | **f pa** barth, john |
| | | **ca** | corporate author keywords | **f ca** senate foreign |
| | | **xc** | exact corporate author | **f xc** ford foundation |
| | | **xt** | first words of title | **f xt** sun also rises |
| | | **xt " "** | exact title | **f xt** "letters" |
| | | **tw** | title keywords | **f tw** sun rises |
| | | **ti** | title key and/or exact | **f ti** sun also rises |
| | | **ut** | uniform title | **f ut** hamlet |
| | | **xs** | exact subject | **f xs** algeria-history |
| | | **su** | subject keywords | **f su** motorcycle |
| | | **se** | series title | **f se** ciba symposium |
| | | **pe** | periodical keywords | **p pe** life letters |
| | | **xp, xpe** | periodical first words | **f xp** life and letters |
| | | **xp " ", xpe " "** | periodical only words | **f xp** "life" |
| | **acfind** | | authority control: links variant forms/names | **acfind pa** barth, john |
| Truncate | | **#** | | **f su** journalis# |
| Boolean | | **and, or, and not** | | **f pa** barth, john **and not xt** sot weed |

**APPENDIX A.** Continued

| Command Type | Command | Label (Index) | Index Description | Example |
|---|---|---|---|---|
| Limit (not in **PE** database) | **and** | **and lan** | language | **f pa** shakespeare **and lan** fre ger |
| | | **and date** | by year or | **f pa** barth, j **and date** 1985 |
| | | **and date** | by decade | **f pa** barth, j **and date** 1980s |
| | | **and date current** | within last 3 years | **f pa** barth, j **and date current** |
| | | **and date recent** | within last 10 years | **f pa** barth, j **and date recent** |
| | | **and form** | format (non-books, books, music scores) | **f pa** shaekespeare **and form** video<br>**f pa** mozart, w **and form** scores |
| | **at / not at** | | by campus location | **f xs** hopi **at** ucb ucla |
| | **set** | | limit session's searches by lan, date, form, lib | **set form** nonbooks |
| Display | **d, display** | **short** | brief listing/call number | **d** [default setting] |
| | | **REView** | one-line listing | **d rev** |
| | | **long** | full bibliographic record | **d long** 1–3 |
| | | **[field]** | specify any part/s of record | **d su ti** |
| | | **marc** | machine-readable record | **d** 1,2 **mar** |
| | | **acf** | authority search results | **d acf rev** |
| | | **browse** | show browse results | **d browse** 6 |
| | in PE database: | **short** | UC holdings only | **d sho** 2–4, 7 |
| | | **all** | holdings all 34 libraries | **d** 2–4, 7 **all** |
| | **set** | | change session's display | **set d lon** |
| | **show** | | show all settings | **show** |
| | **reset** | | return to default settings | **reset display** |

**APPENDIX A.** Continued

| | | | | |
|---|---|---|---|---|
| Moving Around | **ns / ps** | | next/previous screen | **ns** |
| | **b,browse** | | scan author or subject (not title) indexes | **b su** john muir trail |
| | **set, select** | | choose heading/s found by browse | **sel** 1,3, 5–9 |
| | **d** | **b [#]** | resume browse display | **d b** 10 |
| | **save** | | choose items for later printing | **save** 3 5 |
| | **use** | | gateway to other opacs | **use** dartmouth |
| Edit | [add on] | **[boolean + index]** | modify search just done | **or tw** computers |
| | **Bac, backup** | | recover last search | **bac** |
| | **show history** | | numbered list of searches | **show history** |
| | **redo** | | recover any search | **redo** 4 |
| | **with** | | add to a **redo** command | **redo** 4 **with and lan** eng |
| Help | **h, help** | | for current situation | **h** |
| | | | list of available help | **h glossary** |
| | **e, explain** | | explain specific topic | **e** browse |
| | **show** | | news, statistics, info | **show** news, **show** UCLA info |

## APPENDIX B. TRANSACTION LOG CODING FORM

Case No._____

**Total number of FIND commands:**

0 = 0
1 = 1
2 = 2
3 = 3
4 = 4
5 = 5
6 = 6
7 = 7
8 = 8
9 = 9
10 = 10-15
11 = 16-20
12 = 21-25
13 = 26-30
14 = 31-35
15 = 36-40
16 = 41-45
17 = 46-50
18 = 51-60
19 = 61-70
20 = 71-80
21 = 81-90
22 = 91-100
23 = 101-200
24 = 201-300
25 = 301-400
26 = 401-500
27 = 501 +

**MELVYL database: Number of searches**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**MEDLINE database: Number of searches:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**CC database: Number of searches:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**CCT database: Number of searches:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**UPE database: Number of searches:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**TEN-YEAR MELVYL database: Number of searches:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Non-MELVYL systems used?**

0 = No
1 = Yes

**FIND TITLE:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**FIND AUTHOR:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**FIND SUBJECT:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Combined indexes in a search:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Searches with zero retrievals:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Searches with 1 - 50 retrievals:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Searches with 51 - 200 retrievals:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Searches with 201 - 500 retrievals:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Searches with 501 - 1000 retrievals:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Searches with 1001 - 2000 retrievals:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Used wrong index name/omitted index name:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**In wrong database:**

0 = No
1 = Yes

**Illegal combination of index names:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Illegal attempt to truncate:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Omitted FIND command:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Long search:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Typographical errors (including commands & index names):**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

**Validity of subject search terms used in MELVYL catalog and Periodical database:**

0 = Not applicable
1 = No terms were valid
2 = Some terms were valid
3 = All terms were valid
4 = Can't tell

**Effectiveness of searches:**

0 = Zero retrievals for all searches
1 = Some retrievals, no displays
2 = Displayed some records
3 = Displayed records and refined searches

**Use of Lookup Mode in MELVYL catalog:**

0 = Did not use Lookup Mode
1 = Used Lookup Mode

**Use of HELP/EXPLAIN features:**

0 = Didn't use either feature
1 = Used feature(s) and it helped
2 = Used feature(s) to no apparent effect

**Display command errors:**

0 = 0 4 = 4
1 = 1 5 = 5
2 = 2 6 = 6-10
3 = 3 7 = 11 +

# An Overview of Applications of Automation to Special Collections: Rare Books and Art Collections

**Bobs M. Tusa**

*Automation traditionally has been viewed as an inappropriate means of control of the unique items in special collections libraries because of its fundamental requirement of standardization. However, it has come to be viewed as an excellent means of control of and access to these items by curators who view the special collections library as a system precisely because of standardization. This study looks at the issue of standardization in the application of computerized automation—specifically to rare books and art objects—and at some recent examples of those applications in both North America and Western Europe.*

Historically, automation has come late to special collections, even later than to libraries generally. The primary reason for this tardiness, which may turn out to be advantageous as computer applications become over time more refined and more economical, lies in the apparent nature of the two entities: special collections materials are by definition rare or unique items that do not on first sight lend themselves readily to the standardization characteristic of computer technology. For example, the benefits derived from the shared cataloging of current materials via national bibliographic database utilities like OCLC and RLIN are not easily applicable to unique copies of rare books, manuscripts, or archives. Furthermore, standards of bibliographic control become irrelevant when applied to works of art and other museum objects.

Additionally, curators of special collections have been characterized, accurately or not, as obstructionists when confronted with the possibilities of automation: "These attitudes [of curators] seem to reflect a kind of institutional parochialism and lack of vision with regard to the role of special collections as a *national* research tool and not just a local resource or private treasure. This parochialism combines in some cases with a competitiveness with other special collections, a general unwillingness to engage in cooperative projects, and a highly developed chauvinism about the importance of their own collections,"[1] according to Steven Paul Davis. On their side, curators and archivists are frequently mystified by the inability of librarians to comprehend that rare books, manuscripts, and archives are not just peculiar entities whose descriptions should be distorted until they fit those of current monographs.

The major issues in the computerization of special collections are fundamentally a single issue, that of standardization. Although seen as a problem by some special collections curators, standardization via computer technology is, according to those curators and archivists cited in this paper, an excellent means of control of the unique materials found in special collections libraries. They share the view that a special collections library, like any organization, is a structure of interrelated systems with repeated functions that lend themselves to automation. Whether the function has to do with acquiring the items

Bobs M. Tusa is Head of Special Collections, Manuscripts and Archives, Prescott Memorial Library, Louisiana Tech University. Her e-mail addresses are TUSA@LATECH.BITNET and TUSA@VM.CC.LATECH.EDU.

in the collection, establishing a form of inventory control, providing access to and information about the collection, administering the collection, or fitting the collection into its place in larger systems, certain aspects of each function lend themselves to the standardization that is fundamental to the computer. The machines and the software already exist in a bewildering plethora of choices, and new ones are being announced every day. The fundamental issue is that of perspective: approaching the special collections library from a systems point of view and then making appropriate decisions about which automated tools to choose in order to accomplish the objectives of the particular special collections library regarding its clientele and resources. The problem is the solution.

## RARE BOOKS

The recent history of the standardization of bibliographic control of rare books dates from the mid-1970s. According to John B. Thomas, III, librarian of the University of Texas at Austin and chair of the Standards Committee of the Rare Books and Manuscripts Section of the Association of College and Research Libraries, "descriptive cataloging codes for preparing machine-readable records for many types of materials (including rare books) began to be created soon after the first presentation by IFLA [International Federation of Library Associations and Institutions] of the *International Standard Bibliographic Description for Monographic Publications* or ISBD(M) in 1973. The impetus for a code for older materials was the attempted and unsatisfactory use of the MARC format in cataloging projects at the Bibliotheque Nationale, the Bodleian, and the National Library of Scotland in the early 1970s."[2]

The inadequacies of the ISBD(M) for bibliographic control of rare books stemmed from its design objective of control of current materials. It made no provision for exact transcription of title, for collation of every page, or for detailed information about facts of publication. The IFLA remedied these deficiencies in 1975 with the release of *International Standard Bibliographic Description for Older Monographic Publications (Antiquarian)* or ISBD(A). Also influential at this time in the development of a United States standard for rare book cataloging was the Library of Congress' *Bibliographic Description of Rare Books*, which incorporated provisions of the ISBD(A) into *AACR2* cataloging rules. Additional fields for genre, provenance, physical aspects, printer, and details of publication were provided for the MARC format at the instigation in the United States of the Independent Research Libraries Association (IRLA) and carried out by the Standards Committee of the Rare Books and Manuscripts Section of the Association of Research Libraries.

The result of the above efforts is that for the first time there exists an internationally agreed-upon standard for bibliographic control in machine readable form of rare materials with its concomitant benefits of increased access by the international scholarly community, exchange of information among collections curators, and rationalized collection development. The key to success to date has been the willingness of rare book catalogers to embrace ISBD(A) and the MARC format, to work to enhance and change them where necessary, and to cooperate in developing national and international standards.

One approach to an internationally available database of rare books records in MARC format is the *Incunable Short Title Catalogue (ISTC)*, a database of fifteenth-century books and other printed materials being compiled at the British Library. The copy-specific records come from American, British, and European union catalogs and represent "over three-quarters of the estimated total of incunable printing, . . . the largest incunable bibliography ever assembled."[3] The *ISTC* permits searching of each MARC field directly and in combination on the British Library's BLAISE-LINE mainframe subscription service; it is planned to make the database available on disk for use on microcomputers and on CD-ROM.

For special collections libraries that lack access to mainframe computers, the microcomputer has provided automated bibliographic control. One such installation, using the commercially available software Pro-Cite, is that in the Map Department at Texas A&M University. Less a map library than a special collection of maps, pamphlets, books, and serials on the topic of travel and tourism, the department explored automation as the solution to its problems of control of a collection of fragile, uncataloged, and unlisted materials that were heavily used. The choice was made to eschew full cataloging in favor of a short-

record database that would be "easy to learn, possible to edit and update regularly, and provide a printout that would be easy for patrons to use."[4] Pro-Cite, from Personal Bibliographic Software for use on IBM PCs, was chosen because of favorable reviews, low price, and the fact that it was the only commercially available software with an option for maps. Pro-Cite uses the MARC format and provides twenty predefined workforms for completing all or selected fields. All fields and records are variable length. Records may be selected with Boolean operators, sorted, formatted according to a variety of bibliographic styles, written to another file, or printed. Indexes may be created for any field. The department assigned Library of Congress call numbers and subject headings to the items in the collection because these were familiar to their patrons. The data were entered by students working from entry forms designed by the staff. The department was satisfied with the choice of Pro-Cite to solve a delimited problem and did not consider integrating this system with the other systems of the department or sharing the data with other collections.

Stephen Davis offers a warning regarding the introduction of microcomputers into special collections: "While microcomputers may in the future hold out many benefits to special collections, they also seem to have the potential of returning us to the dark ages of purely local practice in terms of cataloging and automation standards. Use of the bibliographic utilities has gradually imposed a basic consistency and standardization upon catalog records—something they never had before in special collections. Given their history, it would not be surprising if some institutions leapt at the chance of doing cataloging directly on microcomputers in order to get some of the advantages of automation but still continue to catalog the way they did a hundred years ago. . . . Microcomputers should generally not be used for cataloging in place of a local or national system unless a mechanism is in place to communicate those holding subsequently to a national database."[5]

## ART COLLECTIONS

The challenges of bibliographic control of rare books pale in comparison to those encountered in the management of two- and three-dimensional art objects. The museum community has approached automation as a possible answer to its need for control, access, and information exchange.

In the United States the effort to establish a national art and architecture database has come primarily from the Research Libraries Group (RLG), with art and architecture as one of the categories in which special membership may be granted to special and independent libraries. Early members of RLG were the Metropolitan Museum of Art; the Museum of Fine Arts, Boston; and the Art Institute of Chicago. The Art and Architecture Program Committee (AAPC) of the RLG was established in 1979.

The achievements of the AAPC are described by Nancy S. Allen of the library of the Museum of Fine Arts, Boston.[6] In 1983 the AAPC began a shared cataloging project for monographic series; in 1986 an exhibition catalog task force was created to study the problem of shared acquisitions and cataloging of these elusive publications; the RLG Art Conspectus was issued in 1981 and revised in 1987, providing data to enable member libraries to rationalize their collection development in line with policies at other libraries; and in 1984, a J. Paul Getty Trust grant funded a three-year retrospective conversion project. Special databases within the RLG database RLIN include the online *Avery Index to Architectural Periodicals*, comprising some 52,000 entries in more than 500 periodicals, and the *Sales Catalog Index Input On-Line* (*SCIPIO*), with almost 90,000 records entered since 1980.

In spite of the fact that RLIN could be considered "the 'defacto' national art library database" (p. 143), an AAPC questionnaire sent to member libraries in 1987 revealed that "cataloging varies between institutions and often between visual formats within one institution, thesauri and subject headings vary, and that relatively little of this material is in machine-readable form" (p. 150). Therefore, the AAPC's Program for Research Information Management (PRIMA) recommended that the following needs be addressed in the future: "1) an increase in the retrospective conversion of bibliographies and indexes to serial literature; 2) more complete online bibliographic control over museum bulletins, exhibition catalogs, artists' books, trade catalogs, artists catalogs, and art newspapers; 3) preser-

vation efforts for the literature of art; 4) links between RLIN and bibliographic database[s] outside the United States; 5) automated access to archival records from repositories worldwide; 6) development and use of a standard MARC format for works of art; 7) automated control of architectural drawings; 8) MARC cataloging of important photograph collections worldwide; 9) access to iconolaries for art automation projects; 10) online access to images of works of art, monuments, architectural drawings, and photographs" (pp. 148–49).

Allen concludes that "the AAPC might be considered a model for cooperation among art libraries. Its success has been due largely to the strength of the Research Libraries Group itself" (p. 152).

One local project using RLIN is AVIADOR, the project of Columbia University's Avery Architectural and Fine Arts Library to provide "integrated access to the contents of the collection regardless of the format . . . [so that] a user should be able to find in one spot an answer to a question such as 'What do you have on Frank Lloyd Wright?' and know that there are 156 books by him, 136 books about him, at least 178 periodical articles on him, and approximately 600 drawings by him."[7] The project, which has been funded by the Mellon Foundation, the NEH, and Eastman Kodak, has as its goal the automated catalog records of some 45,000 architectural drawings with future access via videodisk. Those involved in the planning for the project found that the addition of genre to the MARC record (added at the instigation of archivists), the notion (from librarians) of a uniform title, and standardized subject headings from the Getty *Art and Architecture Thesaurus* had benefitted their efforts, thus emphasizing the similarity of indexing needs of archival and art materials.

Another approach to automating the cataloging of art objects is that of the Vancouver Maritime Museum in Vancouver, British Columbia.[8] The problem of standardization became paramount when confronted with the task of attempting to describe historic watercraft and other three-dimensional art objects in words. Although the curators early made the decision to follow the guidelines of the *Data Dictionary* of the Canadian Heritage Information Network (CHIN), Canada's National Museum computer organization, they had, when faced with artifacts "ranging from rubber life rafts still in their fiberglass canisters to the *Thomas F. Bayard*, a 19th century pilot schooner and sealing vessel, floating in the museum's harbour" (p. 257) and little existing standards work, to rethink museological terminology from the ground up.

In this case, the postindexing capabilities of the computer provided the solution. "The answer seems to lie in the power of the computer to hold a number of variant terms in relation, with the retrieved information depending on the search. That is, through a combination of entry guides for physical data including standards for descriptions and a sophisticated postcoordinated search system supported by a thesaurus and a network of 'see also' references, a museum can record vast quantities of data without normalizing it ahead of time (precoordinating), and still maintain the capability for effective searches" (p. 259). Still, the curators argue for "some kind of standing committee for authority work and thesaurus construction" (p. 259).

An example of an integrated system combining free-text software with MARC format records for a museum/library is the British Architectural Library's (BAL) database of books, periodicals, drawings, photographs, and realia, which will ultimately include images of the items on videodisk. This collection, the private library of the Royal Institute of British Architects, functions as the national architectural library of Great Britain as well as a museological collection. "The BAL could be described as a museum with traditional museum curatorial functions developed in parallel with the documentation functions of a traditional library service."[9] Although *AACR2* standards had been implemented for the cataloging of the library's monographs and serials, a different in-house system with different subject headings had been developed for the nonprint materials.

Automation provided the opportunity to create a single, integrated system of control and access for all the items in both collections. It was not just the automated equipment and software that made the integration possible, it was the process of standardization and systematization the automation required. "It was understood that automation alone could not solve the problems, but that cohesive systems

could be achieved only through adopting common standards" (p. 246).

The system chosen was a free-text software package, STATUS, developed by the United Kingdom Atomic Energy Authority and adapted for use on a Prime 2655 minicomputer. STATUS was attractive for the library's purposes because of its "powerful retrieval capabilities and . . . its ability to cope with a variety of record formats" (p. 246). The research organization British Non-Ferrous Metals had developed several software packages for use with STATUS: a thesaurus, a text editor, and a data entry system. STATUS at the British Architectural Library "consists of a central integrated database (the IDB) with satellite databases linked on the inner circle of satellites through the release of common data to the IDB, and on the outer circle by being searchable on a word-by-word basis achieved through the free-text application of indexing each word . . . " (p. 247). Other satellite databases include management databases for periodicals accessions and for object accessions, and a database of biographical information that serves as an authority file.

Although the system functions as a stand-alone system, because of its MARC format it will be able to link with other compatible systems internationally. However, some MARC fields had to be adapted and others borrowed (from MARC AMC) because no standards for describing art objects existed at the time the system was being developed.

Access to a similar combination of print and object collections has been developed at the Smithsonian Institution via a rigorous in-house analysis of the nature of the collections in terms of systems.[10] In 1987 the seven Smithsonian museums with major art collections, including the National Museum of American Art, the National Portrait Gallery, the Hirschhorn, and the Freer, initiated a change from the individual department systems that had evolved in a haphazard fashion to a single, integrated system that would track objects and print materials throughout their life cycles as well as provide management control and public access. The result was a move from the magnetic tape, batch processing, and two-week turnaround for hard copy of the Smithsonian's SELGEM system of the early 1970s to Infodata's INQUIRE, a text-oriented database package running on an IBM 4381 mainframe computer.

The analytic process that led to this choice was conducted in two phases: in the first phase museum functions were analyzed and models built, and in the second phase the nature of museum data was analyzed. The Smithsonian needed a system that would support "the acquisition of objects, title transfer, shipping, object tracking, conservation, maintenance of collections documentation, and much more" (p. 222). The data analysis involved identifying and defining data elements, grouping them into logical data groups, and specifying the relationships between groups. A third phase consisted of merging the results of the functional analysis and the data analysis, creating matrices to demonstrate the relationships between the data and the functions.

One of the difficulties the Smithsonian team experienced was changing mental habits from thinking in cataloging terms to thinking in terms of data and structures. "The goal of data modeling is to develop systems that are data driven rather than process driven. Processes are subject to change while data tend to be constant" (p. 228).

A promising application of automation to art collections is the use of analog optical disks, alone or in conjunction with digital text databases. While waiting for digital video interactive (DVI) technology, which integrates text and pictures in a digitized environment, to become standardized and economical, some museum curators have chosen the relatively inexpensive route of videodisks as a storage medium for images of their collections.

Pamela N. Danziger, director of Information Research Services at the Franklin Mint, summarizes the nature of the choice: "For some applications where visual images need to be manipulated or high-quality detail must be maintained, such as pages of text, or technical images, such as architectural or engineering drawings, analog optical disk storage is clearly not suitable. But for many other applications, such as photo research files, picture reference, museums and art collections, where the only storage alternatives other than paper are slides or microimage, . . . optical disk storage may be ideal, and eminently practical."[11] The advantages of the videodisk are: (1) compact storage—up to 55,000 images on a twelve-inch platter; (2) immediate retrieval,

accomplished by entering the image's location via a keyboard or keypad; (3) excellent presentation medium, allowing for both monitor display and projection; and (4) ease of use. The disadvantage is the requirement of two screens, one for the database and one for the images. Danziger recommends subjecting the collection to a needs analysis before making the choice for videodisks.

Examples of the application of videodisk technology to museum collections come from two major French museums:

The Louvre has chosen videodisks for three different applications. The first is the production of a videodisk presenting a general introduction to all the Louvre collections, which is used by visitors to the museum and is also for sale. According to Jean Galard, Chef du Service Culturel at the Louvre, "each volume contains, on the one hand a series of animated sequences, each one lasting about a minute, presenting and commenting upon several masterpieces; the other part is a bank of fixed images. The first volume is concerned with 2,500 paintings and drawings, the second 1,000 sculptures and other objects, the third 1,000 works of antiquity. Each work is presented in an ensemble view, with a caption of identification, and in at least five detailed images."[12] Each volume is accessible via an index, accompanied by a detailed catalog, and available in French, English, and Japanese. To ensure international standardization, each is produced according to PAL (European television) standards and to NTSC (American and Japanese) standards.

The second videodisk application has been designed for visitors to use in the area of the Louvre devoted to ancient Greek art. The visitor at a workstation featuring two screens may choose to access the database and the images via several different themes, such as "The Sculptural Program of the Parthenon" or "Greek Civilization." Each program lasts from ten to twenty minutes. Of the 2,300 images on this videodisk, only 1,000 came from the Louvre collections; so the videodisk provides access to an even broader range of art than that available in the Louvre.

The third videodisk was designed as an aid for the Louvre docents. Projected onto large screens, the images orient visitors to the objects in the museum's collections by placing them within the geographical context of the objects' origins: "This [videodisk] contains geographical maps, plans, diagrams, and views of the sites or the monuments of which the Louvre conserves fragments, as well as images of works conserved [in their entirety]" (p. 25).

Another major French museum to apply videodisk technology is the Union Francaise des Arts du Costume, housed since 1986 in the new Musee des Arts de la Mode. The collection comprises 10,000 garments and 32,000 accessory items dating from the eighteenth century to the present. These materials are fragile and are housed in several different locations, thereby making the videodisk an ideal access medium. The museum has produced two videodisks, the first in 1987 containing 24,000 images from their twentieth-century collection (offered for sale under the title "Mode"), and the second in 1989 containing 5,000 images from their eighteenth- and nineteenth-century collections. The videodisks are accessed in tandem with a digitized database.

According to Marie-Helene Poix, the museum's documentaliste, the results achieved with the videodisk were very satisfactory: "Along with its great storage capacity (54,000 images to a side), its very rapid access time, its excellent level of interaction when it is coupled with a microcomputer, the videodisk quickly imposed itself as a necessity, a solution to our needs."[13] The videodisks are used both by the museum staff as a management tool and by scholars.

The combination of database and optical imaging technology is one of the most interesting developments in the application of automation to collections of visual images. The combination of Boolean searching of keywords and surrogate images not only provides much quicker access to the desired images in the collection but also contributes significantly to the preservation of fragile originals by reducing the amount of handling to which they are subjected.

One such combined application has been developed by the National Archives of Canada for the 20,000 editorial cartoons and caricatures in the Canadian Centre for Caricature. The initial planning stage, as described by Gerald Stone, chief of the Information Services Section, Documentary Art and Photography Division, and Philip Sylvain, optical disc advisor,[14] involved consideration of

various means—from photography to microfilming to CD-ROM to analog optical disks—to attain their goal of access plus preservation.

Their initial solution was to combine a database management system with digital WORM optical disks for image retrieval, but they ultimately chose videodisks instead of optical disks in order to achieve a higher resolution for display and printing of the surrogate images. Their database management system is ZIM, a fourth-generation language product of Zanthe Information, Inc.

The database management system permits Boolean searching, variable-length descriptive records, and repeatable fields. Data may be entered directly or by file transfer from MINISIS, the microcomputer-based system used by the National Archives of Canada for descriptions of archival materials. Users may search by subject, artist, publication, place, date, and item number. They may then view the corresponding surrogate images with their descriptive records, zoom in on a selected image, and print newspaper-quality images via a laser printer. Plans are under way to extend this combined system, called ArchiVISTA, to the collections of photographs within the National Archives. A need for an interactive thesaurus is recognized.

An example of integrated text and image software is IMAGEQUERY, the University of California at Berkeley's prototype, which was developed in-house in response to the needs of the curators of the university's several collections of images and objects scattered throughout the campus. The UC Berkeley Image Database Project is described by Howard Besser,[15] assistant professor of library and information science at the University of Pittsburgh, who while at Berkeley was involved with the project.

The challenge was to provide access to fragile material in the University Art Museum, the Architectural Slide Library, the Geography Department's Map Library, and the Lowie Museum of Anthropology's collection of photographs of its objects. Few of the collections had benefited from any kind of automated access and most were outside the jurisdiction of the university library. However, Berkeley was committed to the goal of remote access to all campus materials via workstations. As a result, the curators sat down with the staff of the university's computing center and together thought through the problem.

The result was the prototype of IMAGEQUERY, an integrated software designed to work on a high-speed distributed computer network. The system may be accessed by a variety of microcomputers in bitmapped workstations running X-Windows on the campus network. Searching is by Boolean operators to produce a hit file of surrogate images with shortened textual descriptions. The various collections maintain their identity, requiring the user to be familiar with each collection's set of indexing terms. However, it is anticipated that terms from specialized thesauri such as the *Art and Architecture Thesaurus* and the capability to search across collections will be available in the future.

A much more economical combination of text and image databases was devised by a microcomputer user. Clarita S. Anderson of the Department of Textiles and Consumer Economics at the University of Maryland first set out to implement a grant awarded for the creation of a database of her historic textiles collection. She chose the commercially available software dBASE III+ to meet her criteria of handling up to ten thousand records, performing complex searches with several variables, being easy to learn and enter data into, and being upgradable in a nonworkstation environment.

However, "it soon became apparent that the database would be a more useful research tool if it were enhanced with images."[16] The critical factor in this user's choice of image software was cost. She warns, "Problems with the software for capturing images, the image capture board, the equipment to capture the images, the cost, and the overall quality of the images were much more difficult to solve than the initial choosing of the data management software" (p. 598). Her choice of image software was PicturePower. However, the camera required turned out to be of higher quality and therefore more expensive than the software producer had indicated, and adjustments also had to be made to the software in order to produce an image of acceptable resolution.

The user's final adjustment was to modify dBASE so that the searcher would not have to leave dBASE and enter PicturePower in order to search for images. The resulting com-

bined system, composed of commercially available software and hardware, met the needs of the user. Provision of broader access and compatibility with other like databases were not considerations.

## CONCLUSION

The application of automation to special collections has greater benefits than merely automating existing functions. One important benefit lies in the fundamental standardization required by the computer. Although standardization has been historically a bugbear to curators of rare books and art objects, the existence of commonalities in this group of unique items suggests the possibility of improved control and access to them. As Stephen Davis says, "one of the lessons of the past few years has been that special collections in different areas, such as rare books, graphic materials, manuscripts, maps, music, archival motion pictures, even machine-readable data files, have a great deal in common in terms of specialized access requirements."[17]

Not only improved access can be achieved by automation but also an improved restructuring of all of the functions of the special collections library. Typical of almost every successful automation project has been functions analysis as a first step. Annette F. Waterman of the School of Information Studies at Syracuse University points out that "a manual system simply transferred to a computer does not change the structure of a system."[18] "The first step should be the formulation of a statement of purpose for the collection . . . . A systematic analysis of the collection and its internal workings should be the second step" (p. 61).

Every aspect of the special collections library should be included in a functions analysis, according to Richard M. Kesner, archivist, consultant, and manager of Office Systems and Services for the F. W. Faxon Company: "In developing a plan of action, word processing, financial modeling, and facility management must share the stage with indexing and information retrieval."[19] He identifies seven areas of activity to be considered: "collection development, physical control over collections, intellectual control over collections, reference services, general administration, grants administration, and publications production" (p. 24).

The final benefit of automation for special collections derives from the fact that automation in itself creates a special collection, a new way of looking at the disparate units in the collection by bringing them together in new and unexpected ways. A special collection is an organism, whose totality is greater than the sum of its parts and whose contribution to research frequently arises from its juxtaposition of unique pieces of reality, not just from amassing them in one place. Automation is a very friendly means to this end.

## REFERENCES


1. Stephen Paul Davis, "Bibliographic Control of Special Collections: Issues and Trends," *Library Trends* 36 (Summer 1987): 112–13.
2. John B. Thomas III, "The Necessity of Standards in an Automated Environment," *Library Trends* 36 (Summer 1987): 130.
3. "ISTC," *Papers of the Bibliographical Society of America* 83 (Sept. 1989): 371.
4. Julia M. Rholes and Suzanne D. Gyeszly, "ProCite as a Management Tool for a Travel and Tourism Collection," *Collection Building* 10, nos. 3/4 (1989): 48.
5. Davis, 119.
6. Nancy S. Allen, "The Art and Architecture Program of the Research Libraries Group," *INSPEL* 23 (1989): 141–54.
7. Angela Giral, "At the Confluence of Three Traditions: Architectural Drawings at the Avery Library," *Library Trends* 37 (Fall 1988): 233.
8. John E. Summers and Edward G. Summers, "The Computerized Cataloguing of Historic Watercraft: A Case Study in Information Retrieval in Museology," *Journal of the American Society for Information Science* 40 (May 1989): 253–61.
9. Jan Floris van der Wateren, "Achieving the Link Between Art Object and Documentation: Experiences in the British Architectural Library," *Library Trends* 37 (Fall 1988): 244.
10. Patricia Ann Reed and Jane Sledge, "Thinking About Museum Information," *Library Trends* 37 (Fall 1988): 220–31.
11. Pamela N. Danziger, "Picture Databases: A Practical Approach to Picture Retrieval," *Database* 13 (Aug. 1990): 13.
12. Jean Galard, "Le Videodisque au Louvre: nouveaux moyens, nouvelles pratiques," *Art Libraries Journal* 15, no. 2 (1990): 24; my translation.

13. Marie-Helene Poix, "L'Union Francaise des Arts du Costume, le Centre de Documentation, le videodisque," *Art Libraries Journal* 14, no. 4 (1989): 23; my translation.
14. Gerald Stone and Philip Sylvain, "Archi-VISTA: A New Horizon in Providing Access to Visual Records of the National Archives of Canada," *Library Trends* 38 (Spring 1990): 737–50.
15. Howard Besser, "Visual Access to Visual Images: The UC Berkeley Image Database Project," *Library Trends* 38 (Spring 1990): 787–98.
16. Clarita S. Anderson, "A User's Applications of Imaging Techniques: The University of Maryland Historic Textile Database," *Journal of the American Society for Information Science* 42, no. 8 (1991): 598.
17. Davis, "Bibliographic Control," 117.
18. Annette F. Waterman, "First Steps in Planning the Automation of a Slide Collection," *Art Documentation* 8 (Summer 1989): 62.
19. Richard M. Kesner, *Automation for Archivists and Records Managers* (Chicago: American Library Assn., 1984), 10. ■■

# Communications

## A Comparison of OCLC and WLN Hit Rates for Monographs and an Analysis of the Types of Records Retrieved

**Rosemary E. Ross**

This study was initiated to secure data on the types of records found in OCLC and WLN for new monographs being cataloged at the Central Washington University Library. It examines what the hit rate is in each system for newly acquired monographs. The records are analyzed as to whether they have a Library of Congress (LC) classification number or Dewey Decimal number and whether the LC classification number is complete. Statistics are presented comparing the search results in OCLC and WLN regarding the number of duplicate records found and whether multiple records for various titles are retrieved when a given title it is searched in either database.[1,2] Retrieved records are analyzed as to whether they are serial or monographic records.

Central Washington University's library is a medium-sized academic library of 469,480 volumes supporting the curricular and faculty research needs of this four-year, comprehensive state university. Bachelor's degrees are offered in the arts, business, education, flight technology, home economics, industrial and engineering technology, law and justice, music, various sciences, and social sciences. Master's degrees are offered in the arts, sciences, education, home economics, organizational development, psychology, and resource management sciences. There are 6,435 students and 366 faculty members.[3]

The library uses OCLC for cataloging, but because it is located in the Pacific Northwest, where WLN is used by many libraries, it was possible to study both OCLC and WLN to compare the respective hit rates and the types of records retrieved. The different sizes of their respective databases (OCLC has 24.8 million records, WLN 7.8 million) suggests to some that OCLC will have a much higher hit rate than WLN.[4] This study was designed to test that assumption as well as to compare the types of records found in the two databases.

The results show that, for this library, WLN is as effective as OCLC in supplying full-level records and CIP level-eight records. Of the 427 titles in the sample, full-level records were retrieved for 328 (76.8%) in WLN. In OCLC 284 (66.5%) of the records retrieved were full-level. One (0.2%) minimal-level record was found in OCLC. In WLN fifteen (3.5%) minimal-level records were retrieved. One (0.2%) title in the sample was not found in OCLC, whereas seven (1.6%) titles were not found in WLN.

### SUMMARY OF THE LITERATURE

A review of the literature produced references to several articles reporting studies of the hit rates in OCLC and in WLN for various types of library materials. In 1974 Meyer and Panetta compared the hit rate for English-language approval books in OCLC (92%) to the hit rate for titles found in the Blackwell North America database (72%).[5] Tracy and Remmerde searched a sample of titles in BALLOTS (RLIN), Blackwell North America, OCLC, and WLN and found hit rates varying from 40% to 96% in these four databases.[6]

The November-December 1979 issue of *Library Technology Reports* compared OCLC, RLIN, WLN, and UTLAS and included hit rates for items when first searched in these utilities from several libraries. OCLC had a hit rate of 83%, RLIN 69%, WLN 70%, and UTLAS 68%.[7] In the 1980 study by Metz and Espley, new monographs were searched in OCLC within sixteen weeks of receipt, with a hit rate of 87.1% for all types of records.[8] A comparison of RLIN and OCLC hit rates for monographs and analytics done at the Univer-

Rosemary E. Ross is Catalog Librarian, Central Washington University, Ellensburg, Washington.

sity of California, Davis, produced an overall hit rate in OCLC of 68% and 50% in RLIN.[9]

In the study done by the University of Oregon in 1981, estimates of hit rates for all types of materials were over 90% for OCLC, 70% to 90% for RLIN, and 50% to 70% for WLN.[10] Three years later the 90% hit rate for OCLC was confirmed by Getz and Phelps.[11] Struble and Kohberger's study of monographs investigated the type of record present in OCLC at the time of receipt and six months later. They found LC copy available for 62% of the monographs at the time of receipt. Six months later, when items not found in the first search were searched again, more than 50% of these books had some type of copy, either LC or member input.[12]

In a comparison of OCLC, WLN, and RLIN done by the University of Idaho Library in 1988, the hit rate for copy cataloging was 96% in OCLC and 90% in WLN.[13] Allan researched the hit rate in OCLC, RLIN, Bibliofile, and Dialog for one thousand titles during 1987. Copy was found for 86.3% of the titles searched in OCLC. Allan defines "help" copy as records that varied from the book being cataloged in the country of publication, publisher, or edition or lacked a LC classification number or subject headings.[14] When "help" copy that did not exactly match the book was included in the statistics, the hit rate rose to 97.2%.

## SAMPLE DESIGN

During the past five years an average of 12,800 new monographic volumes were cataloged and added to the library's collections each year. These include books on the humanities, sciences, social sciences, business, education, music, and technology written in English and several foreign languages (primarily French, German, and Spanish). Since most new monographic titles are acquired in the winter and spring months, the sample of titles for this study was selected from November 1989 through April 1990.

Books are purchased from two jobbers, Blackwell North America and Midwest Library Service. To ensure that any potential variation caused by selecting books from either jobber did not bias the results of this study, the sample was divided to represent the average proportion of books purchased from each of the two jobbers during the previous two years. Blackwell (approvals) supplied 67% of the new titles, while Midwest (requests) supplied 33% of them.

There was an impression that books from Blackwell had fewer full-level LC records found in OCLC than did the books from Midwest due to the fact that the majority of books in the backlog had Blackwell order slips in them. No actual data existed to substantiate this observation, however.

Blackwell automatically sends new books on approval, and titles also are selected from its approval forms. All titles requested by the faculty and students are sent to Midwest. The assumptions for this study were that the titles purchased on approval were published more recently and that titles purchased from Midwest had an earlier publication date because the user had seen a review of the book, seen the actual book, or heard about the book from a colleague. In the last case, books ordered from Midwest would be somewhat older, at least by a few months, and this might make a difference in the type of catalog record available from a bibliographic utility.

With a 95% confidence level and a 2% sampling error, the size of the sample was computed to be 427 titles.[15] The sample was divided to represent the two jobbers, Blackwell with 67% (287 titles) and Midwest with 33% (140 titles).

## METHODOLOGY AND WORKFLOW

New books came to the Cataloging Department on separate book trucks according to the originating jobber. To allow the sample from each shipment to be proportional to the size of that shipment, the average number of books per shelf on new book trucks was calculated. From this it was concluded that four books should be selected from each full shelf of books on the new book trucks to select the sample of 427 monographs.

As each truck of new books came to the Cataloging Department, four numbers ranging from one through twenty-five were randomly selected. Those numbers were used to select the books on a given shelf of a book truck. The selected books were removed from the truck. This procedure was done for each shelf of books on each new truck of books until the total number of books needed for the sample was reached.

The selected books were removed from the truck and the title pages and versos photocopied. These photocopies were used to

search in OCLC and WLN for bibliographic records. All titles were searched within two weeks of receipt in the Cataloging Department. The first search for each title was alternated between OCLC and WLN in order to eliminate any advantage of searching earlier in either database, since new records are added daily to both databases. Fourteen days was the maximum number of days allowed between searching in one database and completing the search in the other database.

Each title was searched by ISBN. If no record was retrieved, then an author/title or title search was done. A printout was made of the search results including duplicate and multiple records retrieved and the actual bibliographic record that matched (hit) the title. The data on these printouts were used to compile the statistics for the study.

A retrieved record was considered to be a hit if the author, title, country of publication, publisher, and date of publication matched. For level-eight (CIP) records, if the above data matched but the date of publication varied by one or two years, the record was counted as a hit because there are instances in which the CIP record is created by LC but publication of the book is delayed substantially. This same definition was applied to the term *bibliographic item* as used in this study.

## RESULTS

### Hit Rates

Table 1 shows the hit rate for the 427 new monographic titles searched in OCLC and WLN. Bibliographic records were found for 426 titles in OCLC and 420 titles in WLN.

### Types of Records Found

In both OCLC and WLN the same number of full-level LC records was found, 251, or 58.8% of the sample (see table 2). However, in WLN an additional 56 (13.1%) upgraded CIP records were found for a total of 307 (71.9%) full-level LC and upgraded CIP records. The 125 level-eight (CIP) records found in OCLC made up 29.3% of the sample. In WLN 72 (16.9%) CIP records were found.

*Table 1. Hits for 427 Titles*

| | Titles Found | |
|---|---|---|
| | No. | % |
| OCLC | 426 | 99.8 |
| WLN | 420 | 98.4 |

The difference in hit rates between OCLC and WLN was six records, or 1.4%.

Twelve of the records found in OCLC were United Kingdom MARC (UKM) records, 2.8% of the sample. Only two UKM records were found in WLN for titles in the sample. This is due to the different sources OCLC and WLN use for MARC records. OCLC loads UKM tapes, while WLN does not.[16] The two UKM records retrieved in WLN came into the WLN database on tapes from other utilities that happened to include these two upgraded UKM records.

Records input by member libraries accounted for 33 (7.7%) of the records found in OCLC. Twenty-one (5%) of the records found in WLN were input by participants.[17]

In both OCLC and WLN one Library of Congress minimal-level cataloging (MLC) record was found. The LC minimal-level record in OCLC had been upgraded to a full-level record in WLN. The LC minimal-level record in WLN had been replaced by a UKM record.

### Analysis of Types of Record Found

The different number of CIP records retrieved in OCLC (125) compared to those retrieved in WLN (72) reflects the diverse policies of WLN and OCLC regarding the upgrading of CIP records. OCLC allows member libraries to add the physical description (field 300) to CIP records. A call number and subject headings may be added if the record lacks them, but no other data can be added or changed.[18] If an error is found that a library is not authorized to change, a report form is filled out and sent to the OCLC Quality Control Section for evaluation and correction of the record.

WLN requires that originally cataloged records of new participant libraries and newly hired catalogers in WLN libraries undergo WLN quality control review until the records from that library or cataloger, as the case may be, meet national cataloging standards for description, subject analysis, and MARC tagging.[19] This quality review process takes place for each format that a library uses. After completing this process, libraries may update CIP records so that they become complete, full-level records. CIP records are upgraded by participant libraries only once in WLN due to this process. The upgraded CIP records are replaced when a full-level Library of Congress

*Table 2. Type of Record by Total Number and Percentage of the Sample*

| | OCLC | | WLN | |
|---|---|---|---|---|
| | No. | % | No. | % |
| Full-level LC | 251 | 58.8 | 251 | 58.8 |
| Upgraded level 8 | 0 | 0 | 56 | 13.1 |
| CIP - level 8 | 125 | 29.3 | 72 | 16.9 |
| UKM | 12 | 2.8 | 2 | 0.5 |
| NLM | 1 | 0.2 | 0 | 0 |
| NLC | 3 | 0.7 | 3 | 0.7 |
| Full-level member cataloging | 33 | 7.7 | 21 | 5.0 |
| MLC | 1 | 0.2 | 15 | 3.5 |
| Total records | 426 | | 420 | |

In WLN an additional 14, or 3.3% of the records found, were participant-input MLC records.

record is distributed on the MARC tapes. WLN participants can bring any record that is not full-level into the change file and make corrections or additions to it according to established guidelines. Records in the change file are reviewed by WLN quality control staff. Alternately, a printout of a record with its required changes noted may be sent to WLN's Bibliographic Center for corrections, with no filled-out forms required.

The different policies of OCLC and WLN regarding the upgrading of CIP records account for 44 (10.3%) more full-level records being retrieved in WLN than in OCLC. The higher number of minimal-level records found in WLN reflects another policy difference between the two systems. WLN requires that a minimal-level record be input by a library before a title can be cataloged. This allows participants to know that another library has the title and is going to catalog it and helps to minimize the number of duplicate records input. Such minimal-level records are also used by participants when ordering items through WLN's acquisitions subsystem.

## Place and Dates of Publication

Three-quarters of the titles found were published in the United States. Another 20% of the titles were published in Great Britain. Titles published in Canada and other foreign countries made up 4.5% of the titles found in both databases. Table 3 illustrates the distribution of place of publication in the records found.

In this sample the country of publication did not appear to influence whether a record was found for titles published in the United States, Canada, and Great Britain. The titles published in these three countries but not found in WLN were published by a small press or were reprint, paperback, or photocopy editions. One title published in Warsaw, Poland, was not found in either OCLC or WLN. (The place of publication for the records not found can be seen in table 9.)

Seventy-three percent of the titles were published during the years from which the sample was taken, 1989–1990. An additional 20% were published in 1988. Less than 1% of

*Table 3. Country of Publication in Hits*

| | OCLC | | WLN | |
|---|---|---|---|---|
| | No. | % | No. | % |
| United States | 324 | 75.9 | 317 | 74.2 |
| Canada | 6 | 1.4 | 6 | 1.4 |
| United Kingdom | 83 | 19.4 | 84 | 19.7 |
| Other foreign | 13 | 3.1 | 13 | 3.1 |
| Total records | 426 | | 420 | |

*Table 4. Date of Publication in Hits*

| | OCLC | | WLN | |
|---|---|---|---|---|
| | No. | % | No. | % |
| 1990 | 38 | 8.9 | 39 | 9.1 |
| 1989 | 276 | 64.6 | 271 | 63.5 |
| 1988 | 86 | 20.1 | 88 | 20.6 |
| 1987 | 9 | 2.1 | 7 | 1.6 |
| 1986–1980 | 13 | 3.1 | 12 | 2.8 |
| pre-1980 | 3 | 0.7 | 3 | 0.7 |
| 19uu– | 1 | 0.2 | 0 | 0 |
| Total records | 426 | | 420 | |

*Table 5. Presence of LC Number in Hits*

| | OCLC | | WLN | |
|---|---|---|---|---|
| | No. | % | No. | % |
| Complete LC number | 405 | 94.9 | 398 | 93.2 |
| No or incomplete LC number | 21 | 4.9 | 22 | 5.2 |
| Total records | 426 | | 420 | |

*Table 6. Presence of DDC Number in Hits*

| | OCLC | | WLN | |
|---|---|---|---|---|
| | No. | % | No. | % |
| Dewey number | 403 | 94.6 | 396 | 94.3 |
| Total records | 426 | | 420 | |

the titles were published prior to 1980. Table 4 shows the distribution of publication dates in the records found.

### Classification Number Availability

Complete LC classification numbers were present in 405 (94.9%) OCLC records and 398 (93.2%) WLN records. Holdings in WLN include the library's NUC symbol and its classification number for that title. Thus some records in WLN may lack an 050 or 090 field, but LC classification numbers can be found in the holdings display for those titles. The same situation applies for Dewey classification numbers. Twenty-one (4.9%) OCLC records and 22 (5.2%) WLN records lacked or had incomplete LC classification numbers as noted in table 5.

The percentage of retrieved records with a Dewey classification number was 94.6% in OCLC and 94.3% in WLN. Of the 426 hits in OCLC, 403 contained a Dewey classification number. Of the 420 hits in WLN, 396 records had a Dewey number (see table 6). Twenty-three hits had no Dewey number in OCLC, compared with 24 hits in WLN.

### Duplicate and/or Multiple Records

Of the 426 titles found in OCLC, 118 (27.6%) had duplicate records.[20] In WLN 9 (2.1%) of the 420 titles found had duplicate records. Multiple records were retrieved for 21

*Table 7. Titles Found with Duplicate and/or Multiple Records Retrieved*

| | OCLC | | WLN | |
|---|---|---|---|---|
| | No. | % | No. | % |
| Titles with duplicate records | 118 | 27.6 | 9 | 2.1 |
| Titles with multiple records | 21 | 4.9 | 24 | 5.7 |

(27.6%) OCLC records, and 24 (5.7%) titles retrieved in WLN had multiple records (see table 7).

Table 8 analyzes the types of records that were duplicates. In OCLC 118 (27.6%) and in WLN 9 (2.1%) titles searched had one or more duplicate records. In OCLC three of these duplicates were caused by the presence of an extra full-level member record. In WLN one duplicate was caused by an extra full-level participant record. Duplicates caused by an order record in addition to a full-level record occurred for three titles in OCLC and for two titles in WLN. The majority of all duplicates were caused by records from national-level cataloging sources and not by records input by member libraries.

In OCLC three of the duplicate records also retrieved multiple records. None of the WLN duplicate records also retrieved multiple records. Cases in which more than one record was retrieved, and these records were various editions of the title being searched or for other titles, were counted as multiple hits. Multiples hits found for three reasons: the presence of more than one edition of a title in OCLC or WLN; no match when searching the title by an ISBN, thus necessitating an author or title search; or the ISBN search retrieved more than one record.

### Titles Not Found

One title was not found in either OCLC or WLN. Records for earlier editions of it were in both OCLC and WLN. All no-hit titles were English-language books. Six of the seven titles not found in WLN were new editions, reprints, or paperback editions and had records in WLN for another edition. One title not found in WLN was a University Microfilms photocopy of a doctoral dissertation written in 1984. These titles not found require original cataloging: 1 (.02%) title not found in OCLC and 7 (1.6%) titles not found in WLN. Details of the no-hit titles can be seen in table 9.

### Minimal-Level Interim Records

Both OCLC and WLN had one LC "in process," encoding level 5, record. However, they were for different titles. The additional fourteen WLN MLC records were published in 1988 (3 titles), 1989 (11 titles) and 1990 (1 title). Four of these titles were published in the United States, 7 in London, and 1 in Tokyo.

### Serial Records Found

All titles were assumed to be monographs when ordered. Three titles were volumes of annual publications. Thus serial records were retrieved for them in both OCLC and WLN. A fourth title was cataloged as a serial in OCLC. In WLN it was analyzed and cataloged as a monograph under the distinctive title for that volume.

*Table 8. Analysis of Types of Duplicate Records Retrieved*

| | OCLC | WLN |
|---|---|---|
| Full-level LC & UKM | 101 | 0 |
| Full-level LC & NLM | 8 | 6 |
| Full-level LC & full-level member records | 0 | 1 |
| Full-level LC, NLM, & UKM | 2 | 0 |
| Full-level LC, NLM, UKM, and order records | 1 | 0 |
| Full-level LC and order record | 2 | 0 |
| Full-level member record & UKM | 1 | 0 |
| Two full-level member records | 1 | 0 |
| Full-level member record and order record | 0 | 2 |
| UKM and NLM records | 1 | 0 |
| Two UKM records | 1 | 0 |
| Total duplicates | 118 | 9 |

*Table 9. Titles with No Hits*

| Type of Book | Place of Publication | Date | No. |
|---|---|---|---|
| *Not Found in OCLC and WLN* | | | |
| English-Polish dictionary | Warsaw, Poland | 1988 | 1 |
| *Not Found in WLN* | | | |
| Reprint edition | New York, N.Y. | 1989 | 1 |
| Reprint edition | Bowie, Md. | 1989 | 1 |
| New edition | Seattle, Wash. | 1988 | 1 |
| Paperback edition | New York, N.Y. | 1988 | 1 |
| Small press | Lynchburg, Va. | 1989 | 1 |
| UMI photocopy of a dissertation | Ann Arbor, Mich. | 1986 | 1 |
| Total titles not found | | | 7 |

## CONCLUSIONS

A high percentage of bibliographic records was found in both OCLC and WLN for the new monographic titles purchased by the library—99.8% and 98.4% respectively. A reexamination of the titles that lacked matching records in OCLC or WLN would be of value to determine whether records are now in the databases for them and the type of record they are, and to note the date when these records were added.

There were 332 (77.8%) full-level cataloging records found in WLN, compared to 287 (67.2%) records found in OCLC (including encoding level "blank" from LC, NLM, NLC, member input, and WLN upgraded level-eight [CIP]).

In WLN 73 (17.1%) records required upgrading from encoding level eight (CIP), compared to 126 (29.5%) in OCLC. These included both LC and NLC level-eight records. This result suggests that a change in OCLC's policy regarding the updating of monographic CIP records for books could save hundreds of libraries from making the same modifications to a given CIP record when they use that record for cataloging a new title. Perhaps it would be feasible for all member libraries and Enhance libraries to upgrade CIP monographs, leaving the record as a level eight for later replacement when the full-level record is distributed by LC on its MARC tapes. If OCLC allowed all member libraries to upgrade CIP records for books, a substantial savings of time in cataloging monographic books with LC CIP records would result.

Possibly a pilot project for updating these CIP records could be conducted by OCLC with various sets of rules used by groups of testing libraries. Since there are various sets of rules for updating a CIP record, each set could be treated separately to determine the set most acceptable to OCLC and the library community.

Thirteen records (3.1%) found in OCLC required substantial upgrading (UKM and MLC), compared to fifteen (3.5%) in WLN. Original cataloging was needed for one (0.2%) title not found in OCLC and for seven (1.6%) records not found in WLN. The total of all records needing upgrading or original cataloging in WLN was 95 (22.2%), while the records in the same category in OCLC numbered 140 (32.8%).

Complete LC classification numbers were present in 405 (94.9%) OCLC records and in 398 (93.2%) WLN records. Dewey Decimal classification numbers were found in 403 (94.6%) OCLC records and in 396 (94.3%) of the WLN records.

Further study is needed to reevaluate the duplicate records found in OCLC and WLN. In June 1991 OCLC implemented its Duplicate Detection and Resolution software to detect and delete the duplicate records. Such a study could analyze how many and which of the duplicate records retrieved during this study are no longer in the OCLC database. The WLN duplicate records caused by the presence of an order record could be examined for elements in the records that might be added to the WLN replace algorithm to enhance the replacement process.

The hit rate for new monographic titles

differed only by 1.4% between OCLC and WLN even thought their databases vary substantially in size, OCLC with 24.8 million bibliographic records and WLN with 7.8 million records.

These results demonstrate the high rate of success medium- or smaller-sized academic libraries similar to Central Washington University Library may have in searching for catalog records for new monographs in OCLC (99.8%) and WLN (98.4%). Future studies that compared the hit rate and types of records found in OCLC and WLN for materials in other formats such as serials, music scores, sound recordings, computer files, microforms, and audiovisual materials would be beneficial. If a sample of titles in each format were studied, similar libraries would have significant data to use in evaluating whether OCLC or WLN would best meet their cataloging needs based on their particular mix of types of materials.

## REFERENCES AND NOTES


1. Duplicate records were defined for this study as more than one bibliographic record in the same format in a database for the same bibliographic item.
2. Multiple hits were defined as search results of more than one bibliographic record none of which were duplicate records for the bibliographic item being searched.
3. The student count is the average FTEs per quarter for 1990–1991; the faculty count is the head count of tenure and tenure-track faculty for the same academic year.
4. In November 1992 OCLC had 24.8 million bibliographic records; WLN had 7.8 million records.
5. R. W. Meyer and Rebecca Panetta, "Two Shared Cataloging Data Bases: A Comparison," *College & Research Libraries* 38 (Jan. 1977): 22–23.
6. Hit rates were BALLOTS (RLIN) 39.76%, Blackwell North America 83.66%, OCLC 96.16%, and WLN 56.59%. Joan I. Tracy and Barbara Remmerde, "Availability of Machine-Readable Cataloging: Hit Rates for BALLOTS, BNA, OCLC and WLN for the Eastern Washington University Library," *Library Research* 1 (Fall 1979): 277–81.
7. Joseph R. Matthews, "The Four Online Bibliographic Utilities: A Comparison," *Library Technology Reports* 15 (Nov.-Dec. 1979): 737.
8. Paul Metz and John Espley, "The Availability of Cataloging Copy in the OCLC Data Base," *College & Research Libraries* 41 (Sept. 1980): 432.
9. Kazuko M. Dailey, Grazia Jaroff, and Diana Gray, "RLIN and OCLC—Side by Side: Two Comparison Studies," *Advances in Library Administration and Organization* 1 (1982): 105.
10. University of Oregon Library, "A Comparison of OCLC, RLG/RLIN, and WLN," *Journal of Library Automation* 14 (Sept. 1981): 224–26.
11. Malcolm Getz and Doug Phelps, "Labor Costs in the Technical Operation of Three Research Libraries," *Journal of Academic Librarianship* 10 (Sept. 1984): 218.
12. Carol A. Struble and Paul B. Kohberger, "Statistical Survey to Determine Availability of Cataloging Copy on OCLC," *Cataloging & Classification Quarterly* 7 (Spring 1987): 16.
13. Mary K. Bolin, ed. and comp., *Comparison of WLN, RLIN, and OCLC* (Moscow, Idaho: University of Idaho Library, Utility Review Committee, 1988), 15 (ERIC ED 296 733).
14. Ann Allan, "Chasing MARC: Searching in Bibliofile, Dialog, OCLC and RLIN," *Journal of Academic Librarianship* 15 (Jan. 1990): 341.
15. Donald S. Tull and Gerald S. Albaum, *Survey Research: A Decisional Approach* (New York: Intext Educational Publishers, 1973), 48.
16. Since this sample was taken the British Library agreed that OCLC could merge UKM records with Library of Congress records, but not with records input by members. OCLC Pacific Network, *News Update*, 63 (Apr. 1991): 4.
17. WLN designates libraries using their system as participants; OCLC uses the term member.
18. OCLC Online Computer Library Center, *Cataloging User Guide* (Dublin, Ohio: OCLC Online Computer Center, 1990), 68-69.
19. Descriptions of the review process from two different viewpoints are found in Bruce N. Ziegman, "WLN's Database: New Directions," in *National and International Bibliographic Databases: Trends and Prospects* (New York: Haworth, 1988), 105–7; and in Bolin, *Comparison of WLN, RLIN, and OCLC*, 8–9.
20. Hickey and Rypka estimated that 7 to 9 percent of the monographic records in OCLC were duplicates. Thomas B. Hickey and David J. Rypka, "Automatic Detection of Duplicate Monographic Records," *Journal of Library Automation* 12 (June 1979): 141. Wanninger concluded that 19 percent of the records found in OCLC were duplicates. Patricia Dwyer Wanninger, "Is the OCLC Database Too Large? A Study of the Effect of Duplicate Records in the OCLC System," *Library Resources & Technical Services* 26 (Oct.-Dec. 1982): 359. ■■

# News and Announcements

## Charles W. Bailey, Jr., Receives LITA/Library Hi Tech Award

Charles W. Bailey, Jr., has been named winner of the first LITA/Library Hi Tech Award for outstanding achievement in communicating to educate practitioners within the library field in library and information technology. He receives the award for his work in establishing and maintaining PACS-L and associated electronic publications. The award ($1,000 donated by Pierian Press and a plaque) was presented during the LITA President's Program at the 1993 ALA Annual Conference New Orleans.

To quote from the award citation:

PACS-L began in the late 1980s as a timely means of communication for a few network-connected library systems professionals sharing project information and posing questions to one another. Through continued improvement in the posting process, development of logon and contribution instructions, moderator services, program presentations, and constant availability to answer questions, Charles expanded the audience and participation in PACS-L to include librarians around the world. . . . Today we learn from a broad range of librarian-created network information services, most of which trace their beginning to the educational process inherent in PACS-L.

## VTLS Offers ADA-Support Workstation

To help libraries comply with Title III of the Americans with Disabilities Act (ADA), VTLS Inc. has developed an option that provides visually impaired users equal access to the VTLS integrated library system. Configured with special VGA and voice synthesizer boards, the ADA-support workstation permits screen enlargement as well as voice output.

The ADA-support interface works with any data communications package but looks best when used with the VTLS Intelligent Workstation. The interface comes with a mouse, which can be used to choose one of three modes of screen enlargement: full screen, in place, single line (ruler view), and single line at the bottom of the screen (dual view). Further enlargement or reduction is possible with another click of the mouse. One click of all three mouse buttons together reverses the polarity of white letters against a black background to black letters against a white background. For users who are totally blind, the interface features a voice board that has been customized to read aloud specific parts of the screen. For additional information, contact Gail Gulbenkian, (703) 231-3605.

## Steve Cisler Receives LITA/Gaylord Award

Steve Cisler, information scientist for the Apple Computer, Inc., Library, is the 1993 recipient of the Library and Information Technology Association/Gaylord Award for Achievement in Library and Information Technology.

The award, $1,000 and a citation donated by Gaylord Bros., Inc., recognizes achievement in library and information technology in the areas of distinguished leadership, notable development of applications of technology, superior accomplishment in research or education, or original contribution to literature in the field.

"Steve Cisler is being recognized for his far-reaching work in the development and application of new technologies in libraries and for providing leadership in the field of library technology and computer networking," said Jeanne Somers, chair of the LITA/Gaylord Award Committee.

Somers said Cisler's specific accomplishments and ongoing endeavors include his work with advanced technology retrieval projects at Apple Computer, coordination of the 20,000-member Apple Library Users Group, and management of the Apple Library of Tomorrow program.

"Server, builder, and cohost of the com-

puter conference on the WELL, Steve writes extensively on a wide range of topics including the Internet, electronic free speech, the expanding mission of libraries, and the future of librarianship," Somers said.

Cisler has worked as branch librarian for the Contra Costa County Library in Pinole, Calif., and as library assistant at the Vallejo Public Library (Calif.), and set up the first library for the school in Togo while a member of the U.S. Peace Corps in East Africa. He has a bachelor's degree in history and literature of religion from Northwestern University and a master's degree in library science from the University of California, Berkeley.

### Rob Carlson Named LITA Deputy Executive Director

Rob Carlson, assistant to the president of the Urban Libraries Council (ULC), has been appointed as the first deputy executive director of LITA. Carlson will be responsible for LITA's continuing education program, including regional institutes, preconferences, and national conferences.

"It is a great pleasure to look forward to LITA activities in which Rob's considerable talents and experience will be a major influence and assistance to our members," said LITA Executive Director Linda J. Knutson. "We welcome him back to LITA and ALA."

Prior to his position with ULC, Carlson served LITA as interim program manager, managing the 1992 national conference in Denver. He previously served as ALA's manager of information technology publishing, ALANET system manager, and deputy ALANET system manager.

Carlson also served as micro program liaison officer for the AMIGOS Bibliographic Council and began his career as director of the learning resource center at Laredo State University. He has a bachelor's degree in French and a master's degree in library science from Western Michigan University. Carlson began in his new position on June 24 at the ALA Annual Conference in New Orleans.

# Recent Publications

## Book Reviews

**Lane, Elizabeth S., and Craig Summerhill.** *Internet Primer for Information Professionals: A Basic Guide to Internet Networking Technology.* Westport, Conn.: Meckler, 1993. 179p. paper, $37.50 (ISBN 0-88736-831-X).

*Internet Primer* provides a basic overview of the Internet; what it is, why it is, and how it works. The book is divided into six chapters. The first three chapters give a good introduction to the terminology and technology of the physical Internet, including an explanation of what the Internet is, an overview of the smaller networks that make up the Internet, and an introduction to networking technology and protocols. The value of these chapters lies in the author's clear explanation of her subject.

Much of the material about network applications contained in chapter 4 is already available in greater detail in Brendan Kehoe's book, *Zen and the Art of the Internet;* however, the text presented here covers the basics and is appropriate for Internet novices.

Chapter 5, on network resources, contains a valuable list of information resources for Internet users, but for the sake of the Internet novices, for whom the book was written, the chapter might have been better organized by dividing it into a section on software and hypertext programs available via FTP (file transfer protocol), such as HYTELNET or the various hypercard stacks; a section on printed resources also available via FTP, such as *Zen and the Art of the Internet* or the *NYSERNET New User's Guide;* and a section on online resources, such as CARL. (For some reason, although the CARL dial-up number was given, the Internet address was left out. It is PAC.CARL.ORG). More thorough explanations of both Gopher and WAIS, two software programs available via FTP that have been implemented at a number of sites, would also have been helpful.

Chapter 6 discusses a number of important policy issues associated with the Internet such as who will pay, who will have access, and who will control the data that is made available, all issues of interest to librarians.

Overall, *Internet Primer* is a useful tool for those interested in learning about the Internet. Unfortunately, the large number of typographical errors (at least seven in one section of fifty pages of text alone), and a couple of unnecessary split infinitives within the same number of pages detract from the text. Among the innovations of the Meckler editors is a new way to spell 'entrepreneur.' The new spelling appears twice within the space of a few lines; in one instance it is printed in a paragraph heading in large, boldfaced type.

In addition to the editing and proofreading problems, the covers of the paperback edition began to curl backward as soon as the book was taken from an air-conditioned room, and within twenty-four hours, the covers had rolled completely back upon themselves. The conditions to which the book was exposed were ordinary summer temperature and humidity conditions in a house with no air conditioning. Even after returning the book to an air-conditioned environment and placing it beneath a pile of other books, the covers have remained curled. This book would not survive long under normal library use outside the most ideal of conditions. Meckler has a responsibility both to the authors of this book and to potential purchasers to provide better editing, proofreading, and book quality than this.

This book is a useful overview for novices, managers, students, and others with a general interest in learning about the Internet.—*Kathleen Rehn, The New York Public Library.* ■■

***Networks for Networkers II: Critical Issues for Libraries in the National Network Environment.*** Ed. Barbara Evans Markuson and Elaine W. Woods. New York: Neal-Schuman, 1993. 249p. $45 (ISBN 1-55570-128-0).

*Networks for Networkers II* is a collection

of papers presented at a networking preconference to the second White House Conference on Libraries and Information Services held December 17–19, 1991. The LC Network Advisory Committee (NAC) sponsored the preconference, whose name the present volume bears. The ten papers cover topics ranging from the basic technology underlying the Internet to the ways the networks have changed scholars' use of information. Such broad coverage might make this volume suitable as an introduction to networking. The volume targets librarians, with most papers discussing possible roles for libraries and librarians in the networked world.

Though perhaps fresher at the time of the preconference, many of the insights and much of the information contained in the papers seem like old news. Douglas Van Houweling's presentation of basic networking technology offers nothing new or different from similar descriptions to be found in any of the numerous introductions to the Internet in print today. Clifford Lynch's "Strategic Agenda for Libraries and Networked Information Resources in the 1990s" repeats what he and others have written and said elsewhere. Peter Lyman's "Goals and Policies for a National Resource and Education Network" pits the need for open access to information in a democracy against the claims of intellectual property rights as these issues relate to online information. Though a well-written and clear treatment of the subject, the paper introduces no new issues.

This volume provides a good opportunity to reflect on the value of publishing conference proceedings two years after the conference. Given the distinguished slate of presenters and authors, their work was undoubtedly of great interest to the networking and library communities at the time of the conference in 1991. Rapid publication of these papers would have been of great value to readers. Authors of the caliber of those represented in the present volume did not, of course, keep their pens or voices in check while these proceedings made their way from podium to printed page. The insights of the authors had already effectively spread through the library and networking communities by the time *Networks for Networkers II* was published. In fact, thinking on the issues discussed in the volume has not stood still but has developed and changed. The publication of *Networks for Networkers II* is, then, rather anticlimactic at best and redundant at worst. Though the volume might be recommended as an introduction to networking issues for librarians just beginning to explore this world, it is not recommended for libraries that have already been collecting in the networking area.—*Brian Sealy, University of Michigan.* ■■

***Wide-Area Networks in Libraries: Technology, Applications and Trends.*** Ed. Gregory Zuck and Bruce Flanders. Supplements to *Computers in Libraries,* 58. Westport, Conn.: Meckler, 1992. 152p. $39.50 (ISBN 088736-841-7).

This addition to Meckler's expanding series of monographic supplements to *Computers in Libraries* is particularly timely, given the exponential increase in articles on the Internet and its implications for libraries. Recent federal legislative initiatives toward a national information infrastructure (NII), coupled with the rise in commercial interest in the net, makes the vision of a virtual library at least a little more possible, if not more probable, every day.

This current discussion of wide area network (WAN) technology is a collection of seven essays. The first essay, by Gregory Zuck, gives readers a brief (nine pages) overview of some basic telecommunications terms and concepts in wide area networking. Following this are six more substantial essays on six very specific, and very different, operational WANs. Individually, these projects roam all over the map, both literally (from Kansas to the UK to the former Yugoslavia) and technologically. In fact, the essays' technological range is so broad as to render the book less than useful as a guide to WANs. What redeems the collection is the fact that the very differences among all six projects manage to present readers with compelling evidence of WAN technology's already substantial impact on the mission and methods of libraries.

Foremost among these essays is Bruce Flanders' description of the KICNET project in Kansas, an outstanding example of technical cooperation among public, academic, medical, and school libraries. Len Simutis' description of the OhioLINK project likewise reads well and presents a variation on the statewide theme. Peter Stone's essay on the

UK JANET network does a fine job of presenting the state of that network up to mid-1992, although rapid advances in international networking in the United Kingdom will shortly reduce the piece to historical description. Other essays include a report on a WAN linking Arizona State University campuses and a description of WAN systems employed by CERN (European Laboratory for Particle Physics), which, oddly, makes no mention of World-Wide Web (WWW), perhaps the most interesting offshoot of networking technology to emerge from the high-energy particle physics community.

Perhaps the most interesting piece is Primoz Juznic and Emil Hudomalj's "Wide-Area Network Applications for Libraries in Slovenia." A prefatory note explains that the authors were, at the time of writing, attempting to upgrade a telecommunications network during the ongoing Yugoslav civil war. Save for one passing reference to the "disintegration of Yugoslavia," the authors call no attention to their situation. (A quick check of the Internet reveals a functioning gopher at the University of Ljubljana, the authors' home institution, in June 1993). The essay is notable for documenting conditions under which most North American librarians hope they will never labor. It serves, in a strange way, as a necessary reality check to the often-surreal world of telecommunications.

In short, this volume is a useful addition to Meckler's series, one that can introduce librarians to a range of WAN applications in operation, and sketch, in broad outlines, how the networking of geographically distant libraries inevitably results in substantial operational changes on a local level.—*Patrick Flannery, American University.* ■■

# Software Reviews

**CA-Cricket Graph III, v.1.0.** Computer Associates International, One Computer Associates Plaza, Islandia, NY 11788; (516) 342-5224. System requirements: Macintosh Plus or later, System 6.02 or higher, 1 MB RAM (2 MB with System 7). Price: $195.

**DeltaGraph Professional, v. 2.0.3.** DeltaPoint, Inc., 2 Harris Court, Suite B-1, Monterey, CA 93940; 1-800-367-4020. System Requirements: Macintosh Plus or later, System 6.02 or higher, 2 MB RAM. Price: $295.

With so many charting options available, one might wonder what advantage a stand-alone charting package could offer. After all, most spreadsheets and databases, and even some word processors, provide integrated charting. But these modules are clumsy and inflexible. Dedicated charting programs are easier to use, provide far more flexibility, offer a wider variety of charts, and produce more attractive output than their integrated counterparts.

Computer Associate's CA-Cricket Graph III and DeltaPoint's DeltaGraph Professional are two popular, general-purpose charting packages for the Macintosh. First introduced in 1986, Cricket Graph supplanted Microsoft's Chart to be the unquestioned leader for many years. But it nearly disappeared when Cricket, Inc., declared bankruptcy. Computer Associates purchased rights to all Cricket software in 1990 and released a revised version of CA-Cricket Graph III early this year. DeltaGraph Professional was originally introduced in 1989 and was updated in 1992. In the three years since the last Cricket upgrade, DeltaGraph Professional has become the leading Macintosh charting program. This review compares three areas of software performance: data import and export, data manipulation, and ease of creating and editing a chart.

## DATA IMPORT AND EXPORT

One advantage of dedicated charting software is that it manages the charting of very large datasets gracefully. But data must somehow be input into the packages' data page. While keying a few numbers is fine, it is impractical to type the contents of a large spreadsheet or database. In addition, completed charts are often transferred to any number of documents. Consequently, efficient import of data and export of finished charts is critical to the functionality of a charting package.

DeltaGraph Professional imports not only ASCII and SYLK files, but Lotus, Trapeze, Excel 2.2, 3.0., or 4.0, and even Cricket Graph 1.2 or 1.3 (but not CA-Cricket Graph III)

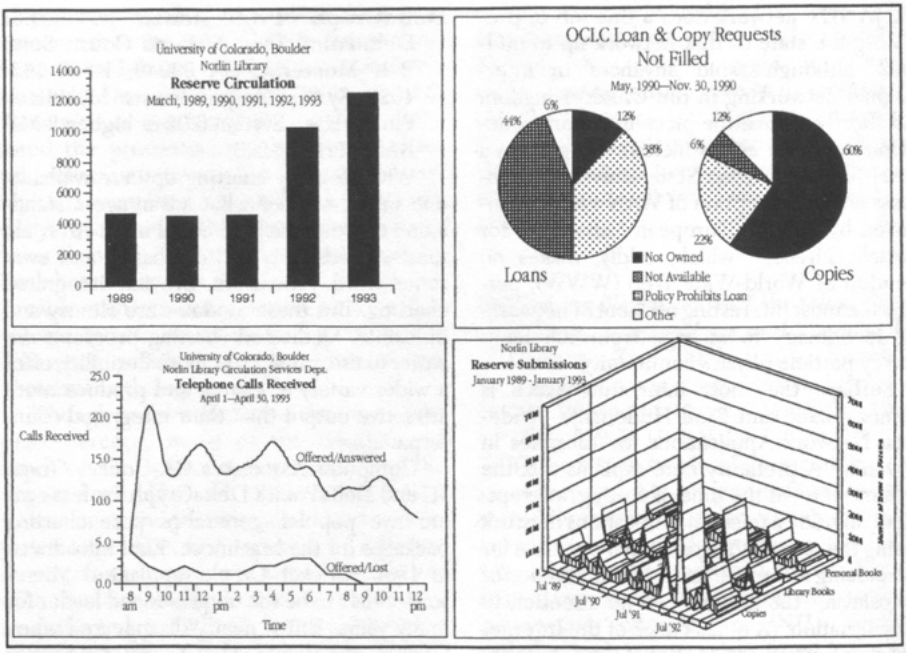



*Figure 1. Examples of charts that may be created and edited in DeltaGraph Professional.*

files. CA-Cricket Graph III only imports DeltaGraph, ASCII, and SYLK files. This requires spreadsheet data always be delivered in ASCII format. All spreadsheets can save in ASCII. But if you are in an environment where you cannot control the format in which data is delivered, the ability to directly open an Excel or Lotus file from within the charting application is valuable.

Ideally, sophisticated users should be able to link spreadsheet data directly to the charting package for continuous import of data and then publish the updated chart directly to an end document. That is, a change in the spreadsheet data automatically updates the charting package and, in turn, the updated chart is sent to a word processing document. This is only possible with DeltaGraph Professional.

CA-Cricket Graph III, while thirty-two-bit clean, does not support Apple Events. DeltaGraph Professional, however, takes full advantage of System 7, including Publish and Subscribe. Consequently, System 7 users running software that also supports Publish and Subscribe can automatically update charts when changes are made to the source data. For System 6 users, DeltaGraph Professional offers a Link feature that automatically relates Excel files to DeltaGraph Professional files. Linking works similarly to Publish and Subscribe, although it is somewhat less flexible. Whenever a column of Excel data is copied into DeltaGraph Professional using the Link command, changes to the Excel spreadsheet are automatically sent to the DeltaGraph Professional chart. Unfortunately, only one column per document can be linked.

DeltaGraph Professional exports data as tab-delimited text. Charts, however, can be exported as Adobe Illustrator, Aldus Freehand, EPSF, and PICT images. System 7 users can take advantage of Publish and Subscribe to export data or text to selected applications automatically. Cricket Graph III only exports data in ASCII and graphics in PICT. However, ASCII export delimiters and filters can be customized so that other applications can more easily interpret the incoming data.

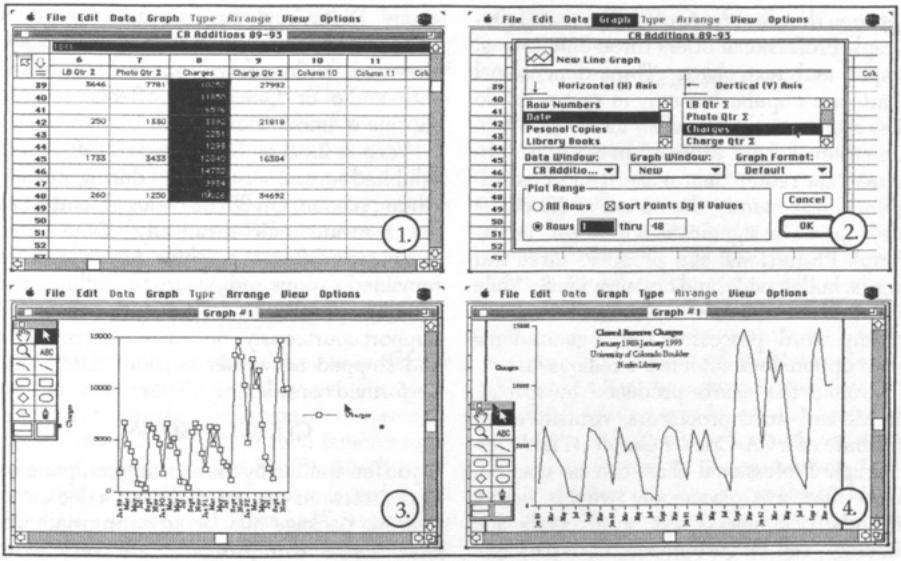



*Figure 2. The process of creating charts in DeltaGraph Professional and CA-Cricket Graph III is very similar. Here, a chart is being created in CA-Cricket Graph III: 1. Data to be charted is selected; 2. The axis' orientation is defined; 3. The default chart is viewed; 4. The chart's elements are edited to produce the final version.*

## DATA MANIPULATION

Often data are collected in such a way that is not immediately graphable. For example, data imported from a spreadsheet may need to be resorted or smoothed. Occasionally it may even be necessary to do mathematical calculations and then chart the results.

Cricket Graph was originally designed for scientific and technical purposes and that heritage is still evident. CA-Cricket Graph III offers an impressive array of statistical, trigonometric, and financial functions that can be applied either to columns or cells. CA-Cricket Graph III can modify or generate new data using dynamic column formulas, mathematical calculations, and transformations.

While not providing the flexibility or sophistication of CA-Cricket Graph III, DeltaGraph Professional does perform basic data manipulation such as sort, smooth, and count frequency. Perhaps because it is less extensive, DeltaGraph Professional's Formula Builder is more intuitive than CA-Cricket Graph's and is probably adequate for most library applications.

Both DeltaGraph Professional and CA-Cricket Graph III offer curve fitting for line, step, and scatter graphs (note the line chart in figure 1). This is achieved by fitting Bezier curves to the data. It is easier to accomplish in DeltaGraph Professional, but Cricket offers more flexibility.

## CREATING AND EDITING CHARTS

Both CA-Cricket Graph III and DeltaGraph Professional use the same basic steps to generate a variety of charts. Data are either entered or imported into a spreadsheetlike page with columns and rows of data. Data are charted by highlighting columns, rows, or sections of the page, and selecting a Graph command from the menu. Editing the chart is accomplished either by selecting menu items or double-clicking on particular elements and responding to a series of dialog boxes (see figure 2).

Both programs are capable of generating standard types of charts—bar, pie, and line—but DeltaGraph Professional offers more variations of the same chart. For example, while CA-Cricket Graph III offers a line chart, DeltaGraph Professional offers four different line charts: a high-low, a range, and a step, as well as the traditional line chart.

Probably the most noticeable difference

between the chart types offered is that DeltaGraph Professional offers three-dimensional graphs and text charts. Three-dimensional charts are popular but only in rare cases do they more accurately display data than a two-dimensional chart. More often 3D charts obscure data rather than make it more accessible. Consequently, the option of 3D charts may not be a significant advantage. DeltaGraph Professional also produces three text charts: bullet, table, and organizational. While bullet charts and most tables can be created in any word processor, the organizational chart option does automate a tedious task.

Unlike the charts produced by spreadsheets and word processors, virtually every attribute of a CA-Cricket Graph III and DeltaGraph Professional chart can be manipulated. Size, axis placement, symbols, labels, legends, tick marks, text, even colors and shadings, can be customized. Both packages include MacDraw-like toolboxes so that geometric objects and text can be added to the chart. While both packages edit similarly, CA-Cricket Graph III's approach is faster and more intuitive. It relies more upon double-clicking chart elements to change attributes rather than, as with Delta Graph Professional, having to search myriad menus for the correct dialog box.

Editing complex charts proved faster in CA-Cricket Graph III than in DeltaGraph Professional. Three-dimensional charts sometimes required several minutes to redraw on a Macintosh LC II. CA-Cricket Graph III, however, redrew even complicated charts quickly and smoothly. It also includes the ability to turn off redraws so that users can work quickly on entry-level machines.

DeltaGraph Professional offers a spelling checker and an impressive presentation component. The spelling checker links to a 117,000-word dictionary and a customizable user dictionary. Like spell checkers in word processors, it finds and replaces misspelled words in any DeltaGraph window. This is particularly valuable when editing text-intensive charts.

DeltaGraph Professional has several features specifically for creating presentations. Its slide show component presents any series of charts, PICT, EPSF, or even QuickTime movies, each with audio effects and unique transitions. Slide shows can be programmed to run automatically or on an operator's queue. DeltaGraph Professional even prints audience handouts from slide show images. However, slide shows, particularly those with audio or QuickTime effects, require enormous amounts of RAM.

Version 2.0.3 of DeltaGraph Professional exhibited several serious bugs during a chart-editing session. On 68000 series Macintoshes the software quit erratically. Even after moving to a different machine, however, there remained serious redraw, text-handling, and printing problems. DeltaPoint's Technical Support courteously confirmed each problem and shipped an earlier version (2.0.2) that performed considerably better.

## CONCLUSIONS

If you feel limited by the charting components of the software you currently use, a dedicated charting package may be an appropriate option. These two programs are very close functionally, and both produce charts quickly and allow efficient editing and data update. They are not interchangeable, however.

CA-Cricket Graph III's origins as a scientific and technical charting package are apparent in the sophisticated mathematical calculations that can be performed on data. For librarians regularly writing reports or analyzing research data, CA-Cricket Graph III is probably the better of the two packages. In addition to excellent data-manipulation capabilities, CA-Cricket Graph III provides fast and easy chart editing and customizable data export commands. However, the lack of System 7 support, most notably of Publish and Subscribe, may be a serious drawback for some spreadsheet users.

If your data regularly originate from Excel, it may be worthwhile to examine DeltaGraph Professional. For an additional $100, Delta Graph Professional offers more charts than CA-Cricket Graph III, a wider range of predefined data-import capabilities, and impressive presentation components. The presentation capabilities are so extensive that if you regularly create on-screen presentations, slides from charts, or audience handouts, DeltaGraph Professional is the obvious choice.—*Scott Seaman, University of Colorado at Boulder.* ■■

e[disk.] Alysis Software Corp., 1231 31st Ave., San Francisco, CA 94122. System re-

quirements: Macintosh Plus or higher; system software version 6.0.5 or later; hard drive, removable cartridge drive, or floptical drive. License: The standard single-user license provides for use on a single Macintosh computer. Single-copy price: $149.95; multi-user price: fewer than 100 copies, $67.50 per copy; more than 100 copies, $60 per copy. Site licensing available.

## DISK DRIVER-LEVEL COMPRESSOR

e[disk] works with the disk driver to compress and decompress files as they are written to and read from a disk. This type of compression differs from that found in such familiar utilities as CompactPro and Stuffit, which compress files in place and are referred to as file-level compressors." Once e[disk] is installed, it executes file compression transparently as the disk is accessed; there is no need to issue a compress" command. e[disk] can effectively double the capacity of the average hard disk, removable cartridge disk, or floptical disk (but not a floppy disk). For disks with a preponderance of highly compressible graphics or database files, e[disk] may even expand capacity three or four times, if the user chooses a high-compression option. The chief attraction of a utility such as e[disk] for acquiring more disk space is that it is less expensive than buying a new disk drive.

## EXPANDING A DISK

Though it takes e[disk] only fifteen minutes to expand an eighty-megabyte hard disk to twice its capacity, the total time for expansion must include the time it takes to back up all files on the disk and to reload files and reinstall software once the disk is expanded, for the e[disk] installation erases all files on a disk. The total time for this one-time installation may be a factor to weigh when considering the alternatives of buying a new disk or using e[disk] on an existing disk. Of course, if e[disk] is used to expand a new disk, backup and reloading will not be a factor.

After running e[disk], the Mac system software will show the expanded disk size. The system software will also display files in their true sizes rather than in a compressed file size. Though files compressed with other file-compression utilities, such as CompactPro, may be stored on an e[disk]-expanded disk, no further compression than that offered by e[disk] can be achieved.

## DISK PERFORMANCE

After installing e[disk], there is some degradation of disk performance, due to the fact that the e[disk] software performs compression routines each time there is a disk read or write. The disk cache feature of e[disk] allows one to speed up disk access under e[disk], though it takes its toll in the increased amount of RAM devoted to system overhead. The disk cache or "delayed write" feature sets aside a user-specified amount of RAM to which to write data; when the software detects no user activity, it compresses data in the cache and writes it to disk. Compression and writing will also occur, regardless of user activity, if the cache becomes full. The larger the cache, the less likely this is to occur. The maximum cache size is ten megabytes, but it should be noted that cached memory is not available for running applications. The *User's Guide* recommends setting the System 7 caching feature to the minimum 32K size and relying on the e[disk] caching feature.

Testing revealed that disk performance varied with different applications. The time required to open the compact word processing program WriteNow 3.0 was virtually the same before and after installing e[disk] with a 512K cache. WordPerfect 2.1 and Claris Resolve experienced 50 percent slower opening times with the expanded disk and a 512K cache. Increasing the cache size to 640K made no difference in the opening times for these applicaions. The time needed to open individual files within applications was not noticeably affected by e[disk].

It should also be noted that one cannot use virtual memory with an e[disk]-expanded hard disk. Alysis hopes that a future maintenance upgrade will include virtual memory compatibility. (This feature may be available by the time this review is published.)

## OTHER FEATURES

e[disk] features a disk optimizer, which scans a disk and makes sure that files are stored in the most efficient manner. The user can choose to have this feature run in the background when the system is not in use or can run it with the Repair command. Using this command, it took about twelve minutes to scan and optimize an eighty-megabyte hard disk in test-

ing. The Repair command also checks the disk's data integrity in addition to performing disk optimization. The Update feature allows the user to install a new version of e[disk] without having to reinitialize the hard disk (and thus back up and reload the disk). e[disk]'s online help feature merely provides an online duplicate of the thirty-four-page *User's Guide*.

## CONCLUSION

e[disk] offers libraries contending with tight budgets and disk-hungry applications a relatively inexpensive way to increase disk space without having to drain equipment budgets with the purchase of new hardware. In testing, no problems with data integrity on a hard disk were encountered. Two factors must figure into a decision on whether to purchase an additional external disk drive or e[disk]. First, disk performance is somewhat degraded. Disk caching offers some relief from this problem but at the expense of decreasing the amount of RAM available for applications. Second, a considerable investment of time may be required to install e[disk] on an existing disk due to the backup and reloading time needed.—*Brian Sealy, University of Michigan.* ■■

# Other Recent Receipts

Association of Research Libraries. *Systems and Procedures Exchange Center: SPEC Kit 184, Interlibrary Loan Trends: Making Access A Reality, May 1992.* Washington, D.C.: Association of Research Libraries (Office of Management Services, 1527 New Hampshire Ave., NW, Washington, DC 20036), 1992. 184p. paper, $40, $25 ARL members (no ISBN).

Association of Research Libraries. *Systems and Procedures Exchange Center: SPEC Kit 186, The Emerging Virtual Research Library, July/August, 1992.* Washington, D.C.: Association of Research Libraries (Office of Management Services, 1527 New Hampshire Ave., NW, Washington, DC 20036), 1992. 184p. paper, $40 $25 ARL members (no ISBN).

*CD-ROM Implementation and Networking in Health Sciences Libraries.* Ed. M. Sandra Wood. New York: Haworth, 1993. 165p. $24.95 (ISBN 1-56024-381-3).

*Collection Management for the 1990s.* Ed. Joseph J. Branin. Chicago: American Library Assn., 1992. 200p. paper, $25, $22.50 ALA members (ISBN 0-8389-0595-1) (ALA Order Code 0595-1-0010).

Cortez, Edwin M., and Tom Smorch. *Planning Second Generation Automated Library Systems.* Westport, Conn.: Greenwood, 1993. 232p. $47.95 (ISBN 0-313-28361-3).

*Developing Library Staff for the 21st Century.* Ed. Maureen Sullivan. New York: Haworth, 1992. 118p. $24.95 (ISBN 1-56024-389-9).

Dewey, Barbara I., and Sheila D. Creth. *Team Power: Making Library Meetings Work.* Chicago: American Library Assn., 1993. 130p. paper, $20 (ISBN 0-8389-0616-8).

Emery, Charles D. *Buyers and Borrowers: The Application of Consumer Theory to the Study of Library Use.* New York: Haworth, 1993. 181p. $26.95 (ISBN 1-56024-183-7).

Gill, Suzanne L. *File Management and Information Retrieval Systems: A Manual for Managers and Technicians.* 3d ed. Englewood, Colo.: Libraries Unlimited, 1993. 267p. $29.50 (ISBN 1-56308-050-8).

Glazier, Jack D., and Ronald R. Powell. Qualitative Research in Information Management. Englewood, Colo.: Libraries Unlimited, 1992. 238p. $35 (ISBN 0-87287-806-6).

Hamilton, Marsha J. *Guide to Preservation in Acquisition Processing.* Acquisitions Guidelines, no. 8. Chicago: American Library Assn., 1992. 34p. paper, $7, $6.50 ALA members (ISBN 0-8389-0611-7) (ALA Order Code 0611-7-0010).

*Information Management and Organizational Change in Higher Education: The Impact on Academic Libraries.* Ed. Gary M. Pitkin. Westport, Conn.: Meckler, 1992. 152p. $42.50 (ISBN 088736-842-5).

*Interlibrary Loan of Alternative Format Materials: A Balanced Sourcebook.* Ed. Bruce E. Massis. New York: Haworth, 1993. 196p. $21.95 (ISBN 1-56024-394-5).

McDaniel, Julie Ann, and Judith K. Ohles. *Training Paraprofessionals for Reference Service: A How-to-do-it Manual.* New York: Neal-Schuman, 1993. 184p. $37.50 (ISBN 1-55570-084-5).

*Networking and the Future of Libraries: Proceedings of the UK Office for Library Networking Conference, April 2–5, 1992.* Ed. John W. T. Smith. Westport, Conn.: Meckler, 1993. 220p. paper, $45 (ISBN 0-88736-863-8).

Olson, Nancy B. *Cataloging Computer Files.* Lake Crystal, Minn.: Soldier Creek, 1992. 123p. paper, $25 (ISBN 0-936996-47-1).

Pennington, Catherine A. *Microcomputer Software Selection for the Law Library: Part 2. Library and Legal Software.* Dobbs Ferry, N.Y.: Glanville, 1992. 146p. paper, $100 (ISBN 0-87802-088-8).

Rowley, Jennifer, and Shelagh Fisher. *Bookshelf: A Guide for Librarians and Systems Managers.* Brookfield, Vt.: Ashgate, 1992. 272p. $79.95 (ISBN 1-85742-008-X).

*Sci-Tech Libraries of the Future.* Ed. Cynthia Steinke. New York: Haworth, 1993. 238p. $29.95 (ISBN 1-56024-447-X).

*Science Librarianship at America's Liberal Arts Colleges: Working Librarians Tell Their Stories.* Ed. Tony Stankus. New York: Haworth, 1992. 227p. $24.95 (ISBN 1-56024-357-0).

*Serials Cataloging: Modern Perspectives and International Development.* Ed. Jim E. Cole and James W. Williams. New York: Haworth, 1993. 415p. $75 (ISBN 1-56024-281- 7).

*Thinking Robots, an Aware Internet, and Cyberpunk Librarians: The 1992 LITA President's Program: Presentations by Hans Moravec, Bruce Sterling, and David Brin.* Ed. R. Bruce Miller and Milton T. Wolf. Chicago: American Library Assn., 1992. 200p. paper, $22, $19.80 LITA members (ISBN 0-8389-7625-5).

*Vendor Evaluation and Acquisition Budgets.* Ed. Sul H. Lee. New York: Haworth, 1992. 142p. $24.95 (ISBN 1-56024-253-1).

Woodworth, Anne, and James F. Williams. *Managing Economics of Owning, Leasing and Contracting Out Information Services.* Brookfield, Vt.: Ashgate, 1992. 204p. $49.95 (ISBN 1-85742-018-7). ■■

# Letters

To the Editor:

I read with interest the communication by Terry Ballard and Arthur Lifshin in the June 1992 issue of *ITAL* entitled "Prediction of OPAC Spelling Errors through a Keyword Inventory." They indicated that because many OPAC records are received from national databases, similar errors should be found in other OPACs, which they said was verified through a look at other large library databases in the New York metropolitan area. I decided to apply their results to another set of libraries and randomly identified a group of ten United States academic libraries of various database sizes all using NOTIS for their OPACs. Ballard and Lifshin's list of most frequently occurring spelling mistakes (p. 144) was searched as keywords in each institution's OPAC. Table 1 represents the number of times each misspelled word occurred in each institution's database.

Table 1 also indicates the size of each institution's database as reported in NOTIS' *1992–1993 Customer Directory.* The percentage of each database represented by these misspelled words is shown. The table is arranged using Ballard and Lifshin's list of frequently misspelled words in descending order. My survey results would require the word list to change slightly from most to least frequently misspelled:

Reseach
Adminstration
Universty
Commerical
Committe
Occurence
Developement
Associaton
Bibiography
Philosphy
Educaton
Questionaire
Pantomine
Responsiblity
Behaviorial
Linquistic
Intergral
Psychanalytic

It would be interesting to examine closely the list of records containing these misspelled words to determine whether they were created locally or were downloaded from a national bibliographic utility. Another test would be to analyze transaction logs from OPACs to discover whether these misspelled words would affect patron use of the OPAC. Ballard and Lifshin's results should be applied to other OPACs to see how universal the problem of misspelled words is to our desire to improve patron access to our holdings.

Perhaps this little study would be appropriate for your communications section.—*Ed Goedeken, Humanities Bibliographer, Parks Library, Iowa State University, Ames.* ■■

*Table 1.*

| Words | AD | PU | IS | CO | NW | AL | TAM | CL | OK | VE | CS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Adminstration | 10 | 56 | 17 | 1 | 87 | 63 | 202 | 11 | 0 | 119 | 3 |
| Commerical | 8 | 32 | 35 | 5 | 37 | 92 | 128 | 0 | 0 | 57 | 14 |
| Questionaire | 8 | 5 | 3 | 6 | 13 | 7 | 11 | 4 | 0 | 7 | 1 |
| Universty | 7 | 92 | 43 | 1 | 6 | 148 | 126 | 39 | 5 | 73 | 3 |
| Intergral | 5 | 2 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 0 |
| Committe | 4 | 35 | 10 | 49 | 36 | 42 | 76 | 17 | 0 | 39 | 4 |
| Developement | 4 | 18 | 12 | 27 | 14 | 19 | 46 | 5 | 11 | 15 | 1 |
| Educaton | 4 | 9 | 4 | 6 | 7 | 18 | 14 | 3 | 0 | 14 | 1 |
| Pantomine | 4 | 1 | 2 | 17 | 4 | 12 | 2 | 1 | 0 | 2 | 0 |
| Occurence | 4 | 15 | 18 | 35 | 19 | 17 | 60 | 17 | 0 | 15 | 0 |
| Philosphy | 4 | 19 | 9 | 11 | 5 | 25 | 14 | 5 | 0 | 15 | 0 |
| Psychanalytic | 4 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 |
| Reseach | 4 | 81 | 45 | 87 | 23 | 97 | 235 | 17 | 0 | 152 | 6 |
| Responsiblity | 4 | 3 | 7 | 5 | 8 | 6 | 0 | 4 | 0 | 4 | 4 |
| Associaton | 3 | 31 | 21 | 5 | 35 | 1 | 30 | 18 | 0 | 18 | 7 |
| Behaviorial | 3 | 7 | 3 | 2 | 4 | 1 | 6 | 0 | 0 | 11 | 0 |
| Bibiography | 3 | 7 | 18 | 5 | 45 | 0 | 0 | 3 | 0 | 31 | 4 |
| Linquistic | 3 | 1 | 3 | 0 | 0 | 4 | 4 | 1 | 0 | 5 | 0 |

| Institution Code | Institution | Database Size | Errors as Percentage of Database (%) |
|---|---|---|---|
| AD | Adelphi University | 310,000 | .027 |
| PU | Purdue University | 575,000 | .072 |
| IS | Iowa State University | 740,000 | .033 |
| CO | Cornell University | 4,700,000 | .005 |
| NW | Northwestern University | 1,300,000 | .026 |
| AL | University of Alabama | 650,000 | .085 |
| TAM | Texas A&M | 1,200,000 | .079 |
| CL | Clemson University | 450,000 | .032 |
| OK | University of Oklahoma | 900,000 | .002 |
| VE | University of Vermont | 1,400,000 | .004 |
| CS | Cleveland State University | 500,000 | .009 |

# Information Technology—the very latest resource guides from ALA Books

## IT's the way of the future!

**New!**
**Headings for Tomorrow: Public Access Display of Subject Headings**
*Martha M. Yee, editor*
*ALCTS Subjects Analysis Committee*

An aid for librarians interacting with systems designers in making decisions on the design of displays of multiple subject headings. Uses subject headings derived from LCSH, but the principles explored can be applied to similar lists of highly structured subject heading strings.

$15.00pbk. 70p.
ALA Order Code 3415-5-0011 1992

**New!**
**FISCAL Directory of Fee-Based Research and Information Services**
*Steve Coffman, compiler*

Copublished by the Los Angeles County Public Library, the fourth edition of the *FISCAL Directory* contains approximately 550 detailed listings of libraries and commercial sources (domestic and international) providing fee-based information services.

$65.00pbk. 650p.
ALA Order Code 2161-2-0011 1992

**202(+) Software Packages to Use in Your Library: Descriptions, Evaluations, and Practical Advice**
*101 Micro Series*
*Patrick R. Dewey*

A reference tool of many uses to librarians in all types of libraries, the second edition of this ALA bestseller brings the reader detailed, up-to-date, user-friendly analyses of more than 250 recommended software packages.

$27.50pbk. 190p.
ALA Order Code 0582-X-0011 1992

**ALA Books**
**American Library Association**
**50 East Huron Street**
**Chicago, IL 60611**
**800-545-2433; press 7 to order**

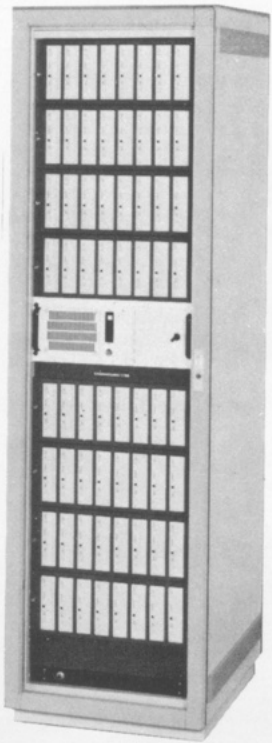
# TODD ENTERPRISES INC.
# NETWORK SOLUTIONS

administrated by **Lotus CD/Networker** software

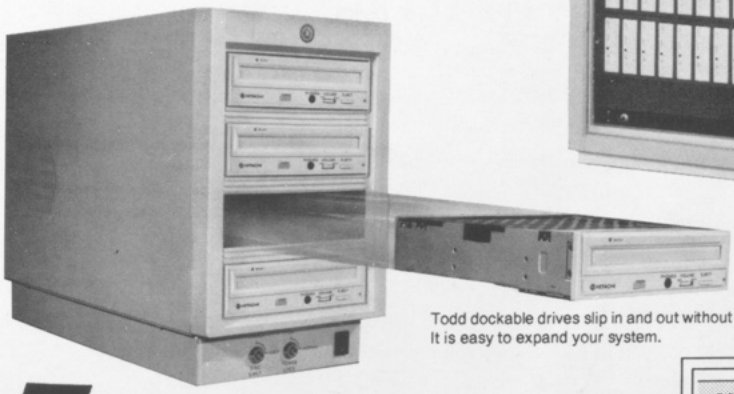
## from 4 to 56+

### CD-ROM DRIVES in ONE CABINET

*Todd massive data retrieval Total Access Systems feature dockable drives and security keylocks. Todd uses Hitachi drive mechanisms which have an access time of under 300 msec and are known for reliability. They have a two-year warranty. The security lock feature of Todd Total Access Systems is useful in public areas. Todd services include hardware design and development, customized systems, and network installation and management. Administrating software for Todd Network Solutions is Lotus CD/Networker. Todd CD-ROM hardware will operate within your workstation environment. Call* **800 445-TODD** *or (718) 343-1040 for more information.*



Lotus CD/Networker software allows Todd's Total Access storage unit, in a single cabinet, to interface with your workstation environment.



Todd dockable drives slip in and out without tools or wiring. It is easy to expand your system.


TODD ENTERPRISES INC.
224-49 67th Avenue • Bayside, NY 11364
718 343-1040 • 800 445-TODD • FAX 718 343-9180


*available on GSA contract*

established 1971
Todd Enterprises
over two decades of EXCELLENCE

# There's More To Know.

Today's patrons expect easy access to information sources via our growing electronic networks. And they want full text and image capabilities. Advances in connectivity, the Z39.50 protocol, and the Internet have created a global network from which to draw information.

# There's More In Store.

Truly open systems, which provide access to this global network, are becoming a reality. Open systems utilizing client/server technology offer greater flexibility through the integration of components from multiple vendors. And easy-to-use graphical interfaces (GUIs) are increasing productivity for both patrons and staff. Open systems are reshaping the way we work.

# There's More To NOTIS.

NOTIS Systems has been a leader in library automation for over 10 years, and today supports over 200 libraries. We're developing open systems that will help you provide expanded patron access and give you greater system flexibility and cost-effective solutions. NOTIS was the first to offer Z39.50-compliant products like *InfoShare*,™ a database access server for UNIX platforms; *ProPAC*,™ a windows-based client for PCs and workstations; and *PACLink/PACLoan*,™ a collection sharing system. Keep an eye on the future. Keep an eye on NOTIS.

NOTIS SYSTEMS, INC.
AN AMERITECH COMPANY

708.866.0150