Volume 12,
Number 2

ISSN 0730-9295

Information Technology and Libraries

June 1993

CONTENTS

ITLBDC 12(2) 169–304 (1993); Inf Tech Libr

ISSN 0730-9295

Information Technology and Libraries

Volume 12, Number 2: June 1993

CONTENTS

ITLBDC 12(2) 169–304 (1993); Inf Tech Libr

Information Technology and Libraries

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

Information Technology and Libraries publishes material related to all aspects of library and information technology. Some specific topics of interest are: Automated Bibliographic Control, AV Techniques, Communications Technology, Cable Systems, Computerized Information Processing, Data Management, Facsimile Applications, File Organization, Legal and Regulatory Matters, Library Networks, Storage and Retrieval Systems, Systems Analysis, and Video Technologies. *ITAL* welcomes unsolicited manuscripts. Submissions should follow the guidelines stated under "Instructions to Authors" on page 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*, Arizona State University, Tempe, AZ 85287. 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 Gibson-Porter, 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. ∞

Editorial: The Literate Librarian

Thomas W. Leonhardt

Early this year we received a newspaper clipping with a note attached to it—a letter to the editor, in fact—from Nancy Sosnik of Raleigh, North Carolina. The letter appears in this issue's "Letters to the Editor" section, but we want to draw special attention to it and to quote a bit from the article.

The clipping that Ms. Sosnik sent me is an article from the December 28, 1992, issue of the *Raleigh News and Observer.* Written by staff writer Trish Wilson, it is titled "Computer Wizards Supplanting Book Lovers at Library Desk: A New Breed of Knowledge Keepers."

The article contains much of the hyperbole one has come to expect in articles about electronic information, along with phrases such as "The New Librarians, a breed of knowledge keepers who are leaving the book stacks to 'hang ten' on a tidal wave of electronic information." All in all, however, it accurately describes some of the ways librarianship is changing.

The author interviewed the deans of the library schools at UNC Chapel Hill (Barbara Moran) and North Carolina Central University in Durham (Benjamin Speller, Jr.) and librarians at Duke University's Perkins Library. All note the demands for new approaches by librarians. Speller contends that "the ideal librarian must be well-rounded in both the sciences and the humanities . . . [s]omeone who is well-balanced, likes to work with people, and is not full of himself." Moran observes about this new electronic information age: "It's like the printing press. I think we're on the brink of a fundamental change in our society, and we don't know where it's going to go."

There is more of this kind of talk, but that is not why we are drawing your attention to the letter and the clipping that accompanied it. We are pleased that the article gained a new member for LITA, but we were bothered by the overall tone of the piece.

The opening two paragraphs of the article read: "If a candidate for a librarian's job walked into Rebecca Vargha's office saying she had a 'great love for working with books,' Vargha would drop the application into the closest paper shredder."

"'I would definitely not hire them,' said Vargha, associate librarian at the National Humanities Center in Research Triangle Park. 'I'm a librarian, but I don't really do books. It's all computer. I think of myself as an information surfer.'"

We don't know Ms. Vargha, but we suspect that there was some hyperbole at play here, too. For one thing, the National Humanities Center houses scholars who have a great love for working with books. In fact, they have a great love for reading books and for writing books, too. The librarians who serve those scholars borrow lots of books for them, and why not? They are within spitting distance of three members of the Association of Research Libraries: Duke University, the University of North Carolina at Chapel Hill, and North Carolina State University.

The implied message of the article is that if we learn how to manipulate computers and find information, we don't need to love books. But for us, implied in the love of books is a love of knowledge and learning, and an ability to discriminate among the information sources and the information so that we and those we serve can choose what is right for them and for us. Information and knowledge may overlap but they are not the same thing.

One underlying current in the article is the question of whether society will need librarians in the future, and Dean Moran addresses that adequately. But no one argues for the user of technology who is both knowledgeable about the computer (trained) and able to think (educated). The love of books alone is not sufficient reason to become a librarian, and it is not sufficient reason to hire someone, but without the love of knowledge and the intellectual curiosity and talent to continue learning, love of and knowledge about computers is not enough, either.

The person who reads with comprehension and writes well is an educated person. We want such people in librarianship, and we want them with the personal attributes that Ben Speller describes. We don't need people who learned all they know surfing on the Internet; a few days of such surfing should convince you of that.

To follow up on Ms. Sosnik's suggestion that this topic (is it "librarians, computers, and books"?) might be a good idea for future discussions in *ITAL*, we invite you to submit guest editorials or even long essays on the subject.

Welcome to LITA, Ms. Sosnik, and thanks for your letter. ■■

Selecting Electronic Alternatives

Richard W. Meyer

Academic libraries today are faced with the prospect of numerous alternatives to traditional printed indexes to the primary literature. Not only are many indexes made available by publishers as CD-ROM products, but most of these products can be acquired on computer tape for local mounting. This happy situation is further augmented by the availability of these products through mediated search services such as Dialog and end user services such as OCLC FirstSearch. Choosing the best alternative becomes a matter of estimating demand and comparing costs for each type of access. These estimates are complicated by lack of information on how often patrons will use a given alternative and by lack of qualitative data. Results of comparisons made at Trinity and Clemson Universities provide some revealing benchmarks that can help to place a quantitative framework around the decision process.

This article deals with extending the online public access catalog (OPAC) by providing access to other resources, in particular to indexes. It is the intention to offer a model that can be applied to determine the choice of alternatives when extending the OPAC includes choosing among CD-ROM, end user access, mediated searching, or locally mounted files, in lieu of printed indexes. This paper does not bear upon the issue of locally generated information available through campuswide information systems; however, the natural extension of these systems should include index access.

A paper published in 1990 described some of the relevant management, cost, and behavioral issues.[1] That article examined financial issues thoroughly enough to suggest that locally mounted files are cost-effective where undergraduate use is heavy or where the price of a research database is very low. That study also pointed out that reducing the information-seeking effort of users will transfer costs to the library. That is, online, locally mounted indexes are easier to search than their printed counterparts, but they may be more costly for the library to provide. Librarians making choices of alternatives are advised to remember that the users pay some costs, and this affects the quality of the results achieved. This point should be kept in mind, because the model that will follow necessarily focuses on the direct costs of the alternatives to the library.

This presentation will not tell the whole story, but it should help in providing a baseline for libraries trying to choose between CD-ROMs and locally mounted files versus print or other search alternatives. It extends the cost study for locally mounted files published in January 1992.[2] That study established the demand curve for index access based on a comparison of activity on mediated searches versus locally mounted databases. For an individual library willing to determine a base cost per search that is acceptable, that study will assist in the process of deciding whether to mount files locally or to purchase an alternative. A brief overview of the points made earlier related to user costs are required to establish the context for the description of the model.

MANAGEMENT AND BEHAVIORAL ISSUES

It should be a major goal of every library to provide access to information at the lowest possible cost to users. This involves reducing

Richard W. Meyer is Director of the Library at Trinity University, San Antonio, Texas.

barriers to user access to information, minimizing the number of bibliographic interfaces, educating users, optimizing interface design, and providing efficient data handling. All of these issues affect the ease of access to information that users experience. To put it in economic terms, each issue must be addressed if the personal investment of time involved in seeking information is to be minimized for each patron. The importance of this is illustrated by the effect on demand for searches when mediated searching is replaced with locally mounted files. Demand goes up by two orders of magnitude, which shows that there is a substantial cost to users for some search methods apart from the payment of hard currency. More on this later.

Certain behavioral issues relate to the issue of nondollar costs as well. Interface design, bibliographic education, and interface standardization need to be carefully considered to maximize awareness of the OPAC as well as to make it as inviting as possible. Giving a friendly personification to the OPAC to improve its invitability to the novice user tends to offset some of the intimidating aspects of online systems. The interface design should take into account other psychological issues as well. For example, student users tend to allocate a fixed amount of time to an assignment such as writing a paper that requires use of the library. Since faster access to information leaves more time for reading and writing, it is important that the interface minimize user effort to find citations. Taking these statements into consideration as fundamental assumptions of user psychology helps to yield lower cost access to information overall through better accommodation to user needs.

THE MODEL FOR LOCALLY MOUNTED FILES

It is a fundamental premise that libraries should provide access to information at no direct cost to the user. This premise worked well in a world in which a fixed number of information packages were purchased with the funds available. Once acquired, a resource could be used by any or all users of the library at no increase in cost.[3]

The introduction of mediated searching, or products purchased on an as-used (or connect time pricing) basis, introduced an economic dilemma in libraries that has engendered the great fee-or-free debate. Providing users with a selection of alternative search methods while adhering to the traditional premise of free information to all, including mediated searching, can overcome much of the concern about unpredictable costs. Therefore, Trinity University intends to establish the basic policy that users may have free access to all search methods, including mediated services. Currently, mediated searching is limited to $10 per student. Campus users are provided free access to FirstSearch with a higher limit, depending on the availability of alternatives. Experience over several years has helped to provide some estimate of user need to establish these allocations for mediated and end user searching as well as significant awareness of user characteristics. Eventually, we expect that use of this model will allow implementation of free access to information. To date, this experience has helped to establish related policies, given some fundamental assumptions.

First, across the range of mediated, CD-ROM, end user, and locally mounted files, a search should be defined consistently. In all these cases this study defines *search* as the sum of all transactions including selection of the database plus each transaction that invokes action by the search engine, where *transaction* means pressing the enter key and receiving a resulting screen. Where connect time is considered, that portion of the time spent reviewing the screens is included. This definition appears to be reasonable since it parallels the charging algorithm used by OCLC FirstSearch and allows determination of average search costs from invoices submitted by online vendors. It also represents the basic definition used at Clemson University for this and previous studies of locally mounted files.

Second, to establish the cost-effectiveness of locally mounted files it is necessary to assume that the benefit per search cannot be approximated accurately. No theory of value exists to pinpoint the value or benefit in dollars to the user of any given activity. The level of importance one student places on a given project cannot be assigned a dollar value, nor is it safe to assume that the value to one equals the value to everyone. However, it is safe to assume that students do not use mediated searching because cost exceeds benefit. Furthermore, lower costs will yield higher demand regardless of access method. That is, when local online searching is made available, student interest will increase. The research noted earlier bears this out. Providing locally mounted databases increases demand for searching of the database a hundredfold over mediated searching.[4]

Third, a marginal cost/benefit trade-off can be established for every alternative. With printed indexes, a purchase is made when the selector decides that the marginal cost is less than the marginal benefit to the user. This choice is likely to be made on subjective information and is, in many cases, based on subconscious impressions of the marginal cost/benefit trade-off. On the other hand, locally mounted files provide the necessary information to calculate readily the cost per search based on experience with mounted files. Fortunately, the cost per search may also be calculated for CD-ROM, mediated, and end user searching. In the case of CD-ROM, in-house experience allowed Trinity to develop cost-per-search calculations and to compare these to mediated and end user search systems. These comparisons can be extended to include some subjective impression of the user marginal benefit. That is, it can be assumed that the marginal benefit of a CD-ROM product exceeds the marginal cost if we have continued to carry the subscription. This experience serves to provide a marker for determining the acceptable cost per search. The rationale behind the model follows in the next section.

THE MATRIX OF ALTERNATIVES

At the outset, we need to recognize that the number of searches likely to take place with any choice of medium in any given library can be determined only by trial. Furthermore, the level of activity will vary from library to library even for institutions of the same size. Also, prices for each product will vary from library to library in relation to population size and other factors. Databases are priced accordingly, but like income taxes, vendor pricing structures are often bracketed so that any two libraries with nearly the same size user population may or may not realize the same matrix. Therefore, each library will have to develop its own model for choosing among index alternatives. In Trinity's case, the model matrix includes data on approximately one hundred indexes, most of which will continue to be purchased in print form. What follows is a rough template based on experience at two libraries and modified by experience reported in the literature.

It is unlikely that many libraries have accurate knowledge of the use of manual indexes. Therefore, it will be difficult to determine the average cost per search when using them. On the other hand, most of the costs for mounting files locally or for CD-ROM, end user, and mediated searches can be calculated. Taking each type of search in turn, the following cost elements need to be considered: purchase of data, storage of data, hardware utilization, software utilization and maintenance, and staff time. The nondollar cost elements that should be considered but cannot be calculated are user effort (perhaps better termed *agony*), in the form of multiplicity of interfaces and learning requirements; friendliness of the interface; power of the retrieval software; response time; availability of campus-wide access; and linkage to library holdings. The spectrum across alternatives that are available runs from printed indexes through CD-ROMs, end user searching, and mediated search services to locally mounted, online files.

When making the choice from this spectrum, one needs to recognize both the direct costs and the nondollar costs. A useful guide derives from comparing the pros and cons of manual searching versus mediated online searching. These may be equivalent in total costs to the user if all users are taken into consideration, which suggests that the number of manual searches is approximately equal to the number of mediated searches. However, experience suggests that manual searches exceed the number of mediated searches by five or ten times for undergraduate students. Also, it should be assumed that there is some price per search across all disciplines that is acceptable for the library to subsidize. This price-per-search figure is implied by the availability of the printed index in the collection and could be calculated if the number of searches per year were known. Given the lack of information on how many times indexes are searched, demand should be somehow estimated for printed indexes. Alternatively, the library can set an arbitrary figure or calculate one based on experience. The previously cited research at Clemson suggests that \$3 to \$5 might be acceptable.

For Trinity, the estimates were based on

Table 1. Average Cost per Search for CD-ROM Indexes at Trinity

Database	Years Covered	Avg. Annual Subscription Cost (in Dollars)	Annual Hardware Cost (in Dollars)	Avg. Annual No. of Searches (italic=estimate)	Cost per Search (in Dollars)
ABI/Inform	1988–92	4,950	253.33	705.5	7.38
Compendex	1970–92	4,920	253.33	128.0	40.42
MLA Bibl	1981–92	1,187	253.33	289.0	4.98
PAIS	1972–92	1,500	253.33	293.5	5.97
PsychLit	1974–92	3,395	253.33	608.0	6.00
ERIC	1988–92	630	253.33	*219.4*	4.03
InfoTrac	1969–92	9,800	253.33	*5,000.0*	1.96
Total		26,382	1,520.00	7,243.0	

several assumptions. In regard to locally mounted files, it was assumed initially that the loading process, storage, and maintenance would cost approximately the same as these costs at Clemson for *ERIC* and IAC's *Magazine Index*. While there could be variation depending on the allocation of computer usage, this variable was excluded from the analysis, since there was no acceptable algorithm to allocate the computer available.[5] The cost of staff time and costs of storage are included and would be very similar across campuses. For example, to support this argument, the maintenance of NOTIS at Clemson requires approximately 2.5 FTE and at Trinity 2.0 FTE. The difference between these two relates to the level of demand made on these systems at the two institutions; Clemson has several special uses of NOTIS that are not duplicated at Trinity. For both campuses, file storage costs are equivalent. A second estimate, shown in table 1, was based on a low-cost stand-alone implementation.

It was also assumed that the number of searches at the two campuses are proportional to the number of people in each user population. This introduces a quality vector that could skew the results. However, since there are nondollar, subjective elements that should be considered in deciding among alternatives anyway, this skewing was not considered. Therefore, to estimate the cost per search of locally mounted files at Trinity, a direct extrapolation was made from Clemson data initially. Since the user population was 3,000 at Trinity and 17,000 at Clemson, the ratio 1 to 5.667 was used for the general IAC index and 1 to 20 for *ERIC* because of a smaller education program. Alternatively, these estimates could be approximated by dividing the total estimated costs of staff time plus storage plus database acquisition plus allocated software costs by the expected number of online searches made of the local files. In the absence of locally loaded files at Trinity, that calculation was estimated for *ERIC* and for the databases where current subscriptions to CD-ROM indexes are taken.

The costs per search for other alternatives were easier to make. Trinity had a base of experience with several CD-ROMs, which are shown in table 1. *ERIC* is included, because of its importance, although Trinity does not subscribe to the CD-ROM version at this time. Costs of hardware and the subscription costs of the product were added and divided by the number of uses. Trinity dedicates two personal computers with CD-ROM drives for indexes. The cost of these machines was amortized over five years and divided equally among the indexes. Use of these has been closely monitored by checking out the actual CD-ROMs to the users each time they use one. There is some element of error introduced by the necessity to log uses, but consistency in the logs over two years suggests this is minimal.

For end user searching, Trinity assumes that OCLC FirstSearch will provide the single alternative on campus. Estimates reported by others were used, since accounts were set up too late in the past academic year to generate useful data for this project.[6] Reported estimates indicate that an average of six transactions is required per search and that these transactions will cost between 45 and 90 cents each on FirstSearch, depending on the quantity ordered. Finally, the costs of mediated searching at Trinity were extracted from historical invoices. The matrix of direct dollar cost for those indexes that Trinity makes available on CD-ROM along with *ERIC* and IAC's *Magazine Index* for each alternative were easy to make. However, some accommodation for nondollar cost should be factored into the decision making effort.

Trinity has a policy of not charging back to the user, up to a limited amount, for online searching on OCLC FirstSearch or through mediated searching. Given this policy, it was relatively easy to illustrate graphically the trade-off in total costs (actual dollars spent plus nondollar costs) for each type of alternative. Figure 1 takes into account the cost elements noted above in addition to some subjective measure of user agony, the availability of linkage to library holdings, and staff interaction time. No apology is made here for the subjective nature of this graph. It is meant only to provide some visualization of the elements that must be considered in weighing the alternatives. Clearly, from this representation, users have the most to gain from locally mounted files and the least to gain from printed indexes. The reverse is probably true from the point of view of the library, absent consideration for the impact of public relations.

With this graphic as background, an examination of the matrix for Trinity makes choosing among alternatives fairly clear at this time. As the availability of reasonably priced search engines in conjunction with low-cost hardware becomes more available, the matrix will shift the results in favor of locally mounted files. In addition, the availability of more resources on the Internet will elevate demand on campus for downloaded, full-text, electronic resources. Hardware and staff to support this demand used simultaneously with locally mounted indexes will tend to lower the cost of the online, locally mounted index alternative.

Calculations of the costs for locally mounting files, shown in table 2, are based on a five-year amortization in most cases. Many vendors price their tapes in two parts: a one-time back-file purchase followed by annual update fees. In those cases, the cost of data was set at the annual fee plus one-fifth of the

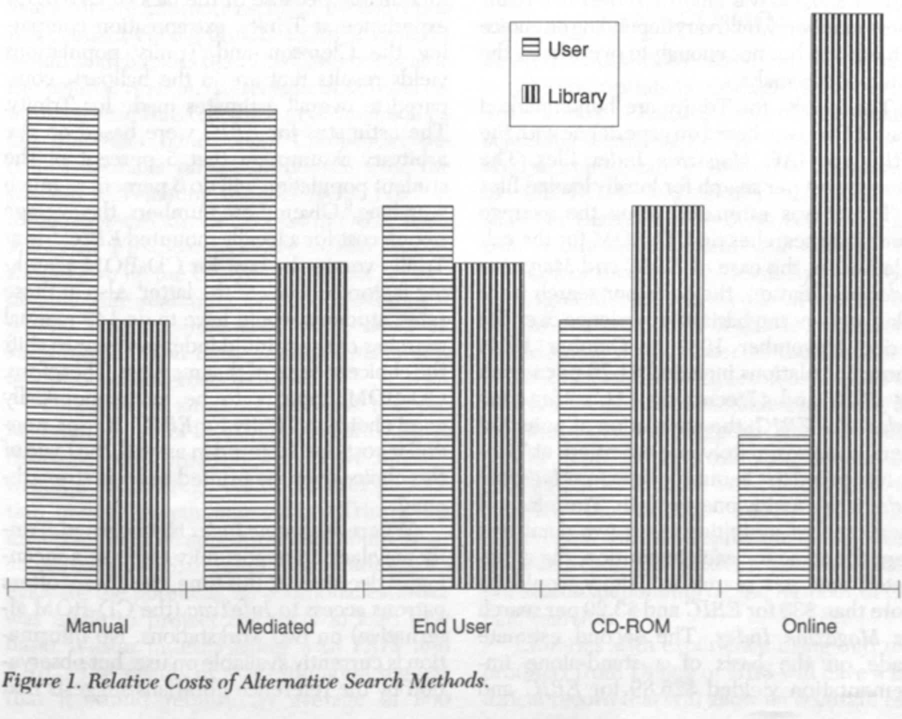


Figure 1. Relative Costs of Alternative Search Methods.

Table 2. Average Cost per Locally Mounted Index at Trinity

Database	Years Stored Online	File Size/ (MB)	Annual Cost of File ($)	Annual Hardware Cost ($)	Annual Load and Maint.($)	Annual Total Cost ($)
ABI/Inform	1988–92	381	10,100	600	3,500	14,200
Compendex	1988–92	1,632	14,369	600	3,500	18,469
PAIS	1972–92	210	10,600	600	3,500	14,700
PsychLit	1987–92	118	12,500	600	3,500	16,600
ERIC	1969–92	571	1,800	600	3,500	5,900
IAC *Magazine Index*	1988–92	319	6,750	600	3,500	10,850
Total		3,231	56,119	3,600	21,000	80,719

cost of the back file. This amortizes the file cost over five years, which is a reasonable means to even out the comparison. Where prices were given as a price per year, the estimate was made for the period 1987–91 or 1988–92, depending on the vendor's information. Additionally, the estimates were based on an installation using BRS Onsite running in the C version on a Unix platform with a limit of ten simultaneous users. For this environment, a one-time expenditure for hardware and software of $36,500 was amortized over five years. These numbers would vary depending on choice of hardware but not enough to overwhelm the results of the analysis.

The results for Trinity are benchmarked against Clemson based on experience with the *ERIC* and IAC *Magazine Index* files. The average cost per search for locally loaded files at Trinity was estimated using the average number of searches on CD-ROM for the calculation. In the case of *ERIC* and *Magazine Index* at Clemson, the costs per search were calculated on the basis of experience over the period November 1989 to October 1990. Those calculations included $1.79 per search for *ERIC* and 47 cents for IAC's *Magazine Index*. For *ERIC*, the population of potential users is approximately one-twentieth at Trinity compared to Clemson, while for *Magazine Index* it is roughly one-seventh. Therefore, if fixed costs of mounting these two databases were the same for each institution, the anticipated cost per search at Trinity would be more than $35 for *ERIC* and $3.29 per search for *Magazine Index*. The second estimate made on the basis of a stand-alone implementation yielded $26.89 for *ERIC* and $2.17 for *Magazine Index*. Table 3 shows cost-per-search estimates across all alternatives for those indexes that Trinity currently takes on CD-ROM, in addition to *ERIC* and IAC's *Magazine Index*.[7]

For *ERIC*, the cost per search for the locally mounted file appears to be greater than the marginal benefit at Trinity and less at Clemson. Similarly, local mounting of *Magazine Index* is viable at Clemson and may be viable at Trinity.[8] While figures for *ERIC* were unavailable because of the lack of CD-ROM experience at Trinity, extrapolation comparing the Clemson and Trinity populations yields results that are in the ballpark, compared to overall estimates made for Trinity. The estimates for *ERIC* were based on the arbitrary assumption that 5 percent of the student population will do 5 percent as much searching.[9] Given these numbers, the average search cost for a locally mounted *ERIC* file at Trinity versus the cost for CD-ROM searching factors in favor of the latter. Also at these rates, students would have to do 149 manual searches of the printed index per year to shift the choice in favor of that medium. Therefore, CD-ROM appears to be an economically good choice at Trinity for *ERIC*. If user, nondollar costs are factored in as well, the logic of this choice over the printed index is strengthened.

Where *Magazine Index* is concerned, Trinity may lack an opportunity to make a meaningful decision. At this time, the library offers patrons access to *InfoTrac* (the CD-ROM alternative) on two workstations. No information is currently available on use, but observation by the reference librarians suggests that

Table 3. Average Cost (in Dollars) per Search for Each Alternative at Trinity

Index	Print	Mediated Online	CD-ROM	FirstSearch	Locally Mounted Index
ABI/Inform		18.72	7.38		20.13
Compendex	3,550/#s	47.54	40.42	4.80	110.16
MLA Bibliography	892/#s	8.72	4.98		
PAIS	495/#s	11.35	5.97	4.80	5.01
PsychLit	1,245/#s	8.74	6.00		27.30
ERIC	550/#s	7.64	4.03	4.80	26.89
IAC *Magazine Index*		13.76	1.96		2.17
Total cost	6,732	1,214	26,382		80,719

it is in the range of five thousand searches per year. At the current price to Trinity, the cost is roughly estimated to be $1.96 per search. Since this is close to the $2.17 estimate noted in table 3, it will be prudent for Trinity to explore the possibility of locally mounting the IAC file. This suggests that loading some Wilson indexes also may merit consideration.

Similarly, cost calculations of the optional choices for these and other databases reveal a pattern that indicates CD-ROM is a prudent choice for *ABI/Inform* and *PsychLit* given the availability of alternatives. *MLA Bibliography* is a prudent choice on CD-ROM if students would search it less than 150 times per year in the printed version. (Reference desk observations substantiate the wisdom of the choice.) On the other hand, when *Compendex* becomes available on OCLC FirstSearch, the wisdom of retaining the CD-ROM product diminishes quickly. In the case of *PAIS*, availability of FirstSearch indicates that Trinity should give up the CD-ROM product in favor of end user searching or local mounting if several files are mounted. Again, however, keep in mind that these comparisons largely ignore nondollar costs to users.

This analysis supports the view that the best candidates for local mounting are indexes that have very broad appeal, active interest within a unique segment of the user population, or very low purchase cost. At Trinity, this means it might be appropriate to mount *Magazine Index*, *PAIS*, or some Wilson indexes. To evaluate the potential, an additional estimate was made to project the costs to load four basic Wilson indexes along with *PAIS* and *ERIC*. The results of that estimation suggest that it would require an average of 900 searches per year on each Wilson database to bring the costs within the range of FirstSearch and 2,400 searches of *PAIS* or 1,400 searches of *ERIC* to bring per-search cost within the range of the CD-ROM or FirstSearch alternatives. Trinity would have to mount the files to test the validity of this estimate. Given our current rate of roughly 5,500 searches per year of *ERIC*, *PAIS*, and *InfoTrac* on CD-ROM, it is likely that a local mounting merits consideration.

CONCLUSION

Several assumptions based on this analysis may be appropriate to establishing a model for any library choosing among alternative searching technologies. The amount of mediated searching can be related to the number of CD-ROM and locally mounted file searches that would be done each year. Locally mounted databases will multiply the number of mediated searches by two orders of magnitude, that is, by about one to two hundred. The number of CD-ROM searches would come much closer to the number of searches of locally mounted files than to the number of manual or mediated searches, but Trinity's experience suggests that CD-ROMs will be searched approximately fifty to one hundred times as often as mediated searches. Further, the average search on OCLC FirstSearch will take approximately six executions and thus cost between $2.70 and $5.40 per search, depending on the number of prepaid searches.

Libraries with experience using end user products from Dialog or BRS will have a historical record that will allow an accurate estimate for this alternative. For the library with records on the number of mediated searches or CD-ROM activity, estimates can be plugged in to the matrix for each database available, using these assumptions. Comparison and choice also should include some reflection of the nondollar search effort, at least subjectively. When dollar costs are nearly equal across alternatives, user friendliness of locally mounted files or CD-ROMs will factor in favor of those over more cumbersome mediated and end user alternatives.

This model lacks precision. However, in comparison to the data-poor environment in which collection development librarians have traditionally chosen printed indexes, some progress has been made. For example, Trinity will purchase *Chemical Abstracts* at a cost of $8,600 for the 1993 calendar year. There are no available data on campus to estimate accurately how many searches will be made of this index during a given year. However, the thirty-six students, six postdoctoral researchers, and seven faculty will have to make thirty-two searches per year each to bring the costs into the $5 per-search range. Given this, it seems likely that the printed index is not as cost-competitive with free online access for all users as it once was, particularly if regard is given to nondollar advantages of searching this index online. Using this model, a hypothetical implementation of four Wilson indexes with *PAIS* and *ERIC* locally mounted on a small Unix platform yields promising possibilities. Clearly, the days have arrived in which electronic alternatives compete effectively with traditional indexes.

ACKNOWLEDGMENTS

The author acknowledges with gratitude the information support of Ches Martin of Clemson University and the vendors. Thanks also to associates Chris Nolan for valuable insights on the text and Craig Likness for the matrix of data and price information at Trinity. Any errors should be exclusively credited to the author.

This paper was originally written for presentation at the Library and Information Technology Association national conference in Denver, September 13–17, 1992.

REFERENCES AND NOTES

1. Richard W. Meyer, "Management, Cost and Behavioral Issues with Locally Mounted Databases," *Information Technology and Libraries* 9, no.3:226–41 (Sept. 1990).
2. Richard W. Meyer, "Locally Mounted Databases . . . Making Information as Close to Free as Possible," *Online* 16, no.1:15, 17–24 (Jan. 1992).
3. This is standard theory of a public good nicely summarized by David Tyckoson, "Electronic Databases, Information Pricing, and the Role of the Library as a Collective Information Utility," *Technicalities* 9, no.6:11–13 (June 1989).
4. Some of the increased activity will be brief searches for individual items or casual interest, which comes about simply because searching is easier.
5. The allocation of machine appears to be left out in other studies reported in the literature, unless the computer is dedicated to the OPAC or index service.
6. Usha Gupta and Lutishoor Salisbury, "Is FirstSearch Really Attractive?" *College and Research Library News* 53, no.7:461–62, 464 (July/Aug. 1992). Data from Clemson show an average of six transactions per search, where *search* is defined as a connect and *transaction* is defined as every keystroke that invokes the search engine into action.
7. Trinity subscribes to IAC's *InfoTrac* product, which is compared to *Magazine Index*. Estimates for the cost per search were not calculated for printed indexes because of a lack of data on the amount of searching.
8. Since the cost of hardware and software is allocated across the mounted databases, full attribution of these costs to one database would greatly increase the cost per search. Bringing locally mounted files into the realm of possibility requires large reductions in fixed costs or more databases against which to allocate them.
9. A comparison of the number of books circulated per student at the two schools suggests this is not a valid number and that the Trinity students will do as much as 10 percent of Clemson's level of searching, but the results of this study are unaffected. ■■

Optical Storage and Retrieval of Library Material

Doris R. Folen and Laurie E. Stackpole

The Ruth H. Hooker Research Library and Technical Information Center of the Naval Research Laboratory has installed an optical disk system consisting of a Sony autochanger, Sun minicomputer, Sun workstations, TDC scanners, printers, personal computers, and various other peripherals. The system stores large portions of the library's collection on twelve-inch optical disks and can be expanded to allow retrieval over the campus network by the scientists of the Naval Research Laboratory. The first segment of the collection to be processed is a technical report collection consisting of 140,000 reports averaging fifty-five pages each. A third of this collection has currently been scanned to disk and is available for retrieval and on-demand printing by the library patron.

This paper describes the development of an optical disk storage capability at the Ruth H. Hooker Research Library and Technical Information Center of the Naval Research Laboratory (NRL). Optical storage was selected for the preservation and maximum protection of the library's immensely valuable collection of technical reports. This collection represents results of research in the areas of physics and engineering since the beginning of the Second World War, and much of this information is unavailable from any other source.

Optical disk technology also provides a viable, sensible solution to many of the serious recurrent problems that plague librarians. The most striking advantage of optical storage is that it is a permanent solution to the space problem; no matter how you treat them, index them, or catalog them, paper products, film products, and so on take up space, and space costs money. Omitting the costs associated with filing, retrieving, and refiling reports, the savings to the NRL in overhead will be more than $100,000 a year when the entire unclassified report collection has been put on disk. Space requirements will drop from 3,600 square feet to only 144 square feet. Optical storage also saves money by eliminating all future filing, retrieving, and refiling associated with physical storage. A paper copy of an item can be printed at the touch of a key and never has to be refiled, since the archived copy remains on the disk. This technology presents a dream scenario in which a librarian can sit in a pleasant environment, identify a paper, report, or picture, and retrieve the item in seconds without leaving the area or the patron.

While the costs of equipment and conversion mandate careful consideration and planning for the implementation of optical storage and retrieval, these costs are not so great that smaller libraries are prohibited from taking advantage of this technology. Small turn-key units can be purchased at reasonable costs, and if libraries collaborate in the scanning, costs can be absorbed in the budgets of even the smaller establishments.

Optical disk technology exists today and is being applied both in the government and commercial sectors. For example, the Internal Revenue Service, which must deal with the phenomenal problem of storing income tax returns both from the private and indus-

Doris R. Folen is a Senior Systems Analyst with Kestrel Associates Incorporated, Arlington, Virginia. Laurie E. Stackpole is Chief Librarian, Ruth H. Hooker Research Library and Technical Information Center, Naval Research Laboratory, Washington, D.C.

trial sectors, has several prototype optical disk storage and retrieval systems currently operational and several others in the planning stage.[1]

The giant insurance company United Services Automobile Association in San Antonio, Texas, has produced a practically paperless operation using optical disk technology; smaller insurance agencies are opting for this technology as well.[2]

The U.S. Navy's "paperless ship project" launched by Admiral Metcalf in 1987 is an undertaking that includes converting vast numbers of documents to optical storage.[3]

The list of those who are applying this technology is a long one, and users and potential users may be found wherever there are large quantities of information to be stored.

Libraries are obvious beneficiaries, since by their very nature they are places where large quantities of information are stored in every kind of format: books, journals, maps, charts, film, and computer databases. All are candidates for optical storage and retrieval.

The National Archives and the Library of Congress each have been working with prototypes for optical storage for more than five years and have proven the importance of this technology for preserving their immensely valuable, but sometimes old and brittle documents.[4,5] Libraries such as those at Cornell University and Carnegie Mellon University have designed optical storage and retrieval systems and currently are engaged in projects to exploit the capabilities of these systems.[6,7] The NRL library has had optical storage and retrieval capabilities fully operational for more than three years. This optical disk system is one of the first of its kind and is still the largest, in terms of volume of material stored, currently operational in a research library.

PLANNING FOR AN OPTICAL DISK SYSTEM

Planning for an optical disk system is no easier now than it was four years ago, although the problems are different. The major concern then was collecting information and advice, since there were comparatively few people who were authorities in this field and very few facilities that had any experience with optical storage, as the technology was so new at that time. Present-day problems are largely due to the proliferation of new equipment and the burgeoning technology.

Optical disk systems are expensive. Right from the beginning they require a commitment of time and money. The first year cost of the NRL library's system was just over $500,000. This included development of a prototype system, installation, maintenance, and the scanning to optical disk of fifteen thousand reports. Additionally, upgrades to improve scanning, printing, and optical character recognition have added another $500,000. Many uncounted hours of staff time went into the original planning as well.

An optical disk system is a project that, once started, cannot easily be abandoned. It will be an integral part of the library for a long time, and future plans have to be considered. Thought must be given to who is going to use the system, who is going to run the system, and who is going to be responsible for its maintenance and future development. The level of staffing must be considered. Professional librarians and support technicians must be assigned to assist patrons and to maintain the system.

Funds for staffing must be budgeted depending on the size and planned use of the system. Large in-house systems require a supervisory librarian with knowledge of large computer systems and a staff that includes, at a minimum, reference librarians trained to use the system, a computer systems analyst or other system administrator, and library technicians or clerks to assist with the report conversion. A small library with minimum turnkey equipment may be able to absorb the entire operation with its existing staff.

Consideration should be given to the time requirements of an optical system. It will take time to plan the system, buy it, install it, run it, maintain it, and scan documents into it. Staff and patrons must be trained to use it, which will also take time.

A study should be made of why an optical disk system is needed: What problems will it solve, how will it be used, how will it change the way staff and patrons do their work? Decisions have to be made on design considerations and trade-offs. For example, is retrieval speed important enough to justify increased system costs, or what is the best compromise between image quality and time and money available for the scanning process?

With currently available optical equipment, images from almost any medium can be processed and sent to the optical disk. Equipment exists for automatically scanning film, microfiche, maps, charts, photographs, and aperture cards. Therefore, decisions have to be made as to what is going to be converted to optical images. For example, the NRL library has one million reports in microfiche format, and they take up quite a bit of space and are a bother to file. However, they are much more space efficient than paper products, so the decision was made to process the paper products.

Vendors must also be considered and selected. Once one or more vendors can be singled out as likely candidates, discussions with them will identify the equipment most suitable for incorporation into an appropriate system. Usually at least three or four configurations slowly emerge for consideration.

Money must be budgeted. This is a chicken-and-egg proposition. The amount of money available will determine the design. On the other hand, a study of what a system could accomplish for an organization might justify a higher budget. At this time, at least $250,000 is needed for the purchase of equipment to begin a large operation; this includes an autochanger for 12-inch disks, minicomputer, scanner, printer, supporting personal computers, and workstations. On the other end of the scale a small operation might cost only $50,000, which could provide a personal-computer operation involving a single disk drive and 5.25-inch disks. System sizes can be looked at in terms of numbers of pages to be stored. A 12-inch optical disk can hold up to 130,000 pages, while a 5.25-inch disk can store approximately 20,000 pages. If a collection consists of millions of pages to be stored, then a large system must be considered.

Ongoing costs in both large and smaller systems involve library staff to help patrons, purchase of equipment upgrades, maintenance of equipment and software, and the cost of continuing document conversion.

A good estimate of the money to be budgeted can only be arrived at after all aspects of an optical storage have been considered and in-house needs established. Costs of hardware and software are not something that a librarian/planner can determine alone, as most equipment and/or systems are sold only through vendors. Extensive talks with vendors who are experts in the field of costs are needed to begin to solidify an estimate.

The future of the system must be considered. Evidence indicates that images can be safely stored on optical disks for up to a hundred years.[8] However, the technology is advancing very quickly. For example, images stored on a disk holding 3.2 gigabytes take up twice as many disks as images stored on recently available disks holding 6.55 gigabytes. Already the NRL library has upgraded its system to accommodate the higher density disks. Multimedia storage, that is the recording of sound, video, and graphics all on the same disk, is a reality and soon will be practical. Scanners and printers that handle color are available. Image scanners and printers that process sheets of paper at more than one hundred pages per minute are to be available at reasonable cost in the near future. Although not all of the technology that vendors and manufacturers promise for the future materializes, much of it does reach the market: provisions for upgrading an optical disk system to incorporate these advances should be part of the overall game plan.

THE LIBRARY'S OPTICAL DISK SYSTEM

The NRL library initially chose Online Computer Systems, Inc., as the OEM (original equipment manufacturer). Online Computer Systems took the library's choice of basic equipment and, working with the library staff on the design, put together one of the first optical disk systems of major proportions to be used in a research library (see figure 1). Currently, Kestrel Associates, Inc., is providing technical staff with responsibility for the maintenance of equipment, report preparation and scanning, and the integration of new hardware and software into the system.

When the library began its optical storage and retrieval project, there were very few vendors in this field. Now there are many well-qualified companies who can provide systems tailored to any library's budget or need. One source of a listing of available vendors is *The Handbook of Optical Memory Systems*, by C. Peter Waegemann, which has a chapter on vendors and consultants and is revised periodically.[9]

Proper selection of equipment is of immense importance. Just the cost alone necessitates a carefully considered decision. Once purchased, the major components become the platform for future developments. Re-

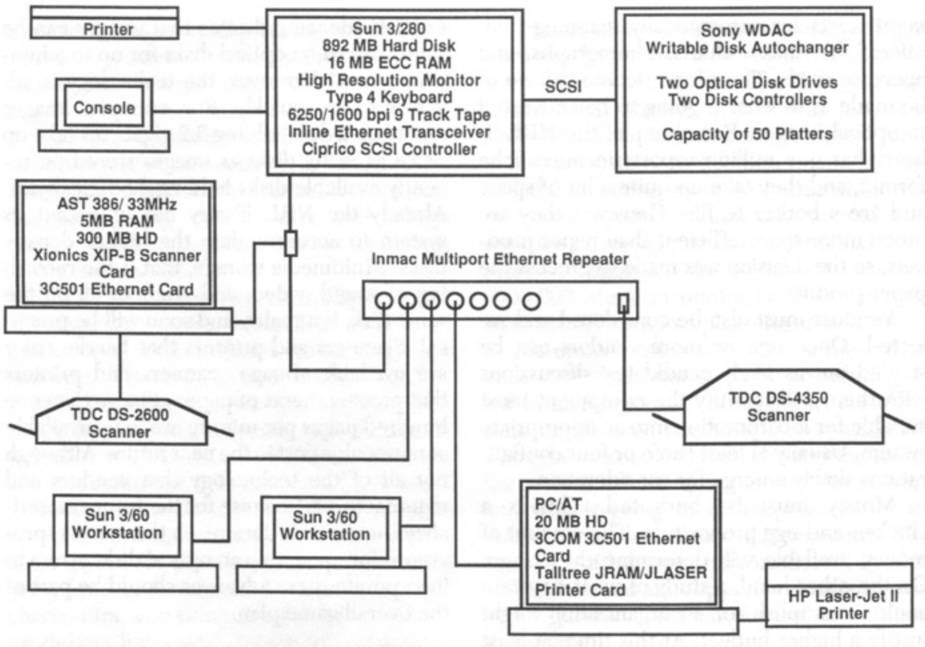


Figure 1. NRL Document Retrieval System.

placement of major components is not easily made.

Many hours were spent at conventions and shows looking at equipment and talking to vendors before the equipment forming the nucleus of the NRL's optical disk system was selected. Dreamware and fantasyware had to be identified as such.

Since the system was to be an in-house, stand-alone archive, standards were not a primary consideration in the selection of equipment. Standards for the twelve-inch optical disk were in the talking stage at that time, as they are still. The library worked around this problem by trying to project standards and by picking leaders in the field as vendors.

A Sony Writable Optical Disk Autochanger, model WDA-610 (see figure 2), and a Sun minicomputer and its workstations form the nucleus of the system. To be consistent with the market, Sony now calls its autochanger a jukebox, which it somewhat resembles in operation. It is designed specifically for use with Sony's twelve-inch optical disks. The jukebox has a footprint of eight square feet and from the front looks remarkably like a two door refrigerator.

One twelve-inch WORM (write once, read many) optical disk can provide 6.55 gigabytes of digital data, enough to store the contents of 130,000 typewritten pages. With the ability to accommodate fifty disks of 6.55 gigabytes, one autochanger provides the equivalent storage space of up to five hundred file cabinets. Up to four autochangers can be daisy-chained to expand storage capacity to 1.312 terabytes of online data on a single SCSI interface. Through the synchronized use of two writable disk drives, the autochanger offers an average disk-to-disk access time of just five seconds.

Connected to the Sony autochanger by a SCSI interface is a Sun 3/280. It has a console, four workstations, a 892-megabyte hard disk, 16-megabyte ECC RAM, online Ethernet transceiver, and a Ciprico SCSI controller. It also has a nine-track tape drive.

One of the in-house scanners is a TDC DocuScan DS-2600. It is capable of scanning pages of different sizes and thicknesses and can scan two sides of an 8 1/2-by-11-inch page in fewer than two seconds at 200 dpi (dots per inch). It is of compact design, having a footprint of 2 feet by 2 feet, and weighs 100 pounds. This scanner is supported by an AST 386/33 microcomputer containing an Xionics XIP-B scanner card. Here the scanned images

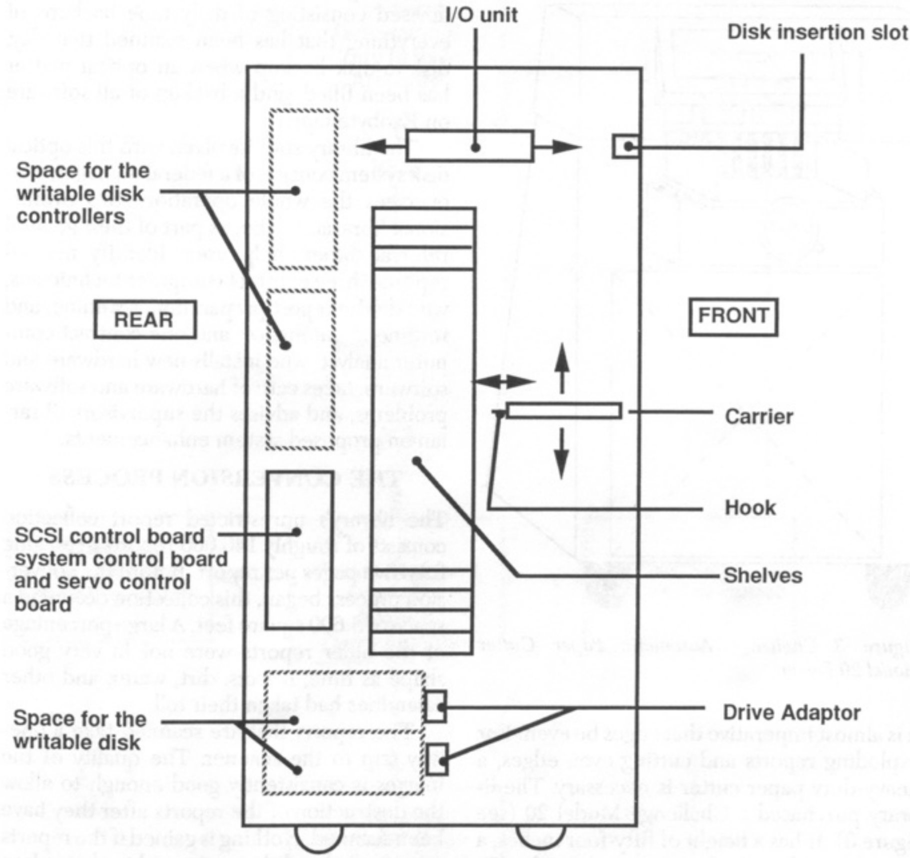


Figure 2. Sony WDA-610 Autochanger.

are compressed and stored on a 300-megabyte hard disk before being transferred to the Sun 3/280, where the images are processed and sent to the Sony autochanger.

In using the TDC DocuScan DS-2600 scanner, the operator slides the sheet to be scanned along the alignment guide. The transport system grips the sheet and carries it on a straight path past the scanning area and into the receiver tray. Thus paper jams are virtually eliminated. Scanning up to two thousand pages in an eight-hour workday is considered possible with this equipment.

A TDC DocuScan DS-4530 scanner has been incorporated into the system as the primary scanner. This scanner scans at 300 dpi, has sheet feeder and monitor, and is capable of sustained scanning at a rate of forty pages a minute. After enhancing the image electronically, this scanner compresses the image into industry-standard CCITT Group IV format. When this scanner is used in conjunction with the TDC DS-2600, the capability of scanning sixteen thousand pages in an eight-hour workday becomes a reality.

When a stored image is to be viewed, it is retrieved by the Sun 3/60, decompressed, and either viewed at a workstation or queued on a PC and printed out on an HP Laserjet printer. The equipment is networked internally on an Ethernet LAN and could be connected to the campuswide network, which is part of the Internet.

All bindings, staples, and so on that hold a report together must be removed before scanning. Documents can be scanned more readily if edges are fairly even, and if an automatic document feeder is used on the scanner

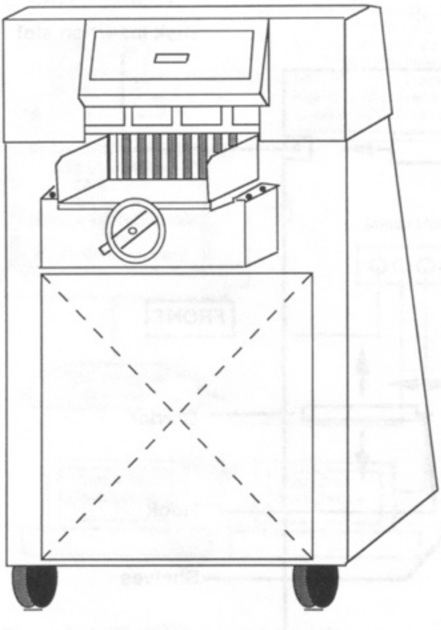


Figure 3. Challenge Automatic Paper Cutter Model 20 Power.

it is almost imperative that edges be even. For exploding reports and cutting even edges, a heavy-duty paper cutter is necessary. The library purchased a Challenge Model 20 (see figure 3). It has a height of fifty-four inches, a footprint of three by four feet, and weighs 530 pounds. It requires two hands on the control panel for operation to keep the operator out of harm's way. A Ferrups 18KVA uninterruptible power supply was installed to provide power in the event of a general blackout. This system provides fourteen minutes of full-load power or thirty-nine minutes of half-load, which gives the entire computer operation time to take itself down in an emergency without any loss of data.

The equipment described above is all housed in the Documents Section of the NRL library. The smaller units, such as the workstations, printers, and paper cutter, are easily moved and are rearranged from time to time. The larger units, such as the minicomputer, autochanger, and emergency uninterruptible power supply, are much more unwieldy and can be moved only with great care and difficulty. Environmental factors such as heat and cold are not a problem.

A comprehensive backup routine has been devised consisting of daily tape backups of everything that has been scanned that day, disk-to-disk backup when an optical platter has been filled, and a backup of all software on Exobyte tape.

The library staff involved with this optical disk system consists of a federal manager, who oversees the whole operation; two professional librarians, who, as part of their general referral duties, help users identify needed reports; three contract computer technicians, who do the report preparation, scanning, and routine maintenance; and one contract computer analyst, who installs new hardware and software, takes care of hardware and software problems, and advises the supervisory librarian on proposed system enhancements.

THE CONVERSION PROCESS

The library's unrestricted report collection consists of roughly 140,000 reports averaging fifty-five pages per report. Before the conversion process began, this collection occupied a space of 3,600 square feet. A large percentage of the older reports were not in very good shape as time, insects, dirt, water, and other calamities had taken their toll.

The reports that are scanned take a one-way trip to the scanner. The quality of the images is consistently good enough to allow the destruction of the reports after they have been scanned. Nothing is gained if the reports are scanned and then returned to their place on the shelf to continue to take up space. There are, of course, some exceptions. Occasionally there is a report that has actual historical value, and the original is considered worth preserving. Those few reports are scanned and then saved in a historical file.

Because this entire collection is to be scanned, no particular care has been taken with scanning order. Reports that have come back from circulation are put aside to scan because there is no sense filing them and then taking them off the shelf later to scan. New NRL-originated reports are scanned as soon as the library receives a copy. Reports are taken from the shelves in sequence but are not checked for misfiling as they are prepared for scanning, because unlike full-sized paper reports, which are filed on a shelf, images do not have to be placed on the disk in any particular order.

Time spent in preparation is well invested.

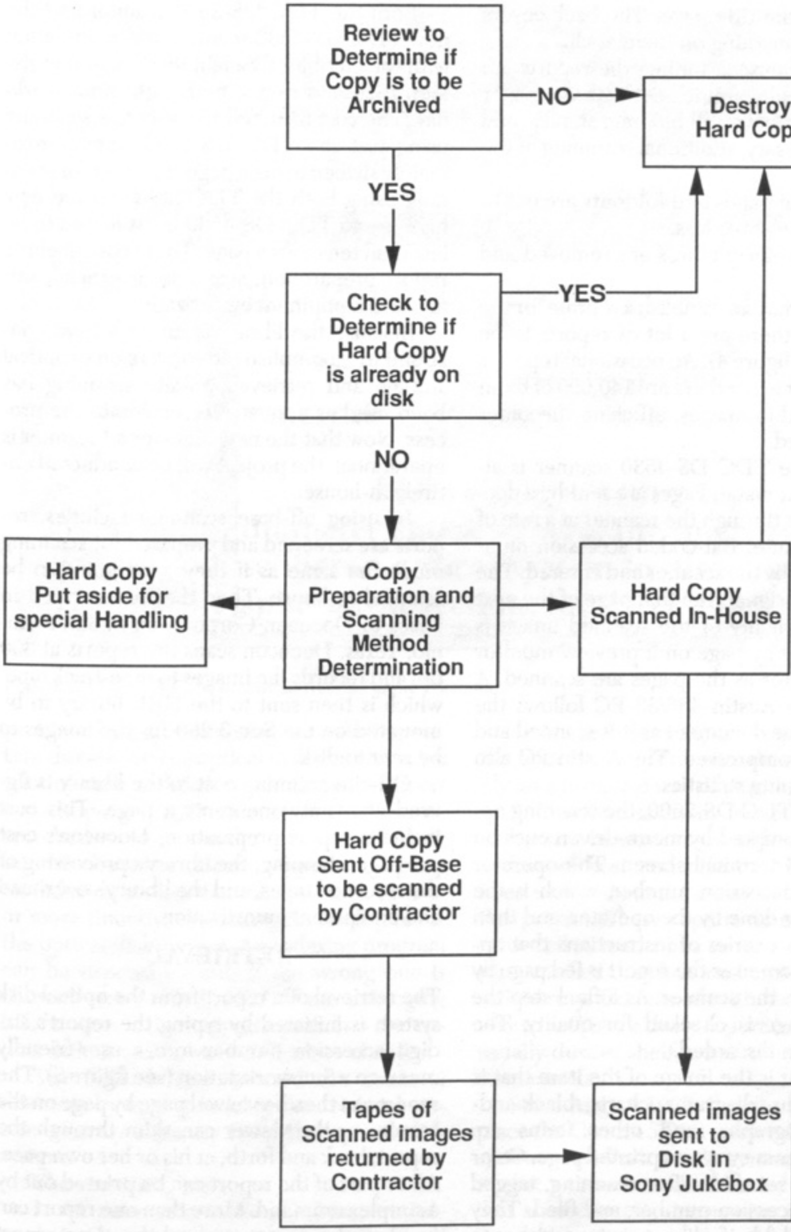


Figure 4. Workflow for Processing Hard Copy for Optical Disk.

Each report is reviewed as a candidate for scanning. Duplicates and reports without technical substance are weeded out. Reports that are oversized or of poor quality are put aside for later consideration.

A report is prepared for the scanner by the following steps:

1. All blank and extraneous pages are removed, especially covers when possible. Usually all of the information on the cover is

repeated on the title page. The back covers seldom have anything on them at all.

2. For in-house scanning, the reports are then exploded by a quick cut with the paper cutter. This removes all binding, staples, and so on. If necessary, additional trimming of the sides is done.

3. Oversize pages and foldouts are cut to 8 1/2 by 11 inches or less.

4. Loose color pictures are removed and filed.

It is important to establish a routine for the work flow if there are a lot of reports to be scanned (see figure 4). An occasional report is one thing, but when there are 140,000 of them to be reduced to images, efficiency becomes the watchword.

Use of the TDC DS-4530 scanner is almost fully automatic. Pages are sent by a document feeder through the scanner at a rate of forty per minute. Bar-coded accession numbers are read by the scanner and entered. The bar code also signals the first page of the next document. Quality of the scanned image is checked page by page on a preview monitor by the operator as the pages are scanned. A screen on an Austin 486/33 PC follows the progress of the document as it is scanned and the images compressed. The Austin PC also records scanning statistics.

With the TDC DS-2600, the scanning operation is prompted by menu-driven cues on the AST 386 terminal screen. The operator keys in the accession number, which is the only indexing done by the operator, and then follows a short series of instructions that appear on the screen as the report is fed page by page through the scanner. As a final step the scanned image is checked for quality. The report is then discarded.

Because it is the image of the item that is stored, graphs, diagrams, charts, black-and-white photographs, and other items are scanned the same way as a printed page. Color pictures are set aside after scanning, tagged with their accession number, and filed. They are then available if a library patron wishes to see them. Fortunately, because of the nature of the reports being scanned, there is not a large percentage of reports with color pictures. The library plans to scan these color pictures for retrieval with the rest of the original report when future technology produces viable and inexpensive equipment for scanning, retrieving, and printing color images.

Both the TDC DS-2600 scanner and the new TDC DS-4530 scanner are in operation with the combined capability of scanning sixteen thousand pages in an eight-hour workday. The cost incurred for in-house scanning using just the TDC DS-2600 was approximately sixteen cents a page. The cost for scanning using both the TDC DS-2600 the new high-speed TDC DS-4530 is estimated to be less than ten cents a page. These costs include report preparation, scanning, overhead, salary, and equipment amortization.

Because the library is under a time constraint to accomplish the conversion to optical storage and retrieval, off-site scanning has been used as a means to accelerate the process. Now that the new high-speed scanner is operational, the project will be conducted entirely in-house.

In using off-base scanning facilities, reports are screened and prepared for scanning much the same as if they were going to be scanned in-house. Then they are shipped in boxes to Docucon Corporation in San Antonio, Texas. Docucon scans the reports at 300 dpi and records the images to nine-track tape, which is then sent to the NRL library to be mounted on the Sun 3/280 for the images to be sent to disk.

Off-site scanning cost to the library is figured at twenty-one cents a page. This cost includes report preparation, Docucon's cost per page, shipping, the library's processing of the returned tapes, and the library's overhead and equipment amortization.

RETRIEVAL

The retrieval of a report from the optical disk system is initiated by typing the report's six-digit accession number into a user-friendly menu on a Sun workstation (see figure 5). The report can then be viewed page by page on the screen, or the viewer can skim through the report, back and forth, at his or her own pace. All or part of the report can be printed out by a simple command. More than one report can be viewed on a screen, and the same report can be viewed simultaneously at more than one workstation.

The ability to identify the report or reports to be retrieved has become a stumbling block for many system designers. Hundreds of thousands of pages of records placed on optical disk can rapidly submerge individual items in a ocean of data. Some designers solve this

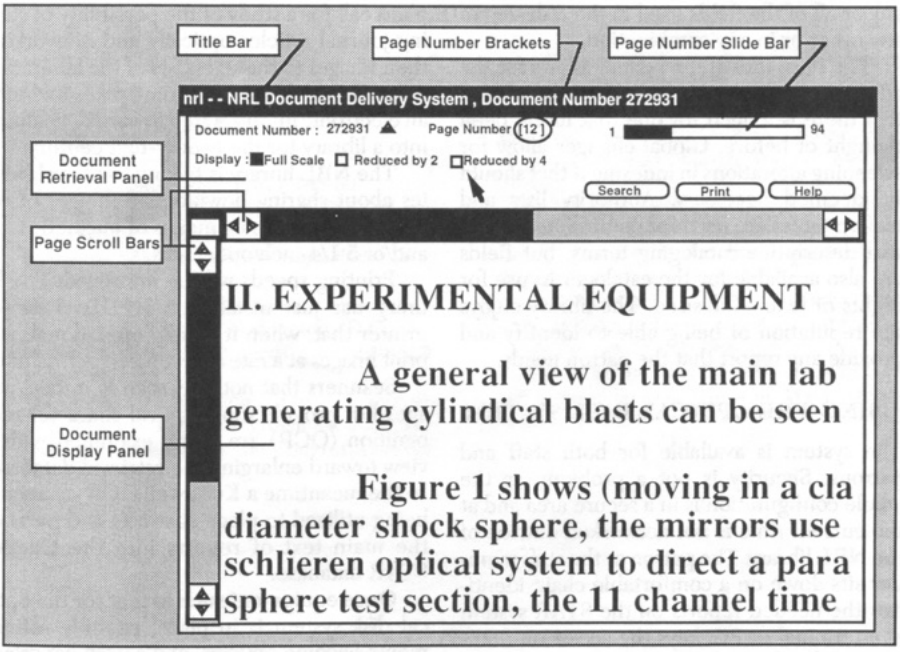


Figure 5. The NRL Document Delivery Sytem Window.

problem by superimposing an indexing system directly on the optical disk, but this takes a lot of time and disk space. Some designers superimpose ASCII characters that can be searched word by word directly on the disk, but this also takes a lot of disk space. Selecting an indexing system can take as much time as or more time than selecting the hardware for the optical disk system. An indexing program can be expensive, and, if the wrong one is chosen, very costly indeed. If the cataloging information is then superimposed on the optical disk, it will be very difficult to change.

The NRL library finessed this retrieval dilemma by opting to keep the retrieval system entirely separate from the optical disk system.

The reasoning was:

1. An automatic retrieval system was already in place.

2. There was a need for the quickest and the most expedient means of getting the reports to optical disk, and a separate retrieval system supported that goal.

3. A separate system is less expensive and easier to manage, input can be done at its own pace, and anything on the system can be changed at any time with little fuss.

4. At the beginning of the planning stage the optical disk system was thought of as simply an alternative way of storing reports.

5. The time saved by combining indexing and retrieval was not considered worth the additional cost and effort that would have been necessary at the time the library started this project. These reports do not have to be retrieved in tenths of a second. Retrieval of a report in less than half a minute is considered satisfactory. Compared with the time necessary to retrieve a dusty report from a real and usually dustier shelf, a half minute seems instantaneous.

Since August 1987, the library has had in place the Cuadra STAR retrieval system. This system provides a very large number of fields so that a report can be indexed in every way that someone might conceive of to identify it—e.g., accession number, title, author, subject, contract, words in an abstract, size, and so on. Searching may be done by using any field separately or combined with others. The fields may be searched in full or by individual words, by subfields, or by masking. Boolean searching of combinations of fields or search results is also possible. The results of all searches can be displayed or printed out by any or all of the fields used in the indexing in any order or in any combination.

The flexibility of the system allows for the addition of a field or subfield if it is discovered that there is a need for one that hasn't been thought of before. Global changes allow for sweeping alterations in indexing if that should be deemed necessary. Authority lists and lookup tables are used for controlling subject and descriptive cataloging terms, but fields are also available for the cataloger to use for flights of fancy if desired. The library enjoys the reputation of being able to identify and provide any report that the patron needs.

USING THE OPTICAL DISK SYSTEM

The system is available for both staff and patrons. Security is not a problem, as the whole configuration is in a secure area and at the current time is not networked outside of the NRL library. The patron or the staff member sits down on a comfortable chair, identifies the needed reports on the STAR system at a terminal or PC, keys the accession numbers of the desired reports into the adjacent Sun workstation, and views and/or prints out any needed pages or reports in a pleasant atmosphere.

PLANNING FOR THE FUTURE

Even though the system used to identify reports is separate from the optical disk system itself, plans are now being made to merge these systems. The indexing system would still be a separate entity but would have incorporated into it the "hooks" used to retrieve the report from the optical disk once the accession number is known.

When this collection consisting of 140,000 reports is scanned to disk, a second collection of 100,000 reports averaging one hundred pages per item will also be put to disk. This will mean another ten million pages of images stored on a separate autochanger to be daisy-chained to the current one. Scanning and printing in color will be considered as the technology becomes more affordable.

Plans are being made for providing images over the campus network to the offices of Naval Research Laboratory scientists. Future plans call for a study of the possibility of storing journal articles optically and networking their images to the scientists of the laboratory as part of the library's ongoing plans to create an electronic library, a step toward developing into a library for the twenty-first century.

The NRL library is talking to other libraries about sharing downloaded images of reports and journals by means of magnetic tape and/or 5 1/4-inch optical disks.

Printing speeds will be improved. The library has just installed a HPIIIsi Laserjet printer that, when it is fully operational, will print images at a rate of seventeen per minute.

Scanners that not only scan in optical images but provide some optical character recognition (OCR) are being examined with a view toward enlarging the retrieval database. In the meantime a Kurtzweil OCR scanner is being utilized to place abstracts and parts of the main text of reports into the Cuadra STAR database.

Once equipment such as that for the optical disk system is in place, possible refinements become obvious, and a wish list grows rather quickly—more workstations for patrons and staff, more information in the retrieval database, faster scanners, faster printers, remote access, and all the new products that this rapidly advancing technology is going to offer.

SUMMARY

The library at the Naval Research Laboratory has designed and installed an optical disk system for the purpose of optical storage and retrieval of its large collection of reports. These reports, averaging fifty-five pages each, are scanned, and their images captured on twelve-inch optical disks stored in a jukebox holding fifty platters. Reports are identified in a separate database containing extensive indexing information pertaining to each report scanned. Once identified, a report can be retrieved from optical storage using the report's six-digit accession number. The images of the retrieved report can be viewed on a screen or printed. Forty thousand reports are currently stored optically, with plans to complete this phase of the project by putting 100,000 additional reports to disk.

REFERENCES

1. Frank Moore, "Internal Revenue Service," paper delivered at the Association for Information and Image Management Conference, Chicago, April 1988.
2. David A. Harvey and Bob Ryan, "Practically Paperless," *Byte* (Apr. 1991): 185-90.
3. Thomas J. Thiel, "Integrated CD-ROM and WORM Optical Disk Systems on the Navy's Paperless Ship," *CD-ROM Professional* (May 1992): 17-26.
4. William L. Hooton, *Optical Disk Storage System Project Report* (Washington, D.C.: National Archives and Records Administration, Mar. 1991).
5. Integrated Automation, Inc., "Facsimile Transmission on the Congressional Cable System" (Final report for the Congressional Research Service, Library of Congress, Washington D.C., Apr. 1987).
6. Ann R. Kenny and Lynne K. Personius, "Update on Digital Techniques," *Advances in Preservation and Access*, vol. 1 (Meckler, 1992).
7. "The Mercury Electronic Library and Library Information System II: The First Three Years," *Carnegie Mellon University Technical Report Series*, no. 6 (Feb. 1992).
8. Masaaki Morita, press release from the American Institute of Image Management Conference, Chicago, April 1988.
9. C. Peter Waegemann, *The Handbook of Optical Memory Systems* (Newton, Mass.: Optical Disk Institute).

The Three *T*'s for a Talking Online Catalog : Technology, Teamwork, Teaching

Wallace C. Grant and Dorothy E. Jones

With the passage of the Americans with Disabilities Act of 1990, there is a renewed emphasis on library collection accessibility for persons with disabilities. Modern computer technology can be adapted so that visually impaired or blind people can independently access today's online catalog. Successful creation and maintenance of a speaking large-print catalog requires a mixture of hardware, software, and caring people. A team with vision and stamina for a continuing promotion and education program is essential. The availability of the adapted online catalog will open new horizons to blind and visually impaired patrons in both the public and academic sectors.

This article is about a new technological service component within a larger library program of library services for persons with disabilities. It is also a case history of the implementation of this new technology, which resulted in the addition of a microcomputer workstation that offers a speaking large-print online catalog, a speaking large-print magazine index, and a word-processing program. Libraries, in widely varying degrees, have for many years provided special services for the blind and visually impaired. Most efforts were directed toward making print materials accessible. For the visually impaired we bought large-print materials or enlarged the images of print material on a screen. For blind persons we addressed a different sensory receptor, buying cassette books, Braille materials, and so on. Some libraries also provided staff assigned especially to offer library assistance to persons with disabilities.

The need to provide service that will make our libraries accessible to persons with disabilities has now been highlighted, underlined, and capitalized by the passage of the Americans with Disabilities Act of 1990. Awareness and concern were expanding steadily even before passage of the act, as evidenced by, and perhaps influenced by, the growth during the past ten years of the body of literature relating to library services for persons with disabilities. This concern has emerged in different ways—concern for accessibility to buildings, to materials, or to programs. Needed changes may dictate new construction and design or alteration and adaptation of building entrances, elevators, service desks, and bathrooms. Other changes involve reassigning personnel, retraining for a wider and more inclusive service concept, changing attitudes, and committing the time to provide services in a special style or format. In libraries we have a clarion call to provide access to the collection—access that, in efficiency and ease, matches as closely as possible the advances made for library users who do not have disabilities. Then there is the responsibility to provide access to the collection in formats that are usable by persons with disabilities.

With the advent of the computer, and all its combinations and permutations, we en-

Wallace C. Grant is Systems Librarian, and Dorothy E. Jones is Coordinator of Library Services for Persons with Disabilities, Northern Illinois Universities Libraries, DeKalb.

tered a new age of service, with new ways to access our collections. Norbert Wiener, in *God and Golem, Inc.*, says that

> the future offers very little hope for those who expect that our new mechanical slaves will offer us a world in which we may rest from thinking. Help us they may, but at the cost of supreme demands upon our honesty and our intelligence. The world of the future will be an ever more demanding struggle against the limitations of our intelligence, not a comfortable hammock in which we can lie down to be waited upon by our robot slaves.[1]

The technology explosion has been of immeasurable help to bright, ambitious persons with visual disabilities. More and more blind and visually impaired persons are going to college. Most of these students are eager for ways to stretch their minds and abilities and are ready for the extra demands new technology will make. Instruments for taking notes in Braille, computers that talk as one types, machines that translate between Braille and print, and Braille printers all offer greater independence for the blind or visually impaired scholar. The computer presents wider horizons for the intelligence and talent that have always been present among people with disabilities. With hard work, persistence, and the use of new technology, disabled persons will create opportunities for education and employment that have been almost nonexistent before.

In libraries, the technology explosion has hit everyone hard and fast. We must take great care that the technology that helps expand patron access does not raise new barriers for the blind. "Many traditional library materials and services have been inaccessible to disabled persons through the years, but technology offers new options for entry into the information age for this segment of library patrons. The very technology that can open doors for the disabled must not possess a new set of unnecessary barriers."[2] New mass-market technology is seldom designed with the needs of disabled users in mind, but an advantage of computer technology is its flexibility and adaptability. "For example, visually impaired persons cannot read a CRT screen to evaluate citations generated from an InfoTrac search. Naturally, they cannot read similar citations from the *Business Periodicals Index*. The technology of the printed source is stagnant; it cannot be altered except through adaptation. The automated product, on the other hand, has the flexibility to permit access for the visually impaired user."[3] According to Pontau, as with other visually oriented formats, the computer needs augmentation by magnification or output aimed at other sensory receptors.

> For handicapped individuals, the primary problems in dealing with computer terminals fall into two categories: manipulation of the keyboard, and dealing with the information displayed. Manipulation problems are usually experienced by individuals with physical handicaps, including the elderly. The second category, dealing with the information presented on the tube, is a problem for both blind individuals and individuals who may have trouble dealing with more complex information. For blind individuals, the problem is primarily one of presenting information in the wrong sensory mode.[4]

As Mates points out,

> While the physically disabled need a way to physically input the information, the blind or visually disabled need some type of output device such as a screen enlarger (which can be as simple as a 'screen-sized' magnifier or a large-print software program) or a speech synthesizer which, when coupled with a screen reading program, converts and reads aloud written ASCII text.[5]

Many colleges, universities, and libraries may have adapted computer labs with word-processing software that makes it possible for visually impaired and blind students to produce papers under the same regimen as other students. However, one of the greatest advances in patron access to library collections, the online catalog, has been unavailable to the blind except through an intermediary person.

In July 1990 at the Northern Illinois University Libraries, specifications and price information were developed for a microcomputer system to provide blind and visually impaired persons access to the online catalog through speech synthesis or enlarged text. This microcomputer is the newest of our adaptive aids for people with visual impairments. The university and the library had been offering service to visually impaired students for years: special-format materials, human readers, talking books, a Kurzweil reading machine, closed-circuit television enlargement devices, a librarian-coordinated service to disabled persons, and a study room for persons with disabilities. We at the library had been providing the tools for independent access to texts of materials contained in the library, but not to the catalog that lists what the library contains. We needed to provide a way for students with disabilities, many of whom are graduate students with intensive research requirements, to conduct, independently, the initial part of their research—exploration into our collection and identification of the materials and information that they wanted to retrieve. The task was to determine the hardware and software mix that would do the job, be within the budget, could be operated easily by faculty and patrons, and could be supported by our systems staff. During the long process of justification, selection, adaptation, installation, and presentation of this new system, we discovered through trial and error that teamwork, technical expertise, and careful teaching are equally important.

THE TEAM

The first step toward getting new technology for persons with disabilities who use your library is to create a strong, committed, caring team. Gather the team together and begin talking about who, what, when, where, why, and how. Then split up the stages of your project and the work to be done—but keep on talking, talking, talking.

We cannot emphasize too much how large a part attitudes of respect and practices of courtesy play in the creation and dissemination of new information techniques within the library. In our time it is particularly true that we librarians cannot be expert in all the aspects of our work and that we must therefore work in teams. The team gathered to put together a computer station that carries database information to be accessed will probably include technical experts, reference librarians, disabilities consultants, institution administrators, equipment vendors, and persons who will use the equipment when it is finished. The various members of the team will be more or less involved at different stages of the project. However, each team member should be present in the awareness of the group all of the time. All members should be kept informed of progress, setbacks, and issues that arise in the course of acquisition, adaptation, installation, and user training. Each member views the project from a specific vantage point and may see problems or possibilities that would not enter the minds of any of the others. Commitment to maintain the communication network is important long after the new piece of equipment is installed. The word "entropy," closely associated with the concepts of thermodynamics, is a scientific term for the amount of randomness or disorder in processes and systems. One symptom of entropy is the tendency of systems to move toward greater confusion and disorder over time. We have all observed systems, both technological and social, that begin with great enthusiasm and eventually degenerate into stagnant centers of bickering or into the oblivion of nonuse because of confusion and disorder. A team that cares and in which everyone is consulted and respected can continue to share concern over a long period of time and can help prevent such a situation.

Here are some examples of areas in which continual communication is necessary:

- Technical team members should be kept aware of the ultimate purpose of the system under construction. They should always be apprised of issues having to do with user education and the attributes of the clientele who will be using the system. They may be able to modify the system, simplify commands, eliminate complicated direction screens, and help tailor the menu to the particular needs of the clientele.
- Team members who are not technically oriented should be kept informed of the technical progress and the problems that arise. They need to be aware of the limitations and powers of the system. For example, public service librarians will probably be the public relations part of the team and need to give accurate information to others. They will also be the teachers of the system and need to have a clear vision of how the unit can serve the patron.
- No changes should be made in the original plan without consultation with the team. Changes might conflict with established goals or with preexisting infrastructure. Together you may find another, better way to get what you need.
- Contact with support groups such as administration or funding organizations must be maintained. It will be to the detriment of all if they become discouraged, disgruntled, or bored.
- The students or other people who will

use the new workstation can help test equipment at several stages along the way. They are absolutely the best people to tell you where they need a sound signal or why the location of a switch is inconvenient.

THE TECHNOLOGY

As Crawford astutely observes, "One catch phrase for human/machine interface design is 'user-friendly.' But what is friendly for one user and one type of use may be hostile to another user and another type of use."[6] The projected user group places design constraints on the new online catalog interface. Each individual online catalog also places unique constraints on the adaptive technology selected. Vanderheider goes on to say:

> Design constraints fall into two categories: user constraints and manufacturer/distribution constraints. The user constraints will vary somewhat depending upon the application for which the special terminals will be used. In general, however, there are four basic user constraints, especially for public access terminals. The system and modifications must be: (1) obvious, (2) easy to learn, (3) easy to set up or connect with, and (4) reliable and easy to maintain. The first two are, of course, general rules; they apply not only to handicapped individuals, but to the population as a whole. The third is appropriate mainly in terms of use by handicapped individuals.[7]

If the method for interfacing the system is simple to connect, use, and maintain, it will be used. If not, it will sit idle.

Enlarging the monitor image is not technically difficult. One can place an optical magnifier in front of the screen, one can get a larger screen or project the image onto a larger screen, and one can use a combination of hardware and software to enlarge the characters present on the screen. The latter has the drawback of limiting the viewing area to a portion of the screen, but allows the greatest flexibility for the user as far as what is enlarged and to what extent.

Speech synthesis has long been a goal of scientists and engineers. VonKempelen built the first speech synthesizer capable of producing both vowels and consonants in 1791. This was a mechanical device that simulated human articulatory organs. J. Q. Stewart made the first synthesizer incorporating an electric structure in 1922. The first synthesizer that actually succeeded in generating continuous speech was the voder, constructed by H. Dudley in 1939. This form of synthesizer is still used today.[8] Human speech is unique and individualized. The mechanical reproduction of speech or its synthesis is complex at best and unintelligible in general. There are four constraints on the design of a speech-synthesis system.

1. The application: A bank funds transfer machine (FTM) only needs a few phrases and can use recorded voices. Telephone information systems can use stored phrases and sort them as needed. Computer text to speech needs a range of techniques.
2. The human vocal apparatus: The human voice is produced by a physiologically complex and sophisticated structure that is trained to reproduce appropriate sounds. These sounds vary from person to person and region to region, so even among humans there are comprehension problems.
3. The structure of language: The variations of grammar and syntax create subtleties in meaning that are all but impossible to imitate. Even when taken within context, the written or spoken language can be misunderstood.
4. The technology available: Until the advent of small, powerful computers, reproduction of human speech by anything other than recordings was impractical. The variations, combinations, and permutations involved were beyond control. Even with the technical capability available today, synthesized speech sounds like synthesized speech.

A computer system consists of three parts: hardware, software, and people. They all must interact properly for the system to be successful in its application. In the grand scheme of systems design, people determine an application, find the appropriate software for the application, and then find the hardware to run the software. This thread of reasoning is frequently broken by preexisting conditions and personal preferences. Some people naturally prefer Macintosh computers to IBMs and vice versa. Institutions use one type of computer rather than another for any number of reasons. Northern Illinois University Libraries is an IBM (and clones) shop, because supporting one type of computer and one type of software is more efficient and more cost effective. Specifications started with an IBM or clone PC and moved on to the rest of the hardware and software to meet the stated needs. The best sources of specifications for systems such as this are similar existing systems you can observe, vendors, in-house technicians, the Computer Science Department, or the college or university computing center. One good list of vendors is contained in Barbara Mates' book *Library Technology for Visually and Physically Impaired Patrons*.[9]

We developed specifications based on our needs assessment by matching our needs with existing hardware and software. We needed a computer that was fast enough and powerful enough to handle speech generation, video enhancement, communications, and information exchange all at once. Speed in this case can be defined in terms of existing computers. The original IBM PC ran at a clock speed of four megahertz; this is very slow. Today's computers run at clock speeds up to sixty-six megahertz, which is very fast. In terms of power, the IBM PC could address one million bytes of memory and move information eight bits at a time. New computers address four billion bytes of memory and move information thirty-two bits at a time. Generally speaking, as the number on the CPU goes up, the speed and power increase. Thus, an 8088 CPU is very slow and less powerful, and an 80486 CPU is very fast and very powerful. Also, as the CPU number goes up, so does the price. Fastest and most powerful are not necessarily best. One must fit the computer to the application to get the most efficient and effective system for the dollar. We determined the minimum microcomputer configuration to be an 80286 processor (PC-AT compatible) microcomputer running at twelve megahertz. The following list identifies the software and hardware capabilities that were desirable for our needs:

Memory

We decided on one megabyte of random access memory (RAM). This is the least amount of memory the system needs to run well. More RAM could have been installed, but additional memory would not have made the system more efficient or effective.

Hard Drive

We chose a forty-megabyte hard drive, as it has enough storage for the software needed to run the system plus additional applications we might want to add in the future. This is a case in which too much to start with is probably not enough in the end.

Floppy Drive

We chose a 5 1/4-inch dual-sided, dual-density floppy drive, as most PCs have that size. We may decide to add a 3 1/2-inch floppy drive later if we have sufficient demand for that form.

Monitor

We selected a nineteen-inch super VGA monitor to give us the size we need for the enlarged print. These larger monitors are significantly more expensive than the standard fourteen inch, but they are better suited to the job.

Modem

We selected a 2400 baud modem for connection with the online catalog. This technology is well tested and provides stable service over a broad range of telephone-line conditions. Reliability was considered more important than speed.

Printer

We chose a nine-pin dot-matrix printer, as it is inexpensive and the print quality suits our needs.

Speech and Video

The special speech-generation and video-enhancement hardware and software were not so easily determined. We in the library had no previous experience upon which to draw, so we viewed and listened to an existing system on campus. If you do not have this option, check with vendors or other campuses. We chose the Vista video board/software and the Vert Plus speech board/software for our system. They are good products that are in use on our campus, so students were already familiar with them and had one less thing to learn.

Having established the system specifications, we contacted vendors for prices and availability of the various components. The total cost of the system as specified came to approximately nine thousand dollars.

PLANNING

When planning for any system, you must leave room for compromise. The components you want may not work together, software may not be compatible, the system may be user hostile, or you may not be able to afford what you want. Plan for change. Also plan for expansion and upgrade if and when you get additional funds. Give thought to what you are willing to substitute or do without. Can you overcome complexity with training? Do you have time to overcome complexity with training? In our case, there was not enough money for the system as specified, so we opted to get a fourteen-inch monitor rather than the nineteen-inch one. This proved workable, but not wholly suitable. The smaller screen does not allow a sufficient quantity of text to be viewed at a time. Scrolling about the screen enlarging segments disrupts reading comprehension. We will replace the small monitor with a larger one when funds become available.

INSTALLATION

We installed the Vert Plus hardware and software according to the instructions in the manual with no problems. The Vista hardware installed easily, but we had problems with the software. Everything was done according to the book, but we could not get enlarged video with the online catalog. After much trial and error, one of our technicians discovered that there was no software interface for the version of PROCOMM, the communications program we were using. We changed PROCOMM versions, and everything worked as specified. This was not an obvious situation, and had it not been for the technician's experience, we would have had a serious problem. Good vendor support is absolutely necessary if you do not have qualified in-house technical support.

Speech Generation and Text Enlargement

The conversion of characters on a screen to understandable speech is a technological discussion beyond the scope of this article; however, in simple terms, the information shown on the screen is contained in a block of the computer's memory that is accessed by the Vert Plus software (see figure 1). Vert Plus software reads the block of memory and converts the content into information that is processed by the Vert Plus hardware into a representation of the human voice, which is then output to the earphones. The user controls the voice output quality and quantity via the Vert Plus software. The user determines whether words are spoken as whole words or spelled out as groups of letters. The pronunciation of words or symbols may be adjusted to the user's taste, as can the output voice's volume, pitch, and rate. The user also has control over the parts of the screen that are turned into spoken output.

Screen enlargement is a similar process. The Vista software accesses the block of computer memory containing the information presented on the screen. The specific data to be enlarged are selected, converted to cover a greater area of the screen, and sent to the

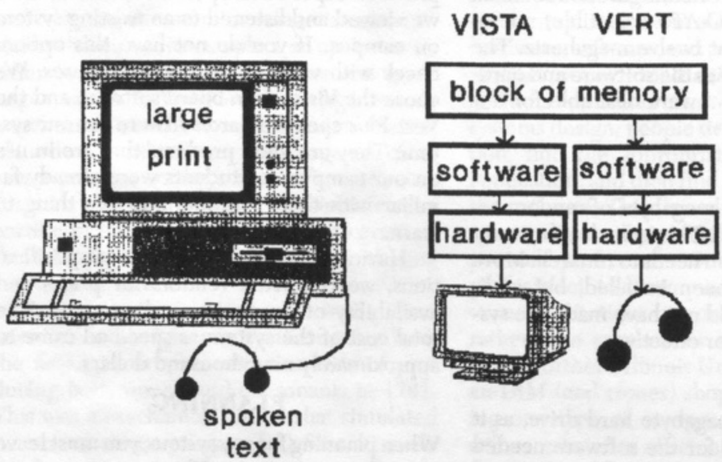


Figure 1. Block Diagram of Speech Generation and Text Enlargement.

Vista hardware. The Vista hardware sends signals to the monitor with both the enlarged and standard images in the proper locations. The user controls the screen-enlargement quantity and quality via the Vista software. The user determines whether the entire screen, a selected portion of the screen, or a single line is enlarged. The user may determine a set level of magnification or may alter the magnification by zooming in or out as needed. The image may be inverted to show either dark words on a light background or light words on a dark background to suit the user's personal taste.

In addition to the software needed to operate the speech and video-enlargement capabilities, we loaded a word processor, a spread sheet, and a communications program, as mentioned above. We have standardized on MS-DOS version 3.3 throughout the University Libraries. This version of DOS works well with this system also.

TEACHING AND ASSISTING

Once the adapted online computer is running *consistently well*, it is time to begin publicizing and teaching. New-system problems must be as well expurgated as possible before you start to publicize. It is unfair to students to ask them to waste their time working through *our* system problems. This is particularly true for students who, because of a disability, must plan study time very carefully and allow more time than the majority student population to read material and write papers.

The instruction program for databases available on adaptive equipment will, like all reference teaching, become a "forever" part of your reference responsibilities. However, unlike most of our reference teaching, it will require concentrated and uninterrupted time with individual patrons. This, of course, has both drawbacks and advantages. It is hard to schedule session times that are convenient for both the librarian and the disabled patron. However, it gives the librarian an opportunity for thorough one-on-one instruction.

Here are suggested steps for a new-database user-training program. The special attributes necessary to or concomitant with a program designed for persons with disabilities will be discussed in the paragraphs following the outline: (1) publicize, (2) schedule, (3) teach (4) maintain, (5) measure and listen, (6) respond.

Publicity

Publicity will need to be of two kinds—publicity regarding the system and publicity regarding the instruction program. Patrons need to be aware of what we have in terms of the machinery. Be enthusiastic but realistic. I believe there is a strong misconception that all people who are into computers are new-age explorers and adventurers who love change and are always ready for new developments. Not true. All of us are by necessity, if not by choice, into computers. We are all human beings, most of whom cling to the familiar. Library technologists, librarians, and students defy group caricature, and reactions to new ways of doing things will have as many variations as there are individuals, from absolute, unthinking rejection to total sheeplike acceptance. Some people tend to become fanatically loyal to a particular brand of computer. Some have impassioned loyalties to favorite software. The students and library patrons who need our adapted computers and databases have the same kinds of loyalties to brands, the same resistance to changes in commands required for use of the system, the same annoyances with variations in the capabilities of both the hardware and software, and the same irritation when results do not, or cannot, match their expectations. However, most will also understand the interdependence of all the systems in the library or the need to build standardization into campus technology programs or other policy-type purchase determinants. Librarians who teach database use to persons with disabilities in an academic setting have a number of advantages and opportunities not available to the general reference librarian. We *can*, in most cases, publicize ahead of time to an identifiable pool of people and ask them to follow a time-and-type specific learning plan. We will not, for the most part, be teaching drop-ins at a reference desk who have had little mind preparation for what they will meet when they come to the library.

In order to publicize the new system, identify as specifically as possible the pool of people who might benefit from adaptive technology. There may be offices on campus with lists of people who have requested special services. These lists will be confidential, but if you have a program of special services in the library, the other special-services offices may be able to get consent for you to use their lists. The librarian working on publicity and giving instruction to patrons should make it very clear that any lists of persons with disabilities that are requested from special-services offices, admissions offices, and so on shall be kept absolutely confidential. Of course, no lists should become service limiting. There are many people with disabilities who might benefit from the new system but who have not declared their disability. To reach this group, you might write articles for campus or community publications.

The preparation of letters, brochures, articles, or other forms of publicity can take a good deal of time and thought. You will have to make decisions about how much and what kind of information is necessary or effective, depending on the publicity format you are using. Below is a list of ideas for inclusion in publicity or notification materials. Obviously, not all of this information can be included in one letter or brochure, but, knowing your own clientele, you will know what is important to them.

1. Describe the computer and the adaptive technology, including brand and model names.
2. Describe what the computer can do. What databases are available? Describe the contents of each database. Is it accessed by menu or special command? Can it be searched by subject, author, title, journal?
3. Describe the instruction plan. Include a telephone number to call for information or to schedule an instruction session. Emphasize the need for instruction. Will you offer a choice between online catalog instruction and other databases, or are *you* going to determine the sequence?
4. Describe any system or software limitations or quirks the patron is likely to encounter.
5. Let people know *when* they can use the computer. Will it be on a first-come-first-serve basis or by reservation? Which uses of the computer have priority? Are you going to reserve the computer for certain uses at certain times?
6. *Let people know you want to hear about their reactions.* Assure the new system users that you want to hear about any difficulties they encounter with the computer, the training procedures, the location of the workstation—-anything. Be sure to mention that you will need their suggestions in order to improve service.

Scheduling

If you are not inundated with eager requests for training, don't be too disappointed. Students are exceedingly busy and are sometimes loath to add one more learning responsibility to their load. A slow initial response to the new workstation may simply reflect the fact that a good program is already in place to assist students to identify and retrieve their library materials. It is hoped that after a few brave souls have called and learned how to identify the library materials they want for themselves, the word will spread and requests for training will increase.

In most cases, it is best to schedule instruction sessions of only one hour. Sometimes students ask to move on during the session, from, say, use of the online catalog to use of other menu items (journal indexes and so on), which you may have loaded onto the same workstation. A good teacher needs to be flexible enough to respond to the abilities and particular learning habits of individual students. However, it seems to be good general policy to instruct in one database at a time and then ask the student to practice using that database before he or she begins another. Most of us who work on reference desks have seen library users sitting at terminals looking for books in journal indexes and looking for articles in the book catalog. Working one on one, we can make sure these particular students know exactly what each database offers and how to get what they need from it. Schedule one session with each student and, at the end of that session, schedule another session if it's needed—either to review and embellish the session just completed or to move on to another database with another set of commands and differently organized results. If training sessions for different databases are separated, the databases are more likely to stay separated and clear in the mind of the student.

Teaching

There are just a few special teaching tips that we are learning as we instruct patrons with disabilities in the use of the new computer.

1. Teach the most concise and direct method of access rather than the easiest. If you have a choice between a user-friendly system with many explanatory screens, and a command system that goes directly to the

information commanded, teach the commands. The students will, for the most part, appreciate directness and speed of access. Listening to screens of explanation when using voice option, or having to scan screen after screen of unnecessary material (remember large-print access diminishes the amount of material visible on a single screen) is time-consuming, confusing, and exhausting. The necessity for concentration is very intense when one is trying to gather information from a computer-generated voice as well as manage the computer and voice from an unseen keyboard.

2. Listen very carefully to your student, and give the student time to try out each step of what you are teaching. Unless you have exactly the same disability as the student, and probably even if you *do*, you will not be able to gauge, by yourself, exactly what the student needs and where the problems of understanding are rooted. Any student who asks for your instruction is motivated to know. If you listen carefully to his or her questions, and if you are also aware of the silences born of confusion, you will be able to travel back over material, rewording instruction and encouraging repetition of the hands-on experience. Try very hard to convince the student that no question is stupid. Each of us has had the experience of knowing that we don't know and yet being afraid to ask a question because our question might betray more ignorance than our teacher is willing to accept. This should never happen between teacher and student. Fear of betraying ignorance sometimes stops the learning process cold. At other times, the student and teacher bypass the problem, continue the lesson, and leave a hole in the knowledge fabric that is likely to create flaws elsewhere in the pattern.

3. If the student can have in mind a subject or research project to work on during the sessions, this will be a help to you. The teacher and the student will also feel good to have accomplished something tangible—the identification of material for a specific paper, for example—during the instruction session.

Maintenance

Maintenance of your user-instruction program may take a little time or a lot of time. Once you have taught one group of users, you must prepare to teach another group. New students will arrive each semester. Always try to incorporate new teaching techniques you've learned from your students as well as from your reading or observations. Toss out things that haven't worked. You may decide to emphasize certain menu items and not others as you observe the success students have working with them. Also, you will undoubtedly find things in the system or the software that you wish were different. Keep track of these things and talk with your technical experts. Many times they can find a way to bring about the changes.

Measuring and Listening

Once students are trained and can use the computer on their own schedule, just as students use the other computers in the library, you will want to know how much and what sort of use the new workstation is getting. Measuring its use can be difficult, particularly if it is in a separate room. You may want to depend on interviews or conversations with your patrons. Patrons are, in the long run, the only dependable guides to what they need, want, and are willing to use. It doesn't matter how wonderful a new technology is if the people don't have the time or won't take the time to learn it and use it. Listen to comments on why users like or dislike the computer, particular databases, or the location of the workstation. Listen if they say that it is too noisy or that having people around makes them nervous. You won't be able to give everyone exactly what he or she wants, but you will be able to make *some* improvements.

Responding

Try to respond to every suggestion or complaint made by a patron—even if the response is, "I'm sorry, there's nothing we can do about that." But don't respond too quickly—don't say no until you have truly explored the problem. Ask your administration, technical people, or whomever else might be appropriate about the possibilities of acquiring, altering, moving, partitioning, and so on. Sometimes the unexpected happens, and, almost always, you will eventually make at least a little progress. As Franklin Roosevelt said, "The test of our progress is not whether we add more to the abundance of those who have much; it is whether we provide enough for those who have too little."[10]

REFERENCES

1. Norbert Wiener, *God and Golem, Inc.: A Comment on Certain Points Where Cybernetics Impinges on Religion* (Cambridge, Mass: M.I.T. Press, 1964), 69.
2. Donna Zufan Pontau, "The Use of Technology for the Visually Handicapped in Libraries," in *Convergence: Proceedings of the Second National Conference of the Library and Information Technology Association, October 2–6, 1988, Boston, Massachusetts*, ed. Michael Gorman (Chicago: American Library Assn., 1990), 146.
3. Ibid., 143.
4. Gregg Vanderheiden, "Modifying and Designing Computer Terminals to Allow Access by Handicapped Individuals," in *Public Access to Library Automation: Clinic on Library Applications of Data Processing (17th: 1980: University of Illinois at Urbana-Champaign)*, ed. J. L. Divilbiss (Champaign, Ill.: University of Illinois Graduate School of Library and Information Science, 1981), 99.
5. Barbara T. Mates, *Library Technology for Visually and Physically Impaired Patrons* (Westport, Conn.: Meckler, 1991), 2.
6. Walt Crawford, "The Online Catalog in the Real World: Making the Necessary Compromises," in *Design and Evaluation of Computer/Human Interfaces: Issues for Librarians and Information Scientists: Clinic on Library Applications of Data Processing Held at the University of Illinois at Urbana-Champaign, April 17–19, 1988*, ed. Martin A. Siegel (Champaign, Ill.: University of Illinois Graduate School of Library and Information Science, 1988), 134.
7. Vanderheiden, 110.
8. Sadaoki Furui, *Digital Speech Processing, Synthesis, and Recognition* (New York: Dekker, 1989).
9. Mates, 185–90.
10. Franklin D. Roosevelt, "Quarantine the Aggressors," 2d inaugural address, in *The Public Papers and Addresses of Franklin Delano Roosevelt*, 1937 vol., "Constitution Prevails," comp. and ed. Samuel I. Rosenman (New York: Macmillan, 1941), 5. ■■

Improved Browsable Displays: An Experimental Test

Bryce Allen

Browse searching of online catalogs and electronic indexes is a popular means of retrieving information, but some subject heading lists are so extensive that browsing them can be difficult and time-consuming. This research tested one way of presenting a browse interface to reduce the amount of scanning of subject headings required. It found that a hierarchical presentation of headings reduced the amount of scanning required by more than 50 percent without any reduction in the effectiveness of the search. Better browsable displays of this sort can improve the efficiency of searching but appear not to alter effectiveness.

Browsing in automated catalogs and indexes is a popular means of finding information, probably because it is relatively uncomplicated. Typically, a user enters a search expression, such as an author's name, a title, or a subject heading. The information system then displays a sorted list of authors, titles, or subject terms, with the user's search expression highlighted. The user is able to scan up and down the list, selecting one or more items from the list for further processing. This additional processing usually involves retrieving and scanning the references that are indexed by the selected terms from the browse list.

The simplicity of this kind of access can be compared with the relative complexity of search systems requiring the use of Boolean operators or identification of building blocks and synonyms. As a result, many users of CD-ROM indexes and OPAC systems rely upon browse searching to retrieve information. There are, however, problems with browse access. One of these problems is that some information systems have so many index entries that it is a time-consuming and difficult task to scan through them to find appropriate terms or headings. This seems particularly true in subject access. In most cases, browse access to subjects means scanning the subject index of the OPAC system or the thesaurus of the CD-ROM system. With large databases such as those found in OPACs, the subject index may be so large that scanning even a relatively restricted section of it is difficult.

Massicotte, noting the results of OPAC evaluations and research into subject access by Atherton Cochrane and others, pointed out that even for a relatively restricted topic it may be necessary for a user to scan through hundreds of subject headings in order to obtain an overview of the topic.[1] She suggested that it would be better to restrict the number of subject headings that users have to scan initially, ideally to a single screen of headings for any topic. Her suggestion was that the browse display be reduced by replacing many headings with broad conceptual categories. For example, the many geographical subdivisions of some topics could be replaced by a single note, "Subdivided by geographical areas."

McGarry and Svenonius followed up this suggestion by exploring two ways in which restricted lists of subject headings could be created to minimize the scanning of subject headings that was necessary for subject browsing in online catalogs.[2] The first was Massicotte's geographical compression, and the second was an even more radical compression they called blanket compression. In

Bryce Allen is an Assistant Professor, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.

blanket compression, all subheadings were to be suppressed in initial displays of subject headings by using an algorithm that would recognize repeating first elements of headings. In addition, they suggested that the problem of divided displays be resolved by incorporating into a single alphabetical sequence the different ways that subject headings are punctuated.

As part of a larger series of investigations, this project explored experimentally the effects of providing this type of enhancement to browsable subject access. In particular, the first element of McGarry and Svenonius' blanket compression was implemented in an experimental system in such a way that subdivisions were initially suppressed, but could be displayed on demand. As McGarry and Svenonius noted, this type of compression is easy to implement in a "semi-automatic" manner, as opposed to the conceptual compression suggested by Massicotte, which would require more detailed programming to implement. This experiment represents a first attempt to determine how changing the format of browse displays might affect information retrieval performance for end users.

METHODOLOGY

Interfaces

Two interfaces, programmed in Z Basic on an Apple Macintosh, presented a list of subject headings to the participants. The list was displayed on the standard Macintosh screen, with twenty-three headings appearing on each screen. Browsing through the display of subject headings could be done in a number of ways. One way was to enter a search expression. The list scrolled to the heading most closely matching the search expression entered. Other options for moving through the list used the keys PG UP or PG DOWN or the up or down arrow. The standard Macintosh scroll bar at the right side of the list was also available to allow the user to move through the list a page at a time or a line at a time.

Headings from the list were selected by highlighting them (either by clicking on the heading or by scrolling to it), then hitting the return key. When this was done, a window appeared identifying the selection that had been made. No references were displayed by the interfaces, because the focus in this experiment was on mechanisms for scanning subject heading lists rather than on operational retrieval systems.

The first interface included only the capabilities mentioned to this point. It presented a list of headings that allowed participants to scroll up and down using a variety of mechanisms and to select those headings considered useful. This list contained 747 different headings, including many with one or more levels of subheadings.

The second interface was identical to the first in all respects, except that it had two buttons at the top of the display, labeled EXPAND and CONTRACT. The initial display showed only the thirty-nine top-level headings (i.e., those headings that consisted only of a main heading without a subheading). Nineteen of these headings were followed by an asterisk indicating that those headings could be expanded. If one of those nineteen headings was highlighted and the EXPAND button clicked, the list was expanded to include all of the second-level headings under that main heading (i.e., all of the headings consisting of the main heading followed by one subheading). Subsequently clicking on the CONTRACT button while the main heading was highlighted contracted the list so that the second-level headings were no longer displayed. Of the 290 second-level headings, 101 were identified with an asterisk, meaning that they could be expanded to show third-level headings (i.e., those consisting of a main heading followed by two subheadings). There was a total of 418 of these third-level headings. For example, it was possible for a participant to scan through the original list of thirty-nine top-level headings and select the heading "Airlines." Clicking on the EXPAND button would display sixty second-level headings, such as "Airlines—Deregulation." The participant could then scan through those sixty headings, select the "Airlines—Deregulation" heading, and click the EXPAND button again, thus revealing a third-level heading "Airlines—Deregulation—Economic aspects." Figure 1 illustrates the main window of this interface, showing the EXPAND and CONTRACT buttons and the initial state of the display before any headings were expanded. The heading "Airlines—Commercial" has been selected.

Figure 2 illustrates the main window after the EXPAND button was clicked. The

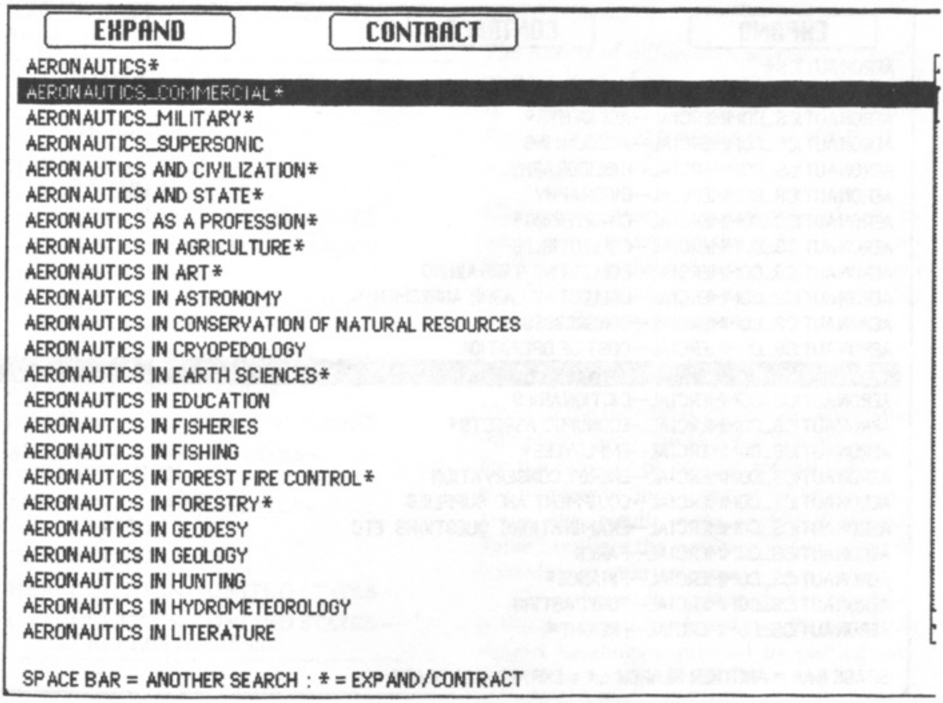


Figure 1. The Expand/Contract Interface Before Expansion

relevant heading "Airlines—Commercial—Deregulation" has now been selected.

Participants

Eighty participants from the student body of the University of Illinois at Urbana-Champaign completed the experimental tasks in this study. These volunteers were obtained by advertising on campus and were paid three dollars for their participation, which lasted about half an hour. The demographics of this group indicated that they represented a broad cross section of the student body.

Materials

The stimulus document for identifying the topic of the search was the title and abstract of the book *Deregulation and Airline Competition* (Paris: OECD, 1988). The abstract contained 103 words, and the Flesch Reading Ease Test indicated a score of 25.7, suitable for most undergraduates. The topic "airline deregulation" was used by Connell, who found that changes in subject cataloging practice made finding exhaustive information on this topic rather difficult.[3]

The list of subject headings, containing 747 different headings, was created by editing the subject index of the online public access catalog in use at the University of Illinois. In preparing the list, only headings beginning with the terms "aeronautics," "airlines," or "deregulation" were selected. This meant that the list was more likely to contain potentially useful subject headings than any operational subject heading list.

Procedures

Participants first read the stimulus title and abstract. They were instructed in writing: "Please read carefully the following information about a book. In a few minutes you will be using subject headings to find additional books on the same topic." Then the participants completed two unrelated pencil-and-paper tests of perceptual speed. Following this testing activity, which was designed to provide a break between reading about the topic and conducting the search and to provide data for other research in progress, the participants were instructed in writing: "A few minutes ago you read a brief description of a

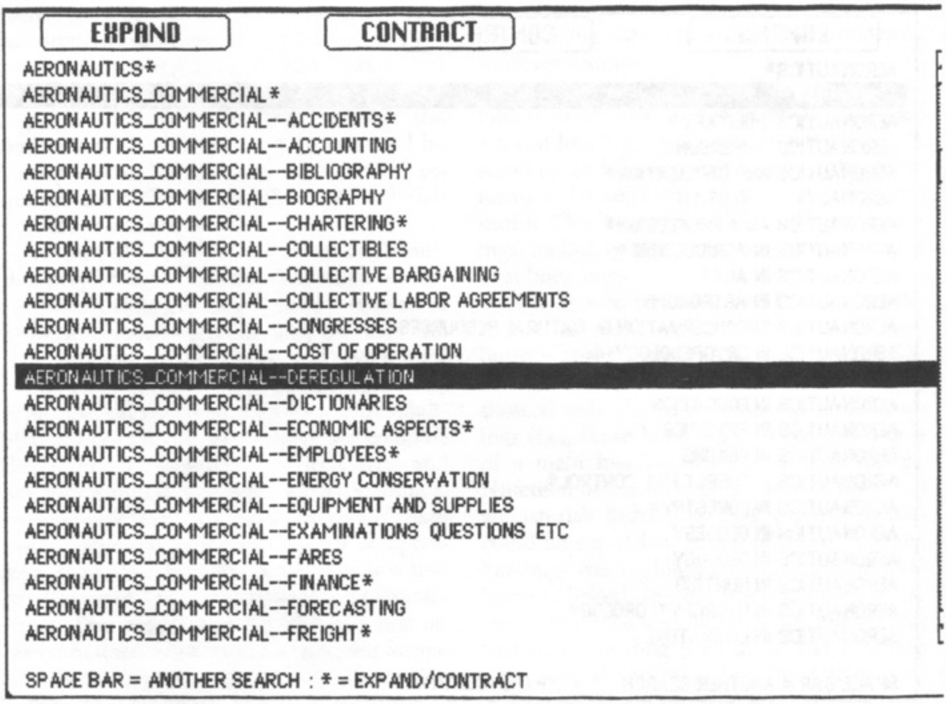


Figure 2. The Expand/Contract Interface After Expansion.

book. Now, assume that you are doing a term paper on that topic for one of your courses. To do your term paper, you want to find more books on the topic. You will be searching a computerized index to find subject headings that you think would lead you to more materials on the topic. Try to find at least two subject headings that you think would be useful." They were then given instructions on how to use the interface to which they had been randomly assigned. No advice was given regarding how best to search for the information, but questions about the working of the interface were answered. Participants were also instructed that once they had found the two subject headings required, they could continue to search the interface for additional subject headings if they wanted. They were told that they should govern the amount of this additional searching activity by their perceptions of how much they normally would search an index of this sort when preparing a term paper. Once the search was completed, participants answered a few questions about themselves on a two-page questionnaire.

Analysis

Dependent Variables

Built into the interfaces were data collection mechanisms that provided the following variables for each search: total time spent on the search, number of search expressions entered, number of subject headings selected, number of clicks of the scroll bar, number of uses of each of the up arrow, down arrow, PG UP key, and PG DOWN key, and, where appropriate, the number of clicks of the EXPAND and CONTRACT buttons. Summing the number of scroll clicks and keystrokes gave the total number of browse actions initiated by the participants. Repeating this sum with the number of PG UP and PG DOWN clicks multiplied by twenty-three gave the number of lines scanned.

The interface programs also recorded which headings were selected by participants as being potentially useful for their term papers. From this recorded data it was possible to ascertain which of the headings were selected most frequently, and from this fre-

Table 1. Most Frequently Selected Subject Headings

AERONAUTICS, COMMERCIAL—DEREGULATION
AIRLINES—DEREGULATION
AIRLINES—DEREGULATION—ECONOMIC ASPECTS
AIRLINES—GOVERNMENT POLICY
AIRLINES—GOVERNMENT POLICY—GREAT BRITAIN
AIRLINES—GOVERNMENT POLICY—UNITED STATES
AIRLINES—LAW AND LEGISLATION
AIRLINES—REGULATION
AIRLINES—SAFETY REGULATIONS
AIRLINES—UNITED STATES—COMPETITION
AIRLINES—UNITED STATES—REGULATION
DEREGULATION
DEREGULATION—UNITED STATES
DEREGULATION—UNITED STATES—COST EFFECTIVENESS

quency-of-selection data it was possible to identify those headings that might be considered particularly relevant to the topic. There were fifteen headings selected by more than ten participants; these are given in table 1.

The number of these headings selected in each search was recorded as a rough measure of the recall achieved, and the number of these headings selected in each search, then divided by the total number of headings selected, was recorded as a rough measure of precision achieved. In addition, the number of headings selected and the number of these "relevant" subject headings selected were divided by the time taken to complete each search. One final dependent variable was given by the questionnaire data: the perceived difficulty of finding subject headings reported by participants.

Independent Variables

Because the focus of this research was on browse interfaces, the independent variable was the interface the participants used. Additional predictor variables assessed through the participant questionnaire included familiarity with the topic searched, with libraries, and with microcomputers.

Control Variables

The effects of differences in academic background, level of academic study, gender, and age were controlled in this study by including these variables in the multiple regressions.

Analysis

All dependent variables were entered individually into stepwise regressions, which included all of the independent and control variables, with categorical variables suitably coded as dummy variables.

FINDINGS

None of the independent or control variables influenced any of the search outcomes, except for the number of browse actions done and the number of lines scanned by participants. Here, the results were highly significant. Table 2 outlines the results of the experiment, according to interface type.

The only result not included in this table is the perceived level of difficulty in finding subject headings expressed by participants. There was no significant difference between the two interfaces, but there was a significant predictor in the knowledge variables. Users with higher self-reported frequency of library use perceived less difficulty in completing the experimental task (Spearman's $R = -.238$, $p < .05$).

DISCUSSION

This experiment showed clearly that an improved browsable subject list of the type suggested by Massicotte and by McGarry and Svenonius is likely to have the desired effect on searching. It substantially reduced the amount of scanning that was necessary to find subject headings, without reducing the number of headings found or the number of potentially relevant headings found and without increasing the perceived difficulty of the search task. But the enhanced interface achieved no improvements in performance. In fact, the searches achieved virtually identical results, regardless of which interface was used.

The differences in the number of browse actions used by participants in finding and selecting subject headings is dramatic, as is the difference in the total number of headings scanned on average by participants using the two experimental systems. Because the inter-

Table 2. Outcomes of the Experiment

	Standard Browse Interface		Expand/Contract Interface	
Search Outcome	Mean	St. Dev.	Mean	St. Dev.
Time taken to complete search (minutes)	5.95	2.96	6.30	2.46
Search expressions entered	1.03	1.09	0.80	0.79
Subject headings selected	10.23	8.11	9.18	6.72
Subject headings selected/minute	1.70	0.67	1.45	0.71
"Relevant" subject headings selected	3.63	2.88	3.78	2.19
"Relevant"subject headings selected/ minute	0.69	0.60	0.63	0.37
"Precision"	0.44	0.37	0.49	0.28
Browse actions	307.48*	297.55	173.70*	125.59
Lines scanned	419.68*	388.74	184.70*	121.51

*p < .01

faces used in this experiment were designed to be identical in every other respect (for example, in the number of headings displayed per screen), these substantial differences can only be attributed to the reorganized subject heading display in the expand-contract interface. This reorganization enabled participants to move reasonably directly to relevant headings, rather than requiring them to scan many irrelevant headings en route.

There were some incidental findings in this experiment. One was the extremely low numbers of search expressions entered. When presented with a direct-manipulation browse interface, most users seemed to prefer to find a single starting point, then to browse from that starting point. Another point worth noting is that knowledge of the topic and familiarity with computers and libraries had no effect on search outcomes, although there was an effect of familiarity with libraries on perceived difficulty. This supports the findings previously reported that knowledge variables tend to influence perceptions of performance rather than actual performance in end-user searching.[4]

There is no question that searching an improved browsable subject index is more efficient than searching standard browsable indexes. There is no evidence to date that this type of index makes subject searching more effective. Additional research, using larger and more realistic subject lists, will be necessary to determine the extent to which search effectiveness may be influenced by such a system. At this point, designers of end-user subject-access systems must balance the additional complexity of creating and maintaining a hierarchical subject heading display system against the clear efficiencies provided by such a system.

REFERENCES

1. Mia Massicotte, "Improved Browsable Displays for Online Subject Access," *Information Technology and Libraries* 7 (December 1988): 373-80.
2. Dorothy McGarry and Elaine Svenonius, "More on Improved Browsable Displays for Online Subject Access," *Information Technology and Libraries* 10 (September 1991): 185-91.
3. Tschera Harkness Connell, "Librarian Subject Searching in Online Catalogs: An Exploratory Study of Knowledge Used" (Ph.D. diss., University of Illinois at Urbana-Champaign, 1991), 287.
4. Bryce Allen, "Cognitive Differences in End User Searching of a CD-ROM Index," in *SIGIR '92: The 15th International Conference on Research and Development in Information Retrieval Proceedings*, ed. N. Belkin, P. Ingwersen, and A. M. Pejtersen (Baltimore, Md.: ACM, 1992), 298–309. ■■

Toward the New Millennium: The Human Side of Library Automation (Revisited)

Kitty Smith

The purpose of this paper is to reexamine some of the human factors of human-machine systems from an individual and organizational behavior standpoint, particularly as they relate to the automation of libraries. Rather than emphasizing the findings of ergonomics' "human factors" research, the text focuses on how important to the success of human-machine systems are individual, social, and organizational responses to automation. It is hoped that what is presented here will serve as food for thought to library managers and others who bear the responsibility for automated products, systems, and services, and to the professional reader seeking further understanding of the role and impact of automation on a traditional human (and frequently bureaucratic) institution, the library.

THE HUMAN CONDITION—WHO IS SLAVE AND WHO IS MASTER?

The fact is that civilization requires slaves. . . . Human slavery is wrong, insecure and demoralizing. On mechanical slavery, on the slavery of the machine, the future of the world depends.—Oscar Wilde[1]

Are we really the masters of our machines, or have we made ourselves slaves of computers? Or do we perhaps need a new paradigm, that of interdependency, to describe our current relationship with our automated technologies? Today, the computer's presence is so pervasive that we barely notice it. People drive cars, find information, transact business, buy and sell, bank money, check books out of the library, pay bills, phone home, get instruction, work, and play with the help of computers. It is a rare day when human beings, their lives, and their occupations are not affected by applications of computer technology. Libraries are no exception. Only ten years ago, the term "library automation" was chiefly identified with large libraries using large systems for handling large accounting or inventory-like tasks, such as circulation and acquisitions. These systems ran on large computers and were accessed via large (but "dumb") terminals. Now, as the result of astonishing advances and the availability of microprocessor, CD-ROM, and other miniaturized technologies, plus interactive data communications capabilities, the world is literally at our fingertips. Library automation is accomplished on a desktop (or laptop) workstation or, through simple phone connections, through a local, nationwide, or worldwide network.

With the approach of a new century, it is natural to pause and reflect on progress achieved in the way we live and work. Contemplating the current electronic age, one might even be moved to offer thanks for the benefits the computer has brought to our lives and workplaces. For many libraries and librarians, the computer has found its niche so swiftly and quietly that we may often take it

Kitty Smith is Assistant Professor, Department of Library and Information Studies, School of Education, University of North Carolina at Greensboro, and Principal, Pathways Consultants, Greensboro, North Carolina.

for granted. Indeed, it may be difficult to remember the days when we occupied precious time with a multitude of servile, routine, labor-intensive tasks, now performed by our computers, which take care of the menial work, freeing us for more creative, "big-picture" responsibilities.

Or so we had hoped.

When the system is down, is slow, contains wrong information, or is not flexible enough to handle the inevitable exceptions, individuals become acutely aware of its limitations. In many instances, it is we who must accommodate ourselves to the demands of the computer, providing answers to its questions, rather than the other way around. For example, for some tasks, the computer requires us to change even the way we think. That is, different conceptual skills are needed when we move from manual systems to automated ones. Kriz and Queijo have described this change as a cognitive shift from physical cues to an abstract, "virtual information space."[2] They compare the process to that of describing to a stranger the route to a destination we ourselves could reach without conscious thought. The task requires an abstraction, the construction of a "mental map." This is the very same process employed when one uses keyboard commands to bring bibliographic information up on a screen instead of handling objects such as catalog cards or paper indexes.

The latest microprocessing technology enables libraries to store and process vast amounts of information in infinitely smaller amounts of space. Hardware, software, and computer communications tools have had an unbelievably swift evolution in the second half of this century. On the other hand, the pace at which human beings adapt these tools to their own purposes, and adapt themselves to new technologies, varies from individual to individual and can affect not only a person's productivity, but also his or her sense of physical, emotional, and social stability and control.

Unrelenting demands for human adaptation, even if only temporary, can be difficult and painful. They can sometimes result in physical symptoms (headaches, backaches, eyestrain) and mental fatigue as well as feelings of powerlessness, anxiety, distrust, self-doubt, and alienation. Psychologists tell us that accumulated anxiety and distrust are experienced as stress (or, more accurately, "distress"). We may use a variety of defense mechanisms to dull the effects of too much stress, but eventually it takes its toll. The body's primitive physical-emotional response to stress (flight or fight) is experienced as fear or anger—fear about being controlled by uncontrollable forces, anger as an attempt to regain some control. More often than we might like to admit, our responses to technological change may tap into deeper fears centered around issues related to people's identities, their control of their own destinies, their worth in the workplace, and their potential for displacement or replacement by automated processes. The computer itself can become the target of misdirected fear and anger.

In the cloud of emotion, the following critical facts may get overlooked:

1. Computers are nonthinking, impersonal machines. They are designed to follow the instructions of their human masterminds. They consistently and tirelessly follow these instructions, even when the instructions are (in human terms) flawed and inflexible. As such, the computer may often serve as both a scapegoat and as an excuse for further bureaucracy and lack of compassion in the organization.[3] (For example, "I'm sorry, but our policy forbids us to check out any more materials to you until we can clear the overdue materials you *claim* to have returned from our computer records.") In reality, of course, the computer itself is not the enemy, merely a sophisticated electronic tool upon which many of us depend.

2. People and computers are components of a larger, more powerful, but often less perceptible, system—the organization. However, even though they seem to have a life of their own, at their core organizations are essentially human.

Now, more than ever before, regardless of the size of our machines, there is an urgent need at the organizational level to consider the human behavioral factors involved in the constructive use of technology. We must inquire within ourselves: Are we using technology to enhance the capabilities of our human resources—or are we using it to further enhance the bureaucratic tendencies of the organization? Is the computer a tool to help people get more and better work done—or is it a vehicle for consciously or unconsciously keeping employees and users in their places in the hierarchical, superior-subordinate power structure?

ATTITUDES ABOUT AUTOMATION

A decade ago, in building a case for human engineering, computer scientist Henry Ledgard noted the preoccupation of the literature of the time with the technology of computerization rather than with its users:

> In the development of any technology there is always a tendency to lose sight of the basic problems that stimulated its introduction. After the first flush of success, interest naturally tends to focus on the technology and the problems it presents. We often forget that the original objective of computer technology was not just to develop more powerful systems but to increase the overall effectiveness of a human problem solver.[4]

While most veterans of library automation would argue that computers in libraries are here to stay, there are some who would acknowledge it reluctantly. Robert Walton has noted a general assumption that automated systems will enter a "positive work environment," one in which the organizational work force eagerly awaits technological change, prepared to adjust its traditional ways of thinking and procedures to accommodate the machinery, fully believing that its potential benefits vastly outnumber its potential harms and limitations. Walton adds, however, that this assumption is primarily the impression of middle and upper management, whereas employee reactions and attitudes differ considerably.[5]

Employee Attitudes

Many employees appear to have a more or less unaware and uninformed, but generally positive, view of the potential benefits of automation, and they expect to take full part in training and use. However, at one end of the spectrum, there can always be found a number who view automation as a panacea for all organizational ills and may be driven to capitalize on opportunities to involve themselves in all phases of its implementation. At the opposite extreme, some employees may be openly resistant, hostile, or indifferent to automation to the point of merely acknowledging its existence or interacting with it only when forced to. Each of these attitude types presents potential challenges for the administrator, whether new computer technology is being introduced to library functions for the first time, or whether the technology is already on the scene. For while people may adapt themselves to an automated workplace over time, their basic human, social, achievement, and growth needs endure and must be fulfilled throughout life. Automating a traditional institution, such as a library, may give it a more contemporary technological image. However, automation in and of itself does not make an institution less traditional and more innovative. In fact, unless the other (human) components of the system are considered as part of the system equation, automation could merely add another layer of rigidity and another excuse for keeping an already bureaucratized institution the way it is. To understand why this situation might occur, it is helpful to look at the attitudes people hold toward traditional institutions, such as libraries.

Attitudes about Libraries

As might be expected, most of the early research on library automation concentrated primarily on the improvement of efficiency, financing, and procurement processes, and very little on the problems of integrating automation into a human environment.

A possible reason for this is that until libraries entered the electronic age, the average person's concept of librarianship was very closely identified with libraries as institutions or places, rather than with the professional persons who operated them. F. Wilfrid Lancaster has asserted that librarianship is among the "most institutionalized of all professions."[6] Physicians are not generally seen simply as people who work in hospitals; lawyers are not merely people who work in the courts. The term "librarian," however, is commonly applied to any person who works in a library, with little distinction made between routine and clerical activity, which the client ordinarily sees being performed by support and professional staff alike, and those tasks requiring professional specialization in the acquisition, organization, and retrieval of information.

Even though many library professionals have worked hard to change the librarian's image from that of guardian of the books to that of information professional, the availability of highly specialized information services, especially in electronic formats, is still a relatively recent addition to many libraries, especially school and public libraries. Moreover, where present, these information services themselves remain identified with their institutional settings, the libraries, rather than with the librarians, apparently with the tacit consent of the majority of professionals. Admittedly, this institutional identification has begun to blur with the availability of direct home or office access to information systems.

Along these lines, Dowlin saw library automation as a valuable tool, offering the library user a combination of books and other materials for knowledge, plus computers for information.[7] The question is evolving from "if" to "how" and "how soon," and undoubtedly will effect a deinstitutionalization of librarianship.

On the dark side, however, futurists such as Robert Theobald and library educators like S. D. Neill have predicted that unless librarians clarify their "linking function" and become even more involved in social change, they and their institutions might evolve into virtual roadblocks to information.[8] These writers see information technology as a means of ending local control of information and expanding the horizons of a free society.

The impending character of significant change is bound to elicit resistance to such change in many individuals who have a stake in preserving the status quo. Those who sense the vulnerability of their personal power are most likely to identify with and cling to institutions, artifacts, and political structures of the past. As Machiavelli proposed in *The Prince* in 1532:

> [T]here is nothing more difficult to plan, more unlikely to succeed with or more dangerous to manage than to take the lead in introducing new methods of government, because the introducer has as enemies all those who profit from the old methods, and as only lukewarm defenders all those who will profit from the new ones.[9]

In light of recent history and the current scene, the role of libraries and librarianship seems destined to continue to change.[10] Automation has not only become a tool for doing things better; it has changed the things we do. Libraries and librarians can no longer choose between changing and not changing (if indeed they ever could). The real choice seems to be either conscious, proactive planning for the integration of change into professional functions, or eventual obsolescence. Changes in individual feelings and attitudes, however, often must be preceded by external events, rather than the reverse, and herein lies the managerial challenge of library automation in the twenty-first century.

THE MANAGEMENT CHALLENGE

The successful introduction of change requires managers to make assumptions and predictions about future outcomes, often with very little relevant data on which to base decisions. This can cause management considerable discomfort, especially when the alternatives consist of continued arbitrary decision making or constant reaction to unplanned events. However, the process of implementing change becomes more acceptable if one takes the broader view of the organization as a consolidation of both human and technical elements. Integration of an automated system into its human environment calls for careful attention to overall system objectives, the separation and allocation of functions to either people or machines, and the design of the human-machine interface (work spaces, work flow, hardware, and software). Recognition of the different capabilities of human beings and machines seems a logical preliminary step in the allocation of functions to one or the other. Jordan, discussing the relative advantages of machines versus people, proposed the following summary: "Men are flexible but cannot be depended upon to perform in a consistent manner, whereas machines can be depended upon to perform consistently but have no flexibility whatsoever."[11] A strategy for successfully integrating such a combination would include maintaining a focus on the operation of the whole (organizational) system, and on developing jobs that are interesting, motivating, and challenging and meet the needs of the human members of the system. This kind of job design process, in turn, includes not only task and work-flow analysis but also recognition of human values, especially regarding personal space, territoriality, density, privacy, convenience, physical comfort, health, safety, security, aesthetics, and opportunity for social interaction. Such factors, plus the opportunity to personalize one's work space, can contribute to a sense of control and status and to a more positive attitude toward the work environment.[12] Optimal functioning of the whole system depends on successful integration of these human elements with the technology.

Fortunately, if one judges from the professional literature of the last two decades, the research focus seems to have shifted to a renewed emphasis on the human side of human-machine systems.[13] (John Olsgaard's excellent evaluative review provides references to numerous other examples.)[14] However, if librarians and their clients are to reap the benefits, more attention must be paid to continued research efforts in this area and to wider application of research findings by managers and practitioners.

CHANGING STRUCTURES: ORGANIZATIONAL, SOCIAL, AND PERSONAL

The rational problem-solving approach described above is partly based on the assumption that many library organizations will retain, to some degree, a traditional bureaucratic or hierarchical structure. This type of structure typically consists of an upper level of decision making, a midlevel of implementing, and a lower level of routine operations.

It is increasingly possible, however, that technology might bring about unexpected changes in traditional organizational structures. At the first national conference of the Library and Information Technology Association (Baltimore, 1983), a panel of librarians forecasted that automation, by making information available at almost any point in the organization (the nearest terminal), would eliminate the need for centralized, hierarchical control, which previously served the purpose of passing information to and from the decision makers at the top. Further, new role structures (e.g., decentralized groups of generalists or administrator-coordinators cutting across departmental and organizational barriers to address identified problems) might obviate the need for a pyramid of supervisory levels.[15] Library automation systems themselves have been proposed as personnel administration tools to break down traditional work flow and staffing patterns.[16] Surprisingly, as *Library Journal* reported, "No one in Baltimore spoke against the idea."[17]

Drawing an analogy between information providers and the producers of consumer products, Betty Turock acknowledged the possible equalizing effects of information technology on organizational structures:

> Flattening the hierarchy, superseding it with organizations configured in horizontal networks, has increased the opportunity not only for local decision-making. . . . [O]rganizations must willingly accept the notion of broadly- based participatory decision-making as a model for customer relations policies or they will be forced to. Organizations that have neglected marketing must begin at the consumer end to conceptualize products and services the consumer wants and not products the organization is equipped to offer.[18]

These types of changes in the work environment imply increased participation in decision making by professionals and support staff previously uninvolved in such activities. Regular cross-training at all levels has been suggested.[19] The manager implementing change through automation has to make sure not only that individual behavior patterns change, but also that alterations occur in organizational structure (e.g., communication patterns, evaluation practices, resource allocation) so that the new patterns of behavior are supported and reinforced.[20] Networking, characterized by interpersonal communications among individuals with common goals or beliefs, has been proposed by John Naisbitt, author of *Megatrends*, as an alternative to the traditional "pyramid" structure.[21] In a similar vein, library educator Barbara Conroy has suggested a set of guidelines for "people networks," a strategy by which librarians can discover their own capacities to initiate and assimilate change, as well realize the benefits of automation.[22]

THE HUMAN IMPACT

In studying resistance to technological innovation in libraries, Sara Fine found that what people fear most about technology is neither job obsolescence, nor system breakdown, nor its expense, but rather that interpersonal relationships will suffer as technology gets more sophisticated.[23]

Fine further indicates that the very attributes that attract people to technology also serve as powerful psychological stimuli to which the individual may respond in a dysfunctional way.[24] The increased speed and volume that automation offers may be "out of sync" with personal perceptions of space and internal timing. Similarly, attributes such as remoteness, rigidity, built-in obsolescence, foreignness, and specialized expertise affect the individual's sense of control over his or her environment, evoke ambivalent feelings and conflicting ideas, and provoke defensive behaviors, including displacement of negative feelings from people and events to the technology itself.

A frequent side effect of changes in internal and external organizational structures (e.g., networking, centralization or decentralization) is the individuals' perceived loss of control over their own performance and knowledge of their place in the organization. As a result, people can feel alienated within the workplace. Computerization may make the work more attractive and reach previously unsolvable problems. At the same time, it may lead to more stressful, rigid, mechanical working conditions, with the use of machinery that is remote and incomprehensible. Some research indicates that the loss of actual social contact with other human beings is experienced as dehumanizing, and this experience may trigger further inhumane behavior toward others.[25]

LEADERSHIP WITHIN THE HUMAN-MACHINE SYSTEM

Cohen and Cohen found that the majority of unsuccessful attempts at automation were characterized by sudden management action without warning, quick reductions in staff, little involvement in planning of change by those most affected, wrong equipment, poorly designed work spaces and work flow, inadequate training, and lower priority to individual work decisions, all of which can evoke the cognitive and affective responses described above.[26] On the other hand, successful human-machine systems are characterized by teamwork, interdependence, personal payoff from changes, and treatment of people as integral parts of the system.

Managerial qualities deemed essential in these kinds of organizations include not only knowledge and enthusiasm about automation, but also good communication skills, patience, commitment to training, sensitivity to staff needs, the ability to accept suggestions and to delegate responsibility—in other words, a "participative" style.[27]

The computer's potential for enhancing human capabilities is tremendous. Once the mesh is successfully achieved, why not capitalize on it by channeling the combined human-machine system's energy (or synergy) creatively to affect the organization positively in many other vital ways? Harnessing of this synergy could result in many other kinds of positive change, including individual development, increased competence, organizational growth, new and reasonable structures and policies, not to mention the possibility of real solutions to real problems. Rather than using automation as an excuse for more rigidity and bureaucratization, the organization could use it as a model and a tool for development of the institution into a more holistic, functional entity.

In these ideal organizational systems, people are recognized, treated as adults, given equal opportunities, options, and a voice in decision making and are aided by high-level staff members, with coordinating responsibilities and the ability to empathize with personal problems. Such leaders recognize that technological innovation involves solving not only technological problems but a host of human behavioral and organizational problems as well.[28]

James Driscoll has proposed that, in addition to their technical and administrative leadership functions, managers should pay closer attention to social and institutional leadership chores.[29] This suggested strategy for change includes support for new organizational structures (such as small work groups and task forces) and a nondirective approach that fosters a climate for innovations to become diffused.

As Richard Rowe and Richard Luce counsel, would-be leaders of information systems teams must actively embrace change; be clear about their missions; orient themselves toward values and goals (and not just means); create opportunities that require continual learning and risk taking; and expect excellence.[30]

This type of leadership constitutes what John Burch calls "the sociotechnical approach," within which the organization's operations are "looked at as a whole rather than in terms of isolated functions. . . . [S]pecial emphasis is given to the interrelationships of human (social) and nonhuman (technical) elements in the 'sociotechnical system.'"[31]

Achievement of the massive potential benefits of such systems, for ourselves and those we serve, will require an almost evolutionary leap of vision. Organizations will need to get beyond current levels of thinking about adversarial human-versus-machine relationships and, with emphasis on the human side of human-machine systems, begin looking outward and forward to the challenges the future holds in store.

CONCLUSION

The introduction and continued adaptation of automation to the library workplace has much to gain from research into human psychological and social values. While there are considerable data available in the areas of human engineering and ergonomics research, there is also a growing awareness of the need for related knowledge in the behavioral research areas that might be applied to the design of human-machine systems.

If the goal of automation is the solution of human problems, human beings are essential to its success. All human behavior is an attempt to cope with the environment as perceived. Human responses ought not to be used as an excuse for failure, but rather as information about specific problems. As Fine suggests, instead of labeling the human response to automation as a barrier, the manager of library automation might very well benefit from the valuable information and feedback it provides. "It seems much more constructive to call that which is good and to move from the strength we already have to the strength we are yet capable of."[32] The need for research in these areas is critical to the theory and practice of the library and information professions, now more than ever.

The new information technologies can be viewed as a mixed blessing. Automation can be used either to isolate us from our fellow human beings, to expand our interactions with one another globally, or, more realistically, somewhere in between. Decision makers face the challenge and responsibility of ensuring that human needs are met and human problems are solved in the course of changing traditional functions. Among our most valuable strengths as human beings in the "revolutionary nineties" is our capacity to plan flexible strategies for changing ourselves and our environment to meet our specific needs. Our success will ultimately be measured not only by what we have achieved, but by how we have cared for ourselves and one another in the process.

REFERENCES AND NOTES

1. Oscar Wilde, "The Soul of Man Under Socialism," in *The Soul of Man and Prison Writings*, ed. Isobel Murray (New York: Oxford Univ. Pr., 1990), 16.
2. Harry M. Kriz and Z. Kelly Queijo, "An Environmental Approach to Library Staff Training," *Library Hi Tech* 7 (1989): 62–66.
3. Mary Freedman and Larry Carlin, "The Computer as Scapegoat," *Library Journal* 110 (July 1985): 44–45.
4. Henry Ledgard, Andrew Singer, and John Whiteside, *Directions in Human Factors for Interactive Systems* (New York: Springer-Verlag, 1981), 1.
5. Robert Walton, "Automation," *Texas Library Journal* 59 (Summer 1983): 35–37.
6. F. Wilfrid Lancaster, *Libraries and Librarians in an Age of Electronics* (Arlington, Va.: Information Resources Press, 1982), 137.
7. Kenneth Dowlin, "I Am Not Willing to Destroy My Library in Order to Change It," *Library Association Record* 85 (Dec. 1983): 448–50.
8. Robert Theobald, "Future of the Librarian," *Public Libraries* 20 (Fall 1981): 74–76; S. D. Neill, "Libraries in the Year 2010: The Information Brokers," *Futurist* 15 (Oct. 1982): 47–51.
9. Niccolo Machiavelli, *The Prince and Other Works*, trans. Allan H. Gibbert (New York: Hendricks House, 1964), 110.
10. Robert D. Stueart, "Libraries: A New Role?" in *Books, Libraries and Electronics* (White Plains, N.Y.: Knowledge Industry, 1982), 93–116.
11. Nehemiah Jordan, "Allocation of Functions Between Man and Machines in Automated Systems," *Journal of Applied Psychology* 47 (1963): 161–65.
12. Ernest J. McCormick and Mark S. Sanders, *Human Factors in Engineering and Design*, 5th ed. (New York: McGraw-Hill, 1982), 489–92.
13. For notable examples, see Janice Kirkland, ed., "The Human Response to Library Automation," *Library Trends* 37 (Spring 1989): 385–542, a compilation of eleven scholarly articles on the topic; Leigh S. Estabrook, Chloe Bird, and Frederick L. Gilmore, "Job Satisfaction: Does Automation Make a Difference?" *Journal of Library Administration* 13 (1990): 175–94; David H. Waters, "The Effect of New Technology on Prestige, Self-Esteem and Social Relationships of University Library Employees," *LASIE* 20 (July–Aug. 1989): 16–22; Kristine Salamon Condic, "The Impact of CD-ROM on Reference Departments," *RQ*

28 (Winter 1988): 203–11; Mary Ellen C. Sievert, Rosie L. Albritton, and Paula Roper, "Investigating Computer Anxiety in an Academic Library," *Information Technology and Libraries* 7 (Sept. 1988): 243–52; Susan S. Steffen, "Living With and Managing Change: A Case Study of the Schaffner Library," *Illinois Libraries* 69 (Feb. 1987): 126–29; Alex A. Koohang and David M. Byrd, "A Study of Attitudes Toward the Usefulness of the Library Computer System and Selected Variables: A Further Study," *Library and Information Science Research* 9 (Apr. 1987): 105–11.

14. John N. Olsgaard, "Automation as a Socio-Organizational Agent of Change: An Evaluative Literature Review," *Information Technology and Libraries* 4 (Mar. 1985): 19–28.
15. Hugh C. Atkinson, "The Impact of New Technology on Library Organization, " in *Bowker Annual 1984* (New York: Bowker, 1984), 109–114.
16. Jo Ann Hanson, "A Model for Organization of the Automated Academic Library," *Technicalities* 2 (Oct. 1982): 7–8.
17. Karl E. Nyren and John N. Berry, "Information and Technology," *Library Journal* 108 (Nov. 1, 1983): 2020.
18. Betty J. Turock, "Technology and the Post-Industrial Society," *Catholic Library World* 55 (Feb. 1984): 298–304.
19. Rod Henshaw, "Library to Library," *Wilson Library Bulletin* 60 (Apr. 1986): 44–45; Stuart Glogoff, "Staff Training in the Automated Library Environment: A Symposium," *Library Hi Tech* 7 (1989): 61–62.
20. David G. Hopelain, "Assessing the Climate for Change: A Methodology for Managing Human Factors in a Computerized Information System Implementation," in *Human Factors in Computer Systems: Proceedings of the Symposium in Gaithersburg, Maryland, March 15–17, 1982*, (Institute for Computer Sciences and Technology and Association for Computing Machinery, Washington Chapter).
21. John Naisbitt, *Megatrends: Ten New Directions Transforming Our Lives* (New York: Warner, 1983), 199.
22. Barbara Conroy, "People Networks: A System for Library Change?" *Journal of Library Administration* 4 (Summer 1983): 75–87.
23. Sara Fine, *Resistance to Technological Innovation in Libraries, Final Report, Parts 1 and 2* (Washington, D.C.: U. S. Department of Health, Education and Welfare, Office of Education, Office of Libraries and Learning Resources, 1981); Sara Fine, "Human Factors and Human Consequences," in *Information Technology: Critical Choices for Library Decision-Makers*. ed. Allen Kent and Thomas J. Galvin (New York: Dekker, 1982), 209–24.
24. Sara Fine, "Technology and Libraries: A Behavioral Perspective," in *Communicating Information: Proceedings of the 43d ASIS Annual Meeting, vol. 17* (White Plains, N.Y.: Knowledge Industry, 1980), 246–49.
25. J. D. Orcutt and R. E. Anderson, "Social Interaction, Dehumanization and the Computerized Other," *Sociology and Social Research* 61 (Apr. 1977): 380–97.
26. Aaron Cohen and Elaine Cohen, *Automation, Space Management and Productivity* (New York: Bowker, 1982), 31.
27. Ganga Dakshinamurti, "Automation's Effect on Library Personnel," *Canadian Library Journal* 42 (Dec. 1985): 343–51.
28. Olsgaard, 25.
29. James W. Driscoll, "People and the Automated Office," *Datamation* 25 (Nov. 1979): 107–12.
30. Richard R. Rowe, "Managerial Competence for Library Automation," *Library Hi Tech* 7 (1989): 105–6: Richard E. Luce, "Developing Managerial Competence in Information Technology," *Library Hi Tech* 7 (1989): 107–9.
31. John L. Burch, "Computers, Humanistic Values and the Library," *ASIS Bulletin* 4 (June 1978): 29–31.
32. Sara Fine, "Technology and the Human Factor," in *Structure and Governance of Library Networks*, ed. Allen Kent and Thomas J. Galvin (New York: Dekker, 1979), 163.

On the Shoulders of Giants: From Boole to Shannon to Taube: The Origins and Development of Computerized Information from the Mid-19th Century to the Present

Elizabeth S. Smith

This article describes the evolvement of computerized information storage and retrieval, from its beginnings in the theoretical works on logic by George Boole in the mid-nineteenth century, to the application of Boole's logic to switching circuits by Claude Shannon in the late 1930s, and the development of coordinate indexing by Mortimer Taube in the late 1940s and early 1950s. Thus, electronic storage and retrieval of information, as we know it today, was the result of two major achievements: the advancement of computer technology initiated to a large extent by the work of Shannon, and the development of coordinate indexing and retrieval by the work of Taube. Both these achievements are based on and are the application of the theoretical works of George Boole.

The theoretical framework upon which electronic information is based was conceptualized some 150 years ago, in an 1854 work of a self-taught English mathematician, George Boole.[1] His work lay dormant until the advent of communication systems in the first part of the twentieth century, when, in 1938, it was found to be applicable to switching circuits and networks and, later, to information storage and retrieval. In that fashion, Boole's nineteenth-century classic became the cornerstone of computer sciences and information technologies of the twentieth century and the major work upon which future achievements in the field were based.

Boole's premature genius is not alone in the history of scientific progress. Some of the world's greatest discoveries and inventions were untimely at their inception. They were either unrecognized, unappreciated, or flatly refuted until some future date when the intellectual, scientific, and other environmental conditions became ripe and receptive to new ideas.

Such was the case, for example, of the Polish astronomer Nicolaus Copernicus, whose revolutionary theory of 1514, placing the sun at the center of the universe, faced great resistance from scientific and church authorities.[2] The theory was not fully accepted until some 150 years after Copernicus' death in 1543, when it paved the way for modern astronomy as well as advances in other physical sciences.

Or take the British physician William Harvey, whose work on the circulation of blood published in 1628 made him famous during his lifetime, but was not accepted by the medical establishment until after his death in 1657.[3]

Another genius was Gregor Mendel, whose experiments on inheritance traits in

Elizabeth S. Smith is Director of the Library, Penn State University Erie, The Behrend College, Erie, Pennsylvania.

garden peas generated no interest during his lifetime. His novel approach, utilizing statistics and mathematical formulas in explaining heredity, was neither understood nor given recognition by the scientific establishment. Mendel presented his findings in 1865 at two meetings of the Natural Science Society of Brno, located in what is now the Czech Republic. The minutes of the meetings indicate that although about forty botanists were in attendance, nobody asked Mendel any questions, nor did any discussion follow his presentation. His science contemporaries listened politely to his papers, but there is no indication that anyone really understood what he was talking about or appreciated the significance and the far-reaching effects of his studies. His work, though, was included in the *Catalogue of Scientific Literature* published in England in 1879, and scholars did make references to it occasionally.[4] It was not until 1909, when his studies were duplicated and confirmed by others, that he was given the title Father of Genetics. From that point on, the whole science of genetics and all future genetic discoveries had origins in his works.

And so we come to Isaac Newton, one of the greatest scientific geniuses, if not the greatest. Newton once wrote, "If I have seen further it is by standing on the shoulders of Giants."[5] That statement portrays the process of scientific progress and Newton's indebtedness to men such as Kepler, Aristotle, Pythagoras, Copernicus, and others.

GEORGE BOOLE AND BOOLEAN ALGEBRA

Do we have giants in library and information science, and, if so, who are they? Whose works have helped delineate the discipline, and whose shoulders have served to carry it forward? Without doubt one of the first and greatest is George Boole, whose achievements and contribution to logic created the foundation for the development of twentieth-century information technologies. Just as many before and after him, George Boole did not receive true recognition until after his death.

George Boole, who did not have the benefits of a formal high school or college education, was born in England in 1815. He was very gifted, with a talent and affinity for mathematics, which he studied on his own by reading the works of Newton and other great eighteenth- and nineteenth-century mathematicians. By the time he was twenty-four years old, he had begun submitting papers to mathematical journals. In 1844, at twenty-nine, he submitted an important paper to the Philosophical Transactions of the Royal Society of London in which he discussed the means by which algebra and calculus can be combined and applied to various disciplines, including logic. The Royal Society recognized his genius and awarded him a medal the same year. Three years later, in 1847, Boole wrote another important paper, entitled "Mathematical Analysis of Logic," in which he argued that logic and mathematics are related, and hence logic should be associated with mathematics, not metaphysics and philosophy, as had been the case thus far, dating back to Aristotle.[6]

In 1849, in spite of his lack of formal academic degrees, Boole was appointed professor of mathematics at Queens College, Cork, Ireland, an appointment based on his publications. In 1854 Boole published his most famous work *An Investigation into Laws of Thought, on Which Are Founded the Mathematical Theories of Logic and Probabilities*.[7] This classic, which became the foundation of communication, computer, and information sciences almost one hundred years later, was aimed at separating logic from philosophy and establishing logic, with the combination of newly invented algebra, as a separate science. The work contained detailed analyses of the processes of human reasoning and fundamental laws that govern operations of the mind.

Although Boole's appointment to professorship certainly demonstrated well-deserved recognition during his lifetime, the scope of applications that his invention would have was not evident until much later. By reforming and advancing the science of logic, he laid the cornerstone for computer science and sowed the seeds for future developments in information technologies.

What makes George Boole the information giant upon whose shoulders future scholars and scientists would stand, developing telephones, computers, new forms of indexing, as well as the processes, methods, and equipment that store and retrieve computerized information? What is Boolean algebra and Boolean logic, and how do they relate to information storage and retrieval? How did George Boole's work become the first in a series of steps leading to computer storage and retrieval of information? The rest of this paper attempts to answer these and other questions.

The greatness of Boole lay in his ability to analyze and describe the mechanics of human reasoning that he called the "operations of the human mind," the principles of which were later applied to manual, mechanical, and electronic retrieval of information.[8] To Boole, the processes of reasoning were either the addition of different concepts, or "classes of objects," to form more complex concepts, or the separation of complex concepts into individual simpler concepts.[9] Thus, any process of reasoning, any mental operation, was the addition or deletion of concepts. Nouns (which he called substantives), adjectives, and other parts of speech were classes of objects represented by symbols such as *X, Y, Z*.[10] Additions and deletions of concepts were represented by mathematical signs: plus (+), minus (–), times (x).[11] One of Boole's own illustrations of mental operation states:

> Thus, if x alone stands for "white things", and y for "sheep", let xy stand for "white sheep"; and in like manner, if z stand for "horned things", and x and y retain their previous interpretations, let zxy represent "horned white sheep" i.e. that collection of things to which the name "sheep" and the descriptions "white" and "horned" are together applicable.[12]

Thus, the instrument of reasoning to Boole was language, which was composed of signs that included words and mathematical symbols. The signs were utilized to depict the mental operations of the addition and deletion of concepts in the form of algebraic equations. Boole, who recognized the complexity of the internal intellectual processes of the mind occurring within complex cognitive structure, validated language as the essential criterion for reasoning and cognitive processes and then represented the processes by mathematical equations. He saw language not only as means of expressing and communicating ideas, wishes, or feelings, but as an instrument in thinking, reasoning, and formulating new ideas. He related nouns, verbs, adjectives, conjunctions, subjects, predicates, and descriptive phrases to the processes of reasoning and represented them by algebraic symbols.[13] Thus, Boole saw the working human mind as a set of mental operations that obeyed certain laws expressed by words and algebraic symbols.

But Boole's algebra differs in two fundamental ways from conventional algebra. Boole modified algebraic rules to fit his needs, devising a system that demonstrates and is compatible with his theory of human reasoning. The first way in which Boole's algebra differs from conventional algebra is that it is binary algebra. This means that there are only two possible values for any algebraic symbol, the values of 0 and 1.[14] Any expression, any symbol, in Boolean algebra can have only one of two values, either 0 or 1. Thus, if $A = 1$, then it cannot be 0, and if $A = 0$, then it cannot be 1. If $B = 0$, then it cannot be 1. Contrasting this with ordinary algebra, *A* and *B* can have an infinite number of possible values, from 0 to infinity. For example: $A = 4$, $A = 50$, $A = 100$, $A = -10$, $B = 4$, $B = 18$, $B = 1{,}000$, $B = -200$, and so on.

The second way in which Boole's algebra differs from ordinary algebra is in the meaning and use of the mathematical signs of +, –, and x. These signs in Boolean logic have the functions of combining and excluding concepts during reasoning, or so-called mental operations. One of Boole's mental operations is addition of simpler concepts to form more complex concepts. Boole describes this process as the "positive operation of aggregating parts into a whole" where each part or concept is different from the other, joined together by the conjunctions "or" or "and," both of which are compared to the plus sign in algebra.[15]

Boole further states:

> Thus the expression "men and women" is, conventional meanings set aside, equivalent with the expression "women and men." Let x represent "men", y, "women"; and let + stand for "and" and "or", then we have $x + y = y + x$, an equation which would equally hold true if x and y represented numbers, and + were the sign of arithmetical addition. Let the symbol z stand for the adjective "European", then since it is, in effect, the same thing to say "European men and women" as to say "European men and European women", we have $z(x + y) = zx + zy$. . . . The above are laws which govern the use of the sign +, here to denote the positive operation of aggregating parts into whole.[16]

How did Boole distinguish between the additions "OR and AND"? Consider each word separately: "OR" and "AND." The

Boolean plus sign is somewhat related to ordinary addition, but it means "or." So, Boolean $A + B$ is read as "*A* OR *B*," the OR being called the OR operator. However, one must keep in mind that the Boolean OR has an opposite meaning from the ordinary meaning of "or" in English speech. When used in ordinary speech, connecting words or groups of words, the conjunction "or" reduces the result. For example, when one says "lakes or rivers," it usually means one or the other, but not both; "or" reduces the sum by including only one option, either lakes or rivers, not both. But the Boolean OR denotes a different meaning. In Boolean logic when one says "lakes OR rivers," it means either lakes or rivers, or both, where OR expands and enlarges the sum by including either one or the other or both.

Boole's second mental operation reflecting addition of concepts is somewhat related to multiplication and represented by an asterisk (∘) or a dot (·), and it means "AND." Although it is analogous to addition, it differs from the Boolean OR addition. The statement A ∘ B is read as "A AND B," and the AND is called the AND operator.

Once again Boolean AND has an opposite meaning from ordinary English usage. In the English language the conjunction "and" increases the result when connecting words or groups of words. For example, "France and literature" usually means both concepts, France and literature, thus enlarging the sum. However in Boolean logic, although "France AND literature" also includes both concepts, doing so reduces the sum. How is that possible? The reduction occurs because combining two different concepts by Boolean AND, as in "France AND literature," results in the formation of a single new concept, "French literature." Putting it differently, by using AND, the concept of "French literature" is obtained, which requires that two other concepts be present, the concept of France and the concept of literature.

Boole referred to the notion of exclusion or separation to explain the differences between OR and AND. He asked the question: "Does the expression 'Scholars and men of the world' include or exclude those who are both?"[17] Thus, if x = scholars, and y = men of the world, then in the statement $x \circ y$, which reads "x AND y," the AND includes both concepts, where scholars are also men of the world. In contrast, the statement $x + y$, which reads "x OR y," shows each concept to be mutually exclusive, so that scholars may or may not be men of the world.

Boole gave numerous examples illustrating the formation of new and sometimes complex concepts using algebra and OR and AND during the process of mental operations. Both OR and AND perform addition operations, but their outcomes are different. OR expands the sum by including both concepts; AND reduces the sum by including both concepts.

The third Boolean operation is that of exclusion of concepts and is called complementation. It has no counterpart in ordinary algebra. Boole referred to it as the negative operation of separating "a part from a whole," and he used the term "except," expressed by a minus sign, to describe the operation.[18] Boole discusses it in the following way:

> Thus we cannot conceive it possible to collect parts into a whole, and not conceive it also possible to separate a part from a whole. This operation we express in common language by the sign *except*, as "All men *except* Asiatics," "All states *except* those which are monarchical". Here it is implied that the things excepted form a part of the things from which they are excepted. As we have expressed the operation of aggregation by the sign +, so we may express the negative operation above described by – minus. Thus if x be taken to represent men, and y, Asiatics, i.e. Asiatic men, then the conception of "All men except Asiatics" will be expressed by x – y. And if we represent by x, "states", and by y the descriptive property "having a monarchical form" then the conception of "All states except those which are monarchical" will be expressed by x – xy.[19]

The concept of complementation in Boolean algebra is represented by a line over an expression, and this means "not." For example, *A* bar ($\overline{A}$) means "NOT *A*"; *O* bar ($\overline{O}$) means "NOT *O*." Since in Boolean algebra, 0 and 1 are the only possible values for any symbol, the 0 and 1 are complements/opposites of each other. The NOT operator generates an inversion, or an opposite value, of the symbol in question. For example, if *O* bar ($\overline{O}$) means "NOT *0*," then it must be equal to 1, because 1 is the complement, the opposite of 0, and the only other value that it can have. Nothing else is possible. Similarly, if $A = 1$, then *A* bar ($\overline{A}$) = 0, and when $A = 0$, then *A* bar ($\overline{A}$) = 1.

In his work *The Laws of Thought* Boole

utilized the notion of "collection of things or classes of objects," often referred to as sets.[21] From that evolved the algebra of classes, or algebra of sets, the principles of which are utilized in computer retrieval of information.

A set or class is a collection of objects or ideas whose individual members have some common characteristics. Within the Boolean framework there are different kinds of sets. For example, finite sets have a definite number of members—say, all the months of a year. Infinite sets have an infinite number of members—for example, all the odd numbers, to which, of course, there is no end. There are also overlapping sets whose members may belong to more than one set—for example, two different organizations having members common to both. Disjointed sets are sets that have no members in common. Equivalent sets are sets whose members can be matched one for one, and equal sets have the same members. Boolean algebra of sets or classes developed into set theory, which is used in solving problems in mathematics and logic.

Algebra of sets is the basis of logic that has been used in developing search strategies for online and CD-ROM searching and is reflected in the command language and syntax of communications software used in Boolean retrieval of records. The Boolean operators OR, AND, and NOT combine sets, forming new sets according to Boolean rules of logic.

In computer searching the OR represents the union of sets, or the addition of sets, shown by the symbol +. It is defined as a set that includes all members of set *A*, and all members of set *B*, and all those present in both *A* and *B*, without repeating any members. For example, set *A* with two members representing art periods "Renaissance, Baroque" is combined with set *B*, consisting also of two members "Classical, Modern." The operation will read "*A* OR *B*" and will include "Renaissance, Baroque, Classical, Modern," thus containing all members of both sets. However, in the union of sets in which some members of one set also belong to another set, so-called overlapping sets, the total number of members is fewer than the sum of the members of both sets. If in the previous example each set also contained a third member, "Romanesque," the union of overlapping sets would yield only five members "Renaissance, Baroque, Classical, Modern, and Romanesque," because the latter would be listed only once. This principle of reduction is characterized in computer searching using OR, when the retrieval of citations containing members of two sets is smaller than the sum of both sets. It occurs because some of the citations contain members that are common to both sets and are only counted once.

The AND operator represents the "intersection (meet) of sets" and is shown by the symbol •. The intersection of sets is defined as a set that includes only members that belong to both sets. For example, if set *A* is "red, blue, yellow," and set *B*'s members are "blue, green, violet," the intersection of *A* AND *B* is the set "blue," because blue belongs to both *A* and *B*. Similarly, the intersection of set *X*, consisting of members "roses, violets, tulips," and set *Y*, consisting of "roses, daisies, violets," results in the set "roses and violets," because these members belong to both groups. In computer searching this principle is characterized by the requirement that all selected concepts joined by AND are present in the citations retrieved.

The third Boolean operation is reflected in the use of NOT, which represents the exclusion or the "complement of a set," shown by the symbol –. For example, set *A* consists of the numbers 1 to 10, and subset *X* consists of even numbers within that set, 2, 4, 6, 8, 10. The set of odd numbers within set *A*—or 1, 3, 5, 7, 9—is called the complement of subset *X*, written as *X*1. Members 1, 3, 5, 7, 9 belong to *A* but not to *X*. Therefore, the complement of subset *X* is the set 1, 3, 5, 7, 9. Translating this to computer searching, the NOT operator removes some concepts from the search.

Boole also discussed propositions and how they relate to the study of logic and reasoning. Later, this came to be known as the algebra of propositions. Propositions are statements that may be either true or false, but not both. Whitesitt explains:

> The following examples are typical propositions:
>
> 3 is a prime number;
> when 5 is added to 4, the sum is 7;
> living creatures exist on the planet Venus.
>
> Note that of these three propositions, the first is known to be true, the second is known to be false, and the third is either true or false (not both), although our knowledge is not sufficient to decide at present which is the

case. Contrast this with the following sentence, which is not a proposition:

This statement you are reading is false. If we assume that the statement is true, then from its content we infer that it is false. On the other hand, if the statement is assumed to be false, then from its content we infer that it is true. Therefore this statement fails to satisfy our requirements and is not a proposition.[21]

A true proposition within Boolean meaning has the value of 1; false proposition has the value of 0. Therefore, the first is a true proposition and defined as 1. The second statement is a false proposition, defined as 0.

Propositions can be combined using AND, OR, or NOT to form new compound propositions. Algebra of propositions is concerned with determining truth values of compound statements from truth values of their constituent simple statements. When AND is used to connect two propositions, the new compound proposition is true only if both propositions are true, and false if either one or both propositions are false. In connecting two propositions with OR, the new proposition is true if either one or the other or both are true, and false if both are false. The combinations of various propositions can be depicted by constructing truth tables that provide truth values of single and compound propositions, using the digits 1 and 0 for "true" and "false." Boolean algebra of propositions and the notion of true and false propositions and dichotomous (binary) classification—such as man, woman; yes, no; present, absent—are reflected in indexing, classification, and storage and retrieval of information and will be explained shortly.

CLAUDE SHANNON AND SWITCHING CIRCUITS

Although George Boole gained scholarly recognition among his contemporaries, his work lay dormant with no practical applications for some eighty-four years. In 1938, Claude E. Shannon, a young research assistant at the Massachusetts Institute of Technology, noticed the similarity between Boolean algebra and telephone switching circuits. Circuits are paths of electrical currents activated by inputs of the circuit that can be depicted in diagram form. Circuit diagrams show switches, inputs, outputs, and gates, which can be closed or opened, allowing or preventing current to flow.

The electric circuits are called bistable, or two-state, devices because they have only two signal states, the open state and the closed state. When the circuit is open, the power is off, and when the circuit is closed, the power is on. Shannon noticed that these states can be easily expressed in Boolean two-value binary algebra by 1 and 0, so that 1 means "on" when the switch is closed and the power is on, and 0 means "off" when the switch is open and power is off. Thus, the presence or absence of power can be expressed by the value of 1 (power on) or 0 (off).

Shannon presented his findings in a paper entitled "A Symbolic Analysis of Relay and Switching Circuits" (an abstract of his master of science thesis), at the American Institute of Electrical Engineers (AIEE) Summer Convention, Washington, D.C., in June 1938.[22] In his paper he explained the usefulness of Boolean algebra in indicating mathematically the logical functions of electrical circuits in automatic telephone exchanges, as well as most other circuits that perform automatic operations. This important paper became the focal point for the emerging computer technologies, without which information storage and retrieval as we know it today would not have been possible.

The application of Boolean principles to switching circuits came to be known as the algebra of circuits, or switching algebra, and involves any two-state (bistable) devices, the simplest of which is the switch. The two-value off-on system is present in telephone switching circuits, electron tubes, electronic computers, transistors, magnetic cores, gas and vacuum valves, semiconductors, and other automatic switching devices and networks that are designed on Boolean algebra. The nature of the binary system depends on the device in question, e.g. closed versus open circuit, charged versus discharged, magnetized versus nonmagnetized, or conducting versus nonconducting.

When circuits are designed, Boolean expression is formulated that mathematically expresses the logical functions required in a circuit. For example, the Boolean expression $A \bullet B = C$, which reads "A AND $B = C$," represents the logical function of an AND circuit, where switch A and switch B must both be closed for the power to be on and the light, represented by C, to be lit. When switches A and B are closed, each has a value of 1, and when light C is on, it also has a value of 1. In Boolean terms, since there are only two values, 1 and 0, for any variable, any variable in $A \circ B = C$ (A AND $B = C$) can have only one of two values, either 1 or 0. Thus, there are four possible combinations for the expression $A \circ B = C$. They are: $0 \circ 0 = 0$; $0 \circ 1 = 0$; $1 \circ 0 = 0$; $1 \circ 1 = 1$. Out of these four, only the last example ($1 \circ 1 = 1$) enables light C to glow.[23]

The OR circuit can be exemplified by the expression $A + B = C$, which reads "A OR $B = C$." This expression represents the logical function of the OR circuit, where light C will shine, represented by 1, if either switch A or B is closed, also represented by 1. Open positions of switches A and B are represented by 0 when the power is off. Light C is also 0 if it is off. Since Boolean algebra dictates that any variable may have only one of two values, 1 or 0, there are four possible combinations for the expression $A + B = C$. They are: $0 + 0 = 0$; $0 + 1 = 1$; $1 + 0 = 1$; $1 + 1 = 1$. Out of these four, the last three enable light C to glow, because if one or all the variables are 1, the final result is also 1, permitting the light to shine.[24]

In OR circuits, the OR function is performed by OR gate, activated when any or all of the inputs are present; AND circuits, where AND function is performed by the AND gate, are activated only when all indicated inputs are present; NOT circuits are activated when NOT function is performed by the NOT gate, expressing an inversion, where if X is absent, then NOT is present. Other combinations, such as AND-OR and OR-AND circuits, also exist.

When the Boolean expression has been derived, a diagram depicting the required logical functions of a circuit may be drawn. Electrical circuits are then built to accomplish the needed functions. Conversely, from a logic diagram that shows inputs, gates, and outputs, one can write a Boolean expression of logical functions of a particular circuit. Thus, electronic computers used in information storage and retrieval are built on the principles of Boole's binary algebra, 0-1, off-on system.

MORTIMER TAUBE AND COORDINATE INDEXING

How is Boolean algebra reflected in the storage and retrieval of information? The application of Boolean principles to indexing and storing information goes back to the late 1940s and early 1950s, when scientists addressed the acute problems of indexing and retrieving scientific literature. These problems were the direct outcome of tremendous increases in research during and after World War II, which generated massive volumes of scientific literature for which the existing indexing and retrieval methods were inadequate. New disciplines, new technologies, and new terminology evolved that did not fit into the existing hierarchical subject-heading classification schemes. The existing periodicals indexes, which were either alphabetical or classified, proved inadequate to accommodate the broad and complex newly emerging knowledge. It became evident that new methods of indexing and retrieving information had to be developed, including machines that would store and search for information.

The need was fulfilled by the work of Mortimer Taube and his associates of Documentation Incorporated. Under contract to the United States Armed Services Technical Information Agency, Taube began to develop new methods of storing and retrieving information. He introduced what he called a "functional approach to bibliographic organization."[25] This new approach to obtaining information was based on the notion that all disciplines can be broken down into basic concepts, or ideas, represented by simple index terms that alone or in combination (that is functionally) would express the contents and concepts of the discipline. During a search, a person would first select index terms that describe the concepts about which information is being sought. The user would then combine the index terms and obtain a search statement representing the desired information.

Taube's work culminated with the invention of a new method of indexing that came to be known as coordinate indexing and later evolved into the theory of coordinate indexing. It is a general theory describing a method of organizing categories of information into index terms. The rationale was that if disciplines can be broken down into single ideas with assigned terms, then machines can be used to organize and search for information in the field. This led to the development of various machines that performed all aspects of Boolean searching.

The term "coordinate indexing" was first used by Mortimer Taube in a paper read by him on September 4, 1951, before the Symposium on Mechanical Aids to Chemical Documentation of the Division of Chemical Literature of the American Chemical Society.[26] It was called coordinate indexing because the person looking for information had to combine or coordinate the index terms that were selected, either free or from a thesaurus, to describe the desired information.

New indexes that developed as a result of coordination came to be known as coordinate indexes, and the process of searching these indexes by combining ideas to find wanted information was called concept coordination. Taube called it bibliographic coordination. In time, index terms acquired a variety of names, such as "keywords," "descriptors," or "locators." Taube called them uniterms.

Thus, coordinate indexing refers to the intersection or combination by *and* of two or more coordinates, sets, or classes, which are bits of information, ideas, or concepts having preassigned index terms. This is based on Boolean algebra of sets and classes and its operations, when sets are joined (the union) with OR or intersected (the meet) by AND and excluded (the complement) by NOT to create more or less complex classes. In coordinate indexing the index terms are equal in weight, in contrast with subject or classified indexes, which are hierarchical. In coordinate retrieval, the index terms are brought into common action, joined (OR) or intersected (AND), the result being a new class that reflects the desired information.

How does coordinate indexing relate to Boole's algebra and logic? Since coordinate indexing is based on the principle that pieces of information are represented by index terms (descriptors) that express the concepts and thus the contents of the information, each significant word in a document is indexed as a descriptor. Therefore, any descriptor either is or is not present or assigned to a document. In other words, there are only two possibilities, the presence of a descriptor in a document or the absence of a descriptor in a document. It is a binary system of evaluation—true-false, on-off—which uses the principles of Boolean algebra. The person seeking information selects the appropriate descriptors that represent the desired information, and only those documents that contain the requested descriptors are retrieved.

The invention of organizing categories of information into index terms (descriptors) and the rationale that any discipline can be broken down into single ideas with assigned terms led to the development of various systems of storing and searching for information. Basically, two types of coordinate indexes evolved as a result of Taube's work. One is the term index, and the other is the item index. In the term index, combination of search terms is achieved by handling specific terms. In the item index, combination of search terms is achieved by handling specific documents.

Taube developed a manual system of term indexing and selecting documents by index terms that he called the uniterm system of coordinate indexing.[28] The procedure involved the coordination or comparison of cards with specific index terms on them. In this system a card representing an index term (not a document) was divided into ten columns headed by digits from 0 to 9. Documents, some of which were assigned as many as twenty index terms, were given accession numbers and recorded on the appropriate cards bearing the index term. The accession numbers were placed in the column representing the last digit on the document number. For example, document number 408 on the subject of thermodynamics was recorded in the eighth column on the index card bearing the index term "thermodynamics." Because documents were arranged by their last digits, it was relatively easy to compare several cards with different index terms for numbers common to all. Finding the same document numbers on different cards indicated that the index terms on the cards were present in the document.

For example, if document number 408 appeared on a card with the index term "thermodynamics" as well as on a card with the index term "gases," it indicated that document 408 contained information on thermodynamics as well as on gases, implying that the topic of thermodynamics of gases was being discussed. Thus the Boolean principle of the intersection (meet) of sets is in operation, retrieving information on the single concept of thermodynamics of gases, created by the combination of two separate concepts, the concept of thermodynamics and the concept of gases.

The best-known example of an item coordinate index was Edge Notch Cards, also called McBee Cards, named after a primary vendor. Here, a card represented a document instead of an index term. A card with a bibliographic citation of a document was assigned a document number and had index terms coded around the periphery of the card by notched-out holes. The index terms represented ideas present within the document. During a search, documents with the desired information were combined by notches and separated from others usually through some mechanical means, such as needles being inserted through the holes, representing the terms to be searched. For example, a document on corrosion of metals had as one of its index terms "iron," represented by notch number 5, and another term, "corrosion," represented by notch number 19. By inserting needles through locations 5 and 19 in a collection of cards, one could separate all cards with notched-out holes 5 and 19 from those whose holes 5 and 19 were not open with a notch. The documents with holes 5 and 19 indicated the presence of the index terms "iron" and "corrosion" in these documents.

Since its inception in the 1950s, coordinate indexing has been utilized to a large extent in mechanical, printed, and electronic forms. Mechanized indexing and retrieval of information was a natural and intentional outcome of coordinate indexing that lent its versatility and adaptability.

In the late 1940s through the 1960s, many varieties of item and term indexes and methods of searching were developed and became popular, especially in industry's special libraries. The first attempts at mechanization were keypunches for tabulating cards and mechanized rotary files for filing index cards. The combination of terms was accomplished by IBM adding machines, Rapid Selectors, and Univacs.[28] These machines were sensitive to punched holes, dots, or magnetic impulses representing concepts or bits of information. Electronic statistical machines, such as the IBM 101, were invented to select keywords from documents. Other machines were invented to prepare appropriate index cards for each keyword. Still others stored, assembled, filed, keypunched, and searched. All of them operated on the principles of Boolean algebra, by which sets are joined (the union) with OR, or intersected (the meet) by AND, and excluded (the complement) by NOT, to retrieve desired information. Later, electronic computers placed these machines on a path to extinction.

As coordinate indexing evolved so did the publication of new printed indexes, of which *Science Citation Index* and *Social Science Citation Index* are examples. They incorporate the principles of coordination in their Permuterm Subject Indexes. Biological Abstracts introduced a coordinate subject index in the 1960s, as did the *Monthly Catalog* of the U.S. government and others. Several variations of coordinate indexes evolved. They were KWIT (keywords in title), KWIC (keywords in context), or KWOC (keywords out of context) indexes, also known as title indexes. Examples include chemical titles and geo abstracts.

The mechanization of storage and retrieval of information was followed in the 1960s by the introduction of computers. Computers in information retrieval were first utilized by the U.S. government, especially the National Library of Medicine, which initiated the first online bibliographic searches conducted by medical librarians in late 1960s. In 1972, Dialog, the first commercial online information-retrieval system, was established, soon followed by others whose benefits were quickly utilized by special and industrial libraries. Later, universities, colleges, and public libraries became their users. The 1980s saw the development of CD-ROM technology, which also utilizes the principles of Boolean binary algebra and logic in storing and retrieving information.

SUMMARY AND CONCLUSION

Electronic storage and retrieval of information was the result of two major achievements: the advancement of computer technology, initiated to a large extent by the works of an American scientist, Claude Shannon, in 1938, and the development of the principles of coordinate indexing and retrieval by an American team headed by Mortimer Taube in 1951. Both of these achievements are based on, and are the application of, the theoretical work of an Englishman, George Boole, whose writings date back to 1854.

Most of the information storage and retrieval technologies are based on Boolean binary algebra and logic and utilize the machine language of 0 and 1, and the operators OR, AND, and NOT. The design of computers, specifically the digital-computer circuitry, is based on the on-and-off (1 and 0), two-value system (binary algebra) invented by Boole, which combines in various ways, performing logic and arithmetic applications. The switches that have only two states, on and off (1 and 0) indicate the presence or absence of electricity by which the computer operates. Boole's three operators OR, AND, and NOT are used to indicate logic circuits in digital computers that perform logical functions, so that the AND gate performs the AND function; the OR gate, the OR function; and the NOT gate, the NOT function.

The process of online searching and computerized information storage is also based on Boolean logic. Operators OR, AND, and NOT are used to combine descriptors, creating search statements that are input into computers that store information on magnetic disks in large mainframe computers. The descriptors that are represented by magnetic spots on a disk are read by the computer, identified as present (true) or absent (false) and related to the binary system 1 and 0, respectively. If a descriptor is present the document will be retrieved; if it is absent the document will not be retrieved. Thus, the computer translates the notion of presence or absence of a descriptor to 1 and 0 respectively when it reads the magnetic spots that represent the descriptors on a disk. The CD-ROM technology that emerged in the mid-1980s utilizes the same principles of Boolean logic in its storage and retrieval.

REFERENCES

1. George Boole, *An Investigation into Laws of Thought, on Which are founded The Mathematical Theories of Logic and Probabilities* (London: Walton and Maberley, 1854).
2. Angus Armitage, *The World of Copernicus* (New York: Signet Science Library Books, 1951).
3. Louis Chavois, *William Harvey: His Life and Times: His Discoveries: His Methods* (London: Hutchinson Medical Publications, 1957).
4. Vitezslav Orel, *Mendel*, trans. Stephen Finn (Oxford: Oxford Univ. Pr., 1984).
5. Frank E. Manuel, *Portrait of Isaac Newton* (Cambridge, Mass.: Belknap, 1968), 145.
6. George Boole, *Mathematical Analysis of Logic, Being An Essay Towards a Calculus of Deductive Reasoning* (Cambridge, Mass.: MacMillan, Barclay, and MacMillan, 1847).
7. Boole, *An Investigation into Laws of Thought*.
8. Ibid., 43.
9. Ibid., 34, 36.
10. Ibid., 30, 32.
11. Ibid., 29.
12. Ibid., 31.
13. Ibid., 32, 64.
14. Ibid., 41.
15. Ibid., 36.
16. Ibid.
17. Ibid., 61.
18. Ibid., 36.
19. Ibid.
20. Ibid., 31, 34.
21. Eldon J. Whitesitt, *Boolean Algebra and Its Applications* (Reading, Mass.: Addison-Wesley, 1961), 41.
22. Claude Shannon, "A Symbolic Analysis of Relay and Switching Circuits," *American Institute of Electrical Engineers Transactions* 57 (1938), 713–23.
23. Monica L. Noell, *Fundamentals of Digital Logic*, ED259902 (Washington, D.C.: Marine Corps Institute, 1984).
24. Ibid., 33.
25. Mortimer Taube, "Coordinating Indexing of Scientific Fields" (Paper delivered at the Symposium on Mechanical Aids to Chemical Documentation, Division of Chemical Literature, American Chemical Society, New York, September 4, 1951).
26. Ibid.
27. Mortimer Taube and Associates, *Studies in Coordinate Indexing* (Washington, D.C.: 1953).
28. Mortimer Taube, "Functional Approach to Bibliographic Organization: A Critique and a Proposal," in *Bibliographic Organization*, ed. Jesse Shera and Margaret E. Egan (Chicago: Univ. of Chicago Pr., 1951), 68. ■■

Special Section: Z39.50—Two Perspectives

RLG's Z39.50 Server: Development and Implementation Issues

Lennie Stovel, Rich Fuchs, and Jui-wen Chang

In 1991 and 1992, the Research Libraries Group (RLG) developed a server that implements the main features of ANSI/NISO standard Z39.50-1992, Information Retrieval Service and Protocol. The Z39.50 server allows users of other computers that run Z39.50 client programs to search RLG's bibliographic, authority, and citation databases. This service will be generally available in mid-1993.

In any development project, some of the issues that the development team has to consider are the operating environment of the developed system; the development tools the team will use; the external design of the system, or how it will look to users; and the internal design, or how the programs will look. Then, of course, the coding, testing, and implementing need to be done.

OPERATING ENVIRONMENT

In determining the operating environment for the server, we considered putting it on a separate computer, distinct from the mainframe. Use of the UNIX operating system on such a computer might have given us some advantages in flexibility of network connections and in maturity of network software. It also would take some of the processing load off the mainframe. In the end, we decided to run the server on RLG's Amdahl mainframe, where the databases are also housed, to avoid running a private protocol between the intermediate system and the mainframe. This also provides a smaller system that is more easily maintained because it lives on a single machine.

Initially RLG planned to run the server over a protocol stack that conforms to the ISO Open Systems Interconnection model for communication between disparate computer systems. The Z39.50 standard was written to fit within this model as an application-layer protocol. RLG has several years of experience with the Linked System Project, running information-retrieval and record-transfer protocols over a stack that conforms to earlier versions of the ISO standards, and we planned to upgrade those programs to the current standards. During the time we were developing the server, however, other institutions began to express more interest in running Z39.50 over the Transmission Control Protocol/Internet Protocol (TCP/IP) suite. Since many institutions, including RLG, were expanding their use of the Internet, we decided that if our server was going to be used, it should also run over TCP/IP. We learned that the mainframe-based TCP/IP support within the Stanford Timesharing System used by RLG could support the multiuser environment we anticipated within the context of its session-handling and time-sharing systems. We then switched our focus to the TCP/IP environment. The agreements reached by Z39.50 implementors in the Coalition for Networked Information's Z39.50 Implementors Testbed (ZIT) specify the use of only one function of the protocol layers between the application layer and the transport layer: the use at the presentation-layer level of the Basic Encoding Rules (BER) found in ISO 8825 for carrying the application protocol data units (APDUs) defined in Z39.50.

DEVELOPMENT TOOLS

Several developers of Z39.50 clients and servers began by using the ISO Development

Lennie Stovel is Manager, Intersystem Applications, Research Libraries Group. Rich Fuchs designed RLG's Z39.50 server and coded the Protocol Machine and the BER encoding and decoding routines. Jui-wen Chang coded the search engine.

Environment (ISODE), a publicly available set of programs and accompanying documentation that implements the upper layers of the ISO communications protocols over TCP/IP for UNIX. Generally, they found this software to be unwieldy and difficult to use. With the test-bed group's agreements, only the BER encoding and decoding routines proved useful. As their implementations progressed, OCLC, the University of California at Berkeley, and Stanford University provided code that has been widely used by others implementing under UNIX.

Having decided to operate the server on the mainframe, RLG chose to use its normal language, Pascal/VS, for development. We looked for commercial ASN.1 compilers—programs that take the abstract syntax definition of the APDUs as input and generate the data structures and the encoding—but none was available for Pascal. So we wrote our own encoding and decoding routines, as did several other developers. These routines are specific to the Z39.50 APDUs and are not as generalized as a compiler would be, but it is not hard to add new PDUs. In addition, our routines are more optimized for speed than compiler-generated code can be.

Similarly, because we were writing in mainframe Pascal, we could not make use of the development efforts of other implementors who were using UNIX. Other institutions doing mainframe implementations chose to use the IBM OSI Communications Subsystem. Because of the differences in computing environments, these implementors have not been able to share code.

Historical precedents for servers exist at RLG, in the LSP context and others; while the Z39.50 server code grew out of this experience, some of the precedents were discarded. The magnitude of the changes between the 1988 and 1992 versions of the Z39.50 standard also provided motivation for writing some new code.

EXTERNAL DESIGN

In a client-server environment, the external design is specified by the protocol used between the client and the server, in this case, Z39.50. The interaction of protocol data units is specified in state tables in the standard. The state tables define, for any given state of the server, which PDUs can be expected from the client or generated by the server. Implementing the protocol is largely a matter of selecting among the options defined by the protocol and determining such things as message sizes and diagnostics.

One major area of effort for Z39.50 implementors is mapping the attributes that specify the parameters of a search to and from the underlying database management software and indexing schemes. Initially RLG took a very broad approach: we tried to support each attribute as well as we could, in some cases using capabilities of our DBMS that are not even used by our mainstream service, the Research Libraries Information Network (RLIN). We are now rethinking that approach, as some of the searching is inefficient and leads to unacceptable response times.

One implementation difficulty caused by the decision to run over TCP/IP was the lack of a clear definition of how to terminate a client-server session. The standard relies on termination services provided by another application-layer standard, the Association Control Service Element. The implementors ultimately decided to use a TCP close to shut down sessions.

INTERNAL DESIGN

Internally, the RLG Z39.50 server has two major modules, a protocol machine and a search engine. This design is fairly common among the implementors. The protocol machine receives and generates the protocol data units. It analyzes incoming PDUs and passes the necessary values to the search engine; it receives information about completed searches from the search engine and composes the response PDUs. The search engine communicates with the DBMS, which actually carries out the searches and retrieves the records. The search engine also communicates with existing routines called Import/Export Services, which convert records from the internal DBMS format to the USMARC format. The protocol machine communicates with the BER encoding and decoding routines, which in turn work with a third set of modules called Communications Services, whose routines "talk" to the mainframe TCP/IP support.

The internal design separates the protocol machine and the search engine from each other; they communicate by messages that could easily be serialized in the same way that

PDUs are serialized. Because of this design decision, it would be possible in the future to port the protocol-machine code to a different system to achieve the advantages of a more robust TCP environment, decreased mainframe processing load, and easier scalability as use grows. Thus the internal design reflects some of the early considerations about the operating environment.

As much information as possible is maintained outside the server code. This includes items such as PDU parameters (message sizes, element set names, version numbers, and options available), record conversion tables, database names and aliases, and, most important, the mapping from the use attributes defined in the standard to RLIN indexes. This design decision enables us to make several kinds of changes in the way the server operates without having to recode or recompile any programs. It facilitates prototyping of new capabilities as well as maintenance.

TESTING

Because the server combined existing components of its operating environment in new ways, testing was a lengthy process of tracking down problems in several different pieces of code, some of it not under RLG's control. This caused significant delays. The test environment included a Z39.50 client that enables testers to control elements of the PDUs that would normally be masked to the end user by a well-designed interface. Thus a tester can generate error conditions for the server to handle. There was a fringe benefit of this work: We have been able to use this client to demonstrate what goes on behind an interface, which clarifies the scope and intent of the Z39.50 protocol.

Following internal testing, RLG opened up the server to testing by members of the test-bed group. By and large their client programs were able to communicate with our server successfully. Many of the problems that did arise revolved around differing interpretations of the Basic Encoding Rules. Also there is a tendency among implementors to read only the ASN.1 definitions of the PDU parameters in the standard and to ignore the textual description of the parameters, which frequently contains guidance for applying the parameters.

Initially the RLG server provided access only to test versions of our database. When we provided access to the full database available to the test-bed group, implementors of client programs encountered new issues in accessing very large databases. It is easy for a client to generate a general search, and some seemingly simple searches retrieve large result sets. Managing these result sets on behalf of the user requires care in programming the client.

An area that still awaits resolution by the community of implementors concerns the mapping of queries, both from native searching languages to the intersite query and the bib-1 use attributes, and from the intersite queries to the searching languages of the database maintainers. Because systems have indexed similar types of data differently, a search entered on one system may not yield the expected results when carried out on another system. Resolving these issues will require a great deal of communication between librarians and other expert users of the systems and system implementors.

IMPLEMENTATION

Since July 1992, we have had literally thousands of connection attempts to our server from perhaps fifty different client addresses. Among the test-bed group, between fifteen and twenty institutions have client programs that have connected successfully and regularly. Many of the connection attempts have been from Wide Area Information Server (WAIS) clients; to date, WAIS supports only the 1988 version of Z39.50, which is not compatible with the version supported by the RLG server.

Users at Pennsylvania State University can search RLIN files using the Z39.50 client in LIAS (see accompanying article). This is the first instance of use of RLG's server by people who are not library staff.

OTHER DEVELOPMENTS

Also in 1992, RLG developed a patron-oriented search service called Eureka. Eureka lets users search RLIN databases without training or documentation. Eureka is implemented as a Z39.50 client that runs on RLG's mainframe, communicating with the server running on the same machine. Additional indexes that RLG created for Eureka will be made available to Z39.50 clients as well. We implemented for Eureka some services that are still under development for

Z39.50, notably Scan (index browsing) and Sort. This system was developed on personal computers, though it runs on a mainframe. We are in the process of porting the Eureka client software to run under UNIX. When this is complete, the client could run on a computer at RLG or at a remote location.

CONCLUSION

While RLG developed its Z39.50 server over a period of two years, there is now experience—and some code—available that will shorten the development time for future implementors. As is frequently the case with development projects, circumstances changed during the development that made us alter some of our initial decisions, and the server is still evolving. Version 3 of the standard was approved in 1992, and we will be upgrading the server to support it. However, the Z39.50 vision of being able to use a single interface to search databases housed on multiple remote systems is becoming a reality, and that is very gratifying. ■■

Z39.50 and LIAS: Penn State's Experience

Sylvia MacKinnon Carson
and Dace I. Freivalds

Pennsylvania State University's experience in developing and implementing Z39.50 into LIAS is explored in this article. The emphasis is on development and implementation issues from the client's point of view, including mapping commands into attributes, cases not covered by Z39.50, processing diagnostics, and connecting to multiple servers. The article wraps up by discussing Penn State's current and anticipated applications for Z39.50.

PENN STATE'S LIBRARY COMPUTING ENVIRONMENT

LIAS (Library Information Access System) has been developed in-house at Penn State since the mid-1970s. It has served as the cataloging facility for the Penn State Libraries since January 1981, and the LIAS online catalog has been available to the public since 1983. LIAS supports all activities associated with cataloging: preorder searching, record input and editing, importing records into LIAS from USMARC or OCLC, online shelflisting, and creation of spine and book labels. LIAS also supports a circulation-control system, the fund-accounting component of acquisitions processing, and management statistics programs. In 1992, LIAS began to offer access to journal article databases such as UnCover and ERIC. UnCover, which indexes more than 10,000 scholarly and popular journals, resides on a mainframe in Colorado; LIAS provides transparent telecommunications access to this service. ERIC, which is produced by the U.S. Department of Education, provides access to more than 750 journals and a vast education collection on microfiche. ERIC is locally mounted in LIAS and searchable with the same commands used in the LIAS online catalog. Currently, a fully automated acquisitions component is under development and is anticipated to be operational in late 1993.

Penn State's LIAS runs on a DEC VAX 10,000 with one gigabyte main memory and seventy gigabytes of disk storage. At the present time it supports more than 600 PCs, Macintosh computers, and dumb terminals.

LIAS serves the entire Penn State Libraries system, including the central campus at University Park, as well as more than twenty branch campuses throughout Pennsylvania. A typical week's activity is about one million transactions.

Z39.50 IMPLEMENTATION HISTORY AT PENN STATE

In 1989, Penn State and the University of California received grants from Digital Equipment Corporation (DEC) to support linking the library systems at these institutions via the Z39.50 protocol. Figure 1 outlines the

---

Sylvia MacKinnon Carson and Dace I. Freivalds are Library Systems Specialists, Library Computing Services, Pennsylvania State University. The authors wish to thank Penn State's Z39.50 gurus (Eric Ferrin, Tun Chin, and Lee Newman) for their assistance in the preparation of this paper.

Z39.50 Client/Server Development at Pennsylvania State University

	1990
	Meetings between Library Computing Services, Penn State and the Division of Library Automation, University of California to discuss various ways to implement Z39.50 and to ensure interoperability between the two implementations. Discussions also took place with other institutions interested in developing Z39.50 clients and servers with the formation of the Z39.50 Implementors Group.
	1991
March	Reviewed available Z39.50 and corresponding ISO documents.
April	Started creating the encoder and decoder of the APDUs into and from ASN.1 using the Basic Encoding Rules.
May	Tested internal consistency of the encoder and decoder. Constructed the mapping of use attributes into LIAS commands.
June	Finished testing of the encoder and decoder. Exchanged APDUs with the University of California to check validity across institutions.
July	Started code to map LIAS commands into use attributes and vice versa. Started integrating the encoder, decoder and command mapping code into LIAS as a server. Used drivers to provide the APDUs for testing.
August	Started integrating the encoder, decoder and command mapping code into LIAS as a client. Used drivers to provide the APDUs for testing.
September	Began to investigate running the client and server over TCP/IP.
October	Internally tested LIAS as both a server and client.
November	Tested our client with e-mailed APDUs from UC.
December	UC DLA and UC Berkeley started testing their clients against our server via the Internet. Started testing our client against UC DLA's and UC Berkeley's servers.
	1992
February	Reintegrated Z39.50 development work with the continuing development work of LIAS, which by this time had added keyword capabilities.
March	Demonstrated the first full Z39.50 implementation for cross-institutional access to online public access library catalogs on disparate hardware platforms at the National Net '92 Conference in Washington D.C.
July	Enhanced client to have several connections open simultaneously.
Summer	Successfully exchanged information with several other institutions, such as RLG, DRA, AT&T.
October	Access to RLIN made possible using LIAS with full Z39.50 client features.
	1993
January	Released EIP (EI Page 1), the Engineering Index available through RLIN, to LIAS users via Z39.50.

***Figure 1.** Z39.50 Client/Server Development at the Pennsylvania State University.*

history of this development effort. The initial effort at Penn State focused on building a server to handle incoming requests from the University of California and integrating a Z39.50 client into LIAS to search the UC database. At this time, Z39.50 was evolving from version 1 to version 2, which complicated the development environment. We were committed to supporting version 2, although it was not yet fully fleshed out. Eventually, the scope of Penn State's project was expanded to other information systems, including a number of abstracting and indexing databases as well as the library utilities OCLC and RLG. At Penn State the analysis was a shared effort by our two senior development staff, along with a full-time analyst-programmer, who also wrote the client-server code.

WHAT IS A Z39.50 CLIENT?

Simply stated, the client (or "origin," in Z39.50 terminology) is the system that sends a search request to another system (the server) that contains the database being searched. Basically, the client translates a user's local command into the Z39.50 standard, encodes this information into a Z39.50 APDU (application protocol data unit) using the ASN.1 Basic Encoding Rules, and ships it to the server. The means of transmission that PSU uses is TCP/IP; however, this is not a requirement of the standard. The abbreviation "ASN.1" stands for "abstract syntax notation one" and lays out the basic ground rules on how to represent data. The server, in turn, decodes the ASN.1 encoded APDU, translates the request into its native syntax, and then executes the request. The reverse takes place when the server sends back the response to the client. The client translates between its native syntax and the Z39.50 protocol when sending out requests to the server; the server translates between the protocol and its native syntax when it sends a response back.

The protocol provides a uniform method for clients to access servers without having to learn the local commands and features of the servers and without concern for the hardware platforms on which those servers run. Because the client and server communicate using one common syntax, they do not have to know the syntaxes specific to every possible peer.

DEVELOPING THE CLIENT AND SERVER SIMULTANEOUSLY

Because Penn State was working on a grant to communicate back and forth with the University of California, there was a need from the beginning for us to develop both a client and a server simultaneously. In many ways, the existing design of LIAS made our client-server development effort easier. For one thing, we started off with a mature OPAC that was written in such a way that it was easy to insert new code modules. We did not have to tear down the existing system, but simply add to it. However, there is a flip side to this story. Because LIAS already existed, certain structures were in place that posed problems that had to be specially dealt with. For example, LIAS returns entry lists, which are actually the indexed elements of records, in response to a number of commands. This was not readily replicated in other systems.

Although the client and server are different executables, they share many common routines. For instance, both access the encoder/decoder and the configuration file described in the following paragraphs. It is also important to note that the client has a dual nature. It sends searches to other databases and also acts as the server for the local system. In both cases, it uses a common search engine.

Encoder/Decoder

The information being transferred between the client and server is an agreed-upon encoding of data using ASN.1. It is necessary to translate from ASN.1 to the system's native syntax. To accomplish this, two choices exist: either use an existing ASN.1 compiler, or write your own. We chose the latter path and wrote our own encoder/decoder, with the Encoder handling data on the way to the server and the Decoder handling data on the way back.

At the point that we began Z39.50 development, the existing ASN.1 compilers were somewhat shaky. We wanted to be able to make rapid changes as the standard and our work progressed. By writing our own encoder/decoder, we had more control over our development.

Configuration File

Penn State uses a configuration file to contain selected data about each server that the Z39.50 client accesses. The data includes the IP address and login, the database name, commands that cannot be executed on that server, and types of records supported on the server (full or brief). The client consults this configuration file whenever a connection is made to the server. The configuration file provides information that will be accommodated when the Explain Service, proposed for version 3, is added to the Z39.50 protocol. Explain will enable a client to find information about particular servers, such as what access points and data elements are available.

Adding Clients and Servers Today

Penn State's development of both the client and the server basically paralleled the development of the standard. Consequently, it sometimes happened that different clients and servers had different interpretations of what certain elements of the standard were all about. Most of that is under control now, and developers would be spared that inconsistency at this time. Additionally, developers have been most willing to share information and code.

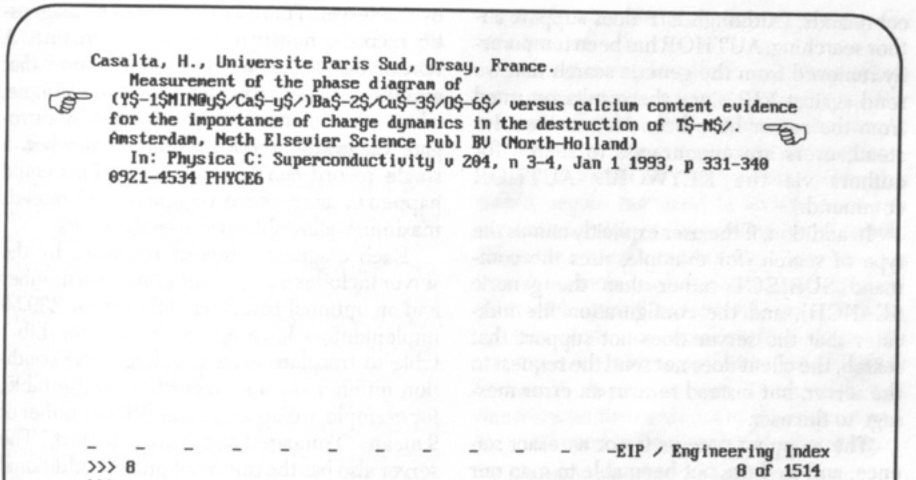


Figure 2. Display of Super- and Subscripts on an EIP Record in LIAS.

SPECIFIC DEVELOPMENT ISSUES

Some of the issues that had to be resolved during the development of the LIAS Z39.50 client are described in this section. They include mapping commands into use attributes, handling cases not covered by the standard, processing diagnostics, integrating new LIAS features into Z39.50, and connecting to simultaneous multiple servers. This is by no means an exhaustive list, but just a few of many issues.

Mapping Commands into Attributes

One of the first things we had to do in developing the LIAS Z39.50 client was to map LIAS commands into attributes. These attributes may specify, for example, the type of term or use (subject, name, title, and so on), its position (first in field, any position in field), its structure (phrase, word, and so on), whether or not it is truncated, and its completeness. The client must decide how best to translate its local commands into these Z39.50 attributes in order to preserve what the command does and what the user expects it to do; Z39.50 works by mapping local searches on the client to attribute combinations, which are mapped back to local searches by the server.

For example, the primary command that LIAS uses is a generic command called SEARCH. This command allows users to simultaneously search on all combinations of author, title, subject, and series. The user does not have to know or specify what index he or she is searching on. However, SEARCH requires that the user input the request in exact word order rather than as keywords. This type of multi-index, word-order-dependent search is not easily supported in the University of California or other server environments that we are dealing with. Therefore, in order for this generic search to be understood by the servers, the LIAS Z39.50 client maps it to a logical OR of AUTHOR, TITLE, SUBJECT, SERIES. Furthermore, when a user inputs the SEARCH command, the LIAS client checks the configuration file mentioned earlier to see if a server supports a particular type of search. The client then sends out the largest supported subset of the search types. For instance, the profile for the server called EIP/Engineering Index Page 1, maintained by RLG, indicates that it does not support subject or series searching. If the user enters a generic search against EIP, the client translates it to a title only, without bothering to send the AUTHOR, SUBJECT, or SERIES commands. (Although EIP does support author searching, AUTHOR has been temporarily removed from the generic search that we send against EIP, since the results returned from the server have been inconsistent. Instead, users are encouraged to search for authors via the KEYWORD AUTHOR command.)

In addition, if the user explicitly names the type of search (for example, uses the command SUBJECT rather than the generic SEARCH), and the configuration file indicates that the server does not support that search, the client does not send the request to the server, but instead returns an error message to the user.

The mapping process is not an exact science, and we have not been able to map our commands 100 percent. For example, even with more than 200,000 attribute combinations, we could not map our TITLE AS SUBJECT command (which searches only USMARC subfield 6XX$t) into the use attributes, since this type of term is not supported in the Z39.50 protocol.

Cases Not Covered by Z39.50

The client may also have to determine how to handle specialized cases not covered by the standard. For example, when we started testing the LIAS client against RLIN's EIP database, we retrieved some interesting-looking (or, some might say, nasty-looking) records. We contacted the people maintaining the server (RLG) and found that in this database chemical formulas containing superscripts and subscripts are expressed in a format specific to EIP, that is, a local format, not the USMARC record syntax. As shown in figure 2, the superscripts and subscripts are returned to LIAS as dollar signs. A situation like this is not covered by the Z39.50 standard. It could be assumed that it is the server's responsibility to handle such cases, but there is no hard and fast rule. The server may decide to do nothing about a particular problem. In that case, the client may decide to write special code. By the way, we have not yet decided what to do about this, but we know we may have to do something, since our users won't understand a record like this as is.

Processing Diagnostic Records

It may be interesting to discuss how a client handles diagnostic (error) records returned by the server. There are two types of diagnostic records: nonsurrogate and surrogate. A nonsurrogate diagnostic record is one that applies to an entire search—for example, when a command is not supported. A surrogate diagnostic record is returned when a single record can't be retrieved. This could happen in cases where a single record exceeds maximum-allowable-size specifications.

Each diagnostic record returned by the server includes a diagnostic condition number and an optional text field. All current Z39.50 implementors have agreed to use the bib-1 table to translate what the diagnostic condition number means. According to the table, for example, a diagnostic condition number of 9 means "Truncated words are too short." The server also has the option of putting additional information in an unformatted ASCII text field. For instance, in the previous example the server may choose to indicate which words are too short.

To be user friendly, the client must determine what the diagnostic record means relative to the user and display it accordingly. For example, when doing a subject search on a database with no subject keys, one server returned the nonsurrogate diagnostic record "Malformed query." While a seasoned developer may understand such a diagnostic, the typical untrained user would be left quite perplexed. There may be additional information in the unformatted ASCII text field that might help the user find out what the server didn't like about the search, but the server is what decides whether to include the additional information field, so it might not be there. If it is there, the client could display it, but we haven't yet decided how we want to do that.

We decided that the LIAS client should interpret some of the diagnostic records returned by the servers to make them more informative to the user. For example, a server might return the diagnostic condition number meaning "Unsupported search." LIAS displays this error message as "Server X does not support that type of search." This is an attempt to make the user aware that LIAS is not rejecting the search, but that the remote system is rejecting the search. A server might also return the diagnostic condition number meaning "Permanent system error." LIAS displays this as "Server X had a permanent system error; please contact LCS." If a server has a permanent system error in response to a LIAS command, there may be a bug in the LIAS client or in the remote server, and Library Computing Services should look into it.

Integrating New LIAS Features

One of the more interesting aspects of Z39.50 development at Penn State was the initial lack of keyword, Boolean, proximity, and truncation search features in LIAS, since these features were typically in place in the servers we planned to access. All of these features were scheduled to be implemented at about the same time that Z39.50 was to be implemented and released in LIAS. This was both a help and a hindrance. On the one hand, it was one less thing to worry about as we initially built the LIAS client and server; on the other, Z39.50 development staff had to keep these features in mind so they could be shoehorned into the LIAS Z39.50 client and server.

Connecting to Simultaneous Multiple Servers

One relatively recent development on the client side of LIAS is the ability to connect to multiple servers simultaneously. The initial development focused on connecting to one server, the University of California. After this was successfully implemented, we built on the code to allow connections to multiple servers and can now have up to ten connections open at one time. By the way, ten is an arbitrary number that can be changed if necessary.

The LIAS client is also designed in such a way that it is possible to issue a one-shot search to another server while remaining connected to the current server. For example, a LIAS user can use the command @EIP to connect to the RLG server and conduct any number of searches. Then that same user can conduct a single search in the LIAS database by issuing the @PSU command. Meanwhile, the connection to RLG remains active, although the next command the user specifies is executed in the local PSU database. This is particularly helpful when users find journal titles in the EIP database and want to make a quick check in LIAS to see if that journal is available in the local library's collection.

Users are also able to refresh themselves as to which connections are open by issuing the command DIL (display libraries). LIAS displays a complete list of databases that are currently open for the user requesting the information.

Having multiple open connections at one time raises some interesting questions. For example, should it be possible for the user to send a single search request to all open connections? From the user's point of view this would negate the need to switch from one server to another and repeat the same search. From the system's point of view, however, this would be technically difficult, raising questions such as how and whether to collate responses, how to accommodate responses coming in from servers at different rates, how to handle duplicates, and so on. Although there are no firm answers to these questions, they will need to be discussed before long.

IMPLEMENTATION ISSUES

During the implementation and testing of Z39.50 in LIAS, considerable time was spent talking to other developers, debugging each other's servers, negotiating, and so on. Early on, for example, our development staff had "gentlemen's agreements" with other developers, agreeing not to perform certain searches that could bring the servers to their knees! Likewise, working with a three-hour time difference between Pennsylvania and California, we had to find mutually acceptable times for testing our clients and servers. Below, specific problems faced in implementing Z39.50 in LIAS are identified.

Network Delays and Perceived Delays

Early in our testing, the network itself was often a problem. Response time across the United States varied greatly, from a couple of seconds to a half minute or more. Our users tended to blame the local implementation rather than the network delays. The upgrading of the Internet has made this less of a problem.

The protocol itself also introduces perceived delays. While LIAS retrieves and displays one record at a time with a resultant quick response time, Z39.50 retrieves multiple records, with the associated retrieval delays for the whole set before anything is returned to the client and displayed to the user.

User Expectations

After almost two years of design and development, perhaps the biggest problem faced in introducing Z39.50 at Penn State was demon-

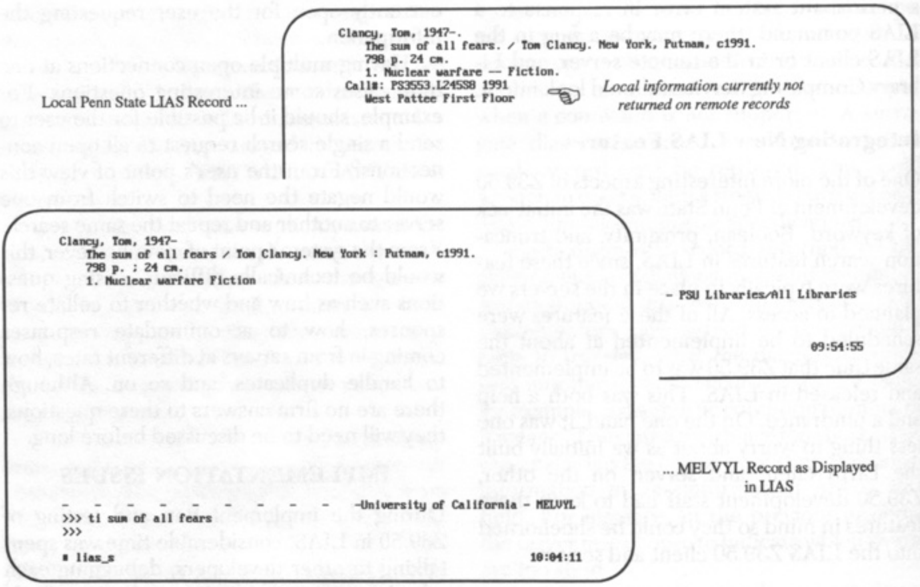


Figure 3. Similarity of Records from the Local and Remote Databases.

strating to users that something had actually occurred! To our users, searching the University of California databases was just like searching the local Penn State database, which was, incidentally, the goal of the project. The users understood that, but still expected to see something radically different—why else would we have spent two years on it? The commands were the same, and, as figure 3 illustrates, the displays were almost the same. The weaknesses of LIAS were still there along with its strengths. The sameness amplified the problems with attributes and differences in indexing mentioned earlier. The user perceives the interface to be the same, and expects that the indexing, retrieval, and response time should also be the same, which isn't true now and probably never will be.

LIAS users are encouraged to report errors they find in their local catalog via a command called OOPS. Since the remote databases look and feel just like LIAS, users expect to be able to report errors in those databases as well. However, the data in remote catalogs is not ours to change, nor is there a vehicle through which a user can request a change. We must emphasize to users that although the databases look and feel the same, there are some very real differences, too.

System stability also becomes an issue in terms of user expectations. The user may depend on a highly reliable local system. With the ability to access more and more remote systems, the likelihood of experiencing downtime increases—there may be network server problems, problems with the remote mainframe, and so on. In any case, this will likely always be perceived by the user as a local problem, with the local system apt to be judged accordingly.

Training and Documentation Issues

Integration of Z39.50 into LIAS is not an end unto itself. Before meaningful access is possible, users must be trained. Training users, and particularly staff who serve as intermediaries to Z39.50 databases, will not be an easy task. The perception that indexing and retrieval is the same across all servers will be one of the thorniest issues trainers will face. They will need to familiarize themselves with the indexing and retrieval conventions of the other database so that they can interpret the search results for users and have ready answers for the question "But I can do this in LIAS—why not here?"

It will also be necessary to help users understand which commands in the local system may be especially effective for accessing remote databases. For example, LIAS has a command called BRIDGE that takes users from a specific access point (author, title, subject, series) in a single record to all other records containing that same access point. BRIDGE can be used in local LIAS, and it also can be used to cross over into a remote Z39.50 database and reexecute a search without rekeying. Such powerful capabilities have to be emphasized so the full functionality between systems can be exploited.

Additionally, Z39.50 only provides the tools to search within the server databases. The burden of instructing users which databases to search for a given topic will fall to the trainer librarians. For that matter, how will the user even know which databases are available in each server? Short term, we will address these issues locally. Long term, however, the Z39.50 Explain Service proposed for version 3 of the standard will provide answers to these questions and make trainers' lives a little easier. Explain will allow the client to obtain details of the nature of the implementation of a server, to learn which databases are available for searching, to learn what attributes sets and diagnostic sets are used by the server. Penn State also plans to use local help files and paper documentation to raise user awareness about Z39.50 and what it can do for them. Now, as we end the development phase, the documentation aspect is beginning.

APPLICATIONS

Penn State has achieved its initial goal, which was to build a client-server environment to communicate with the University of California. Additionally, we have successfully connected to a number of other servers, including OCLC, RLG, Data Research Associates (DRA), AT&T, and Acadia University.

In fall 1992, access to the RLIN database was provided as our first Z39.50 application. In January 1993, we increased our Z39.50 access to RLIN databases by making RLIN's version of EIP accessible to Penn State students, faculty, and staff. This is the first fully publicly accessible use of the Z39.50 protocol between dissimilar mainframes. Response to EIP has been very positive. We will of course continue to make additional OPACs and databases available in the future; this is very simple to do with stable servers.

Cataloging opportunities also provide interesting possibilities at Penn State. Once our client receives a record, to LIAS it is exactly analogous to a USMARC record. Therefore the record can be used as cataloging copy or an acquisitions record and added to our database if we so desire. Although Penn State is not currently doing this, it does present ownership questions that will eventually have to be dealt with as more clients have this capability.

Currently, implementors are aiming at handling bibliographic data. However, full text and images will eventually be accommodated. A mechanism is provided by Z39.50 through which the sharing of bibliographic, full text, and image data could radically change the way libraries interact. For years, we have talked about the library without walls, and this development could certainly be key to that concept. Fulfilling this, however, will take more than the writing of code. There are also major philosophical changes to be considered in terms of what's yours, what's mine, and what's ours. That could be the biggest Z39.50 challenge of all. ■■

Special Section: Second Annual Library Directors' Conference—Linking Multimedia Digital Libraries: The Changing Infrastructure

Introduction

Barbara L. Scheid

VTLS Inc. held its second annual Library Directors' Conference on October 4–6, 1992, at Mountain Lake Resort in Mountain Lake, Virginia. The theme of the conference, "Linking Multimedia Digital Libraries: The Changing Infrastructure," was an extension of last year's theme. The conference focused on the changing infrastructure for information access and the emerging technologies required to support it.

The purpose of the conference was to bring library directors up to date on networking issues and technological trends. The conference was informal and was attended by library directors and assistant directors from public, academic, corporate, and special libraries located all over the world. The conference provided ample opportunity for attendees to network with colleagues and recognized experts in the library and information fields.

The invited speakers discussed topics ranging from public policy issues to developing multimedia OPACs. Attendees also participated in breakout sessions on the changing roles of librarians, publishers, and users of information. These sessions are not summarized in the proceedings.

On the first day of the conference, Robert C. Heterick, president of EDUCOM, gave an overview of the potential values of the Blacksburg Electronic Village project. Peter Young, director of the National Commission on Libraries and Information Services, read and elaborated on Representative Rick Boucher's paper, which reviewed current legislation on NREN and perspectives on home access. Representative Boucher was unable to attend because Congress was in session. Vinod Chachra, president of VTLS Inc., spoke on developing multimedia OPACs and served as moderator for the panel discussion. Diana G. Oblinger, program manager for IBM, demonstrated how multimedia information access systems transform the college classroom. Eugenie Prime, manager of Corporate Libraries for Hewlett-Packard, concluded the first day's general sessions by addressing the corporate imperative for virtual libraries.

On the second day of the conference Paul Evan Peters, director of the Coalition for Networked Information, gave an overview of the evolving network infrastructure and evolving networked information environment. Walt Crawford, senior anlayst at The Research Libraries Group, Inc., and president of LITA, talked about future trends in OPAC design and unified user interfaces. Peter Young and Frank R. Bridge Jr., president of Frank R. Bridge Consulting, Inc., served on a panel with the other speakers that summarized the conference. The panel discussion was titled "Creating the Future with the Initiators of Change." The panel members first reviewed what they thought was the most pertinent issue in the information age and then answered questions from the audience. The panel discussion covered issues ranging from public policy and privacy to copyright and equal access.

VTLS chose the guest speakers because

Barbara L. Scheid is Public Relations Coordinator for VTLS Inc.

of their expertise and did not require them to write formal papers. Summaries of their presentations, as well as Representative Boucher's formal paper, are included in the proceedings. ■■

The Blacksburg Electronic Village: A Field of Dreams

Robert C. Heterick

For several years visionaries have proposed that if community networks were built, citizens would quickly assimilate them into their daily life cycles, and the world of networked information would explode with a vitality that would quickly define the information age. The information transport providers—regional Bell operating companies, local and interexchange carriers, cable TV and wireless companies—have generally ignored the visionaries. They have followed the first law of wing walking: never let go what you have hold of until you have hold of something else. They have sought something more than vision and anecdotal evidence to support the notion that if they build it, the consumers will, in fact, come.

The Blacksburg Electronic Village was conceived as a test bed of the "field of dreams" scenario. A partnership, composed of the local municipality, the local university, and a number of information infrastructure and service providers, is being assembled to create the defining experiment. This experiment will be different than previous trials. Telephone companies (telcos) and regional Bell operating companies (RBOCs) have, for some time, been engaged in technology trials—many of which have as their goal "fiber to the home." Computer vendors have searched, so far in vain, for an elusive home market for their products. Universities have experimented with access through inclusion of parts of the local community in the campus network, the most far-reaching manifestation being the "freenet" model. Municipal governments have generally not been concerned with issues of citizen access beyond control issues such as franchise, rate structure, right of way, and easement granting. The Blacksburg Electronic Village was created to focus on citizens and their access to a rich and robust variety of electronic information sources.

Robert C. Heterick is President of EDUCOM and formerly Vice-President for Computing and Information Systems at Virginia Tech, Blacksburg.

THE TOWN AND THE UNIVERSITY

I have to say in all fairness that Blacksburg is an unusual situation. There are approximately thirty-four thousand individuals in the town of Blacksburg, which is home to Virginia Polytechnic and State University (a.k.a. Virginia Tech). Two-thirds of the population already have significant experience with the world of information networks. The university is interested in the project because more than half of its students (eleven thousand) live in the town of Blacksburg. While the campus already has an extraordinary communication infrastructure, the town has a very weak communication infrastructure. The town would like to see the economic development and improved quality of life that a project such as the Electronic Village promises.

ACCESS, NOT TECHNOLOGY

The Blacksburg Electronic Village is about access, not technology. We have identified a number of off-the-shelf technologies that will undergird the experiment. The telephone industry is well along in planning for a narrowband Integrated Services Digital Network (ISDN). While we may need to hedge our bets that this particular technology provides too little bandwidth, it is nonetheless sufficient to start the experiment. Adding conventional analog technology will provide a low-bandwidth safety net for users for whom higher bandwidth access is prohibitively expensive. Clearly, an end-to-end digital infrastructure will be superior and offer the maximum opportunity and clearest migration path to higher bandwidth and more complex services. The presence of a digital telephone switch in Blacksburg could make available a wide range of telephone services not now available to subscribers.

The software platforms envisioned for the Electronic Village are also off-the-shelf modules. The major shortcoming of extant software that will be used is its lack of integration into a coherent, easy-to-use package. Integration efforts will be focused on constructing a relatively intuitive interface for the user. It is this interface that users will come to think of as the Electronic Village.

In designing the Electronic Village we have been faced with tradeoffs. The most difficult tradeoff we have found is that you can make information accessible or you can make it secure, but you cannot do both. This is a difficult problem for society to cope with over the next decade. Many of you sitting in the audience understand this problem because you have to consider fair use versus copyright laws. This tradeoff permeates the whole computer communications revolution.

THE VILLAGE AS A LEARNING EXPERIENCE

The overarching issues in the Blacksburg Electronic Village are political and economic. It is this set of problems that provides the testbed for study and whose solutions will define the measure of success for the project. We know that there exists the technology to build the infrastructure. We know that the software modules exist to form the core of services. What we don't know are the services consumers will use; the price elasticity of various services; how services should be priced and billed (by subscription, measured by access, by time, etc.); what represents a minimal core of services to develop a critical mass of users; how infrastructure costs can be defrayed over time in the current regulatory environment; the efficacy of multiple infrastructure providers; the appropriate dispersion and pricing of higher bandwidth services; how to organize and manage the effort; and a host of similar issues.

There is a tendency on the part of observers, even participants, to oversimplify the project. The telcos, probably because of their long history as regulated monopolies, tend to want to view every issue as a technology problem. They focus solely on the first cost of providing services to the exclusion of long-term market issues. For the village project to be a success, both consumers and providers must find cost-price issues amenable. Consumers must willingly acquire services because of perceived value and providers must meet profit expectations with the pricing of their services. However, the pricing needs to reflect prices of full, nationwide deployment, not just a small trial as exemplified by the Blacksburg Electronic Village.

The news media, hooked on the cachet of an interesting story line, tend to focus on flashy ideas—fiber to the home being the predominant one. But fiber to the home will only make economic sense when regulations, which are the consequence of the Bell divestiture, change. These regulations, which Representative Boucher addressed in his paper, are anticompetitive. They were not intended to be that way; but they have turned out to be so. When this political problem is addressed, not only will the desirability of fiber to the home be obvious, but the economic viability will be manifest. For nearly all of the services currently even contemplated, the bandwidth of the current copper premises infrastructure, appropriately utilized, could suffice.

The Town must assume a much more activist role in guiding and directing the deployment of infrastructure within the community. In the current case of numerous, vertically integrated infrastructure providers, the Town experiences continuing and unnecessary disruptions to its vehicular and pedestrian systems as these multiple organizations require permits to interfere with traffic while installing and maintaining their duplicating infrastructures. The quality of some of these infrastructures tends to be lower than desirable, raising further the societal costs.

A NEED TO SIMPLIFY

The university and software developers in general have a tendency to presume higher technology literacy on the part of the participants than seems warranted. We must learn to find less complex and more robust strategies for integrating the Village software than heretofore have been typical. This problem likely extends beyond the domain of software into the hardware itself. The "standard" personal computer will be found too complex (and perhaps too expensive) for some citizens. We will continue a dialogue with hardware vendors who have expressed an interest in developing "information appliances" for the consumer electronics marketplace.

A FOCUS ON THE CITIZEN

What sets the village project apart from many previous trials and testbeds is its focus on citizen applications and the inclusion of an entire community in the experiment. Blacksburg is a geographically compact area that includes nearly all of the normal political, social, and commercial interactions of the participants. One of the principal measures of success of the project will be the number of citizens who elect to participate. Deployment choices that would limit the number of participants are extremely poor choices and contravene the global intent of the project. Such choices might be in the infrastructure, the user interface, or pricing strategies.

CHOICES, COSTS, FORMATS

We expect to offer a rich variety of choices for connecting to the village ranging from slow (9.6 Kbs) analog modems, through ISDN (64 Kbs), to Ethernet (10 Mbs) on campus and in high-density residential complexes. The cost of connectivity will reflect a similar variety, ranging from a $200 modem and a $5 monthly access fee, through a $20 monthly ISDN line, to a $20 monthly Ethernet connection on campus and an as-yet undetermined monthly fee for an Ethernet connection in apartment complexes. These pricing issues are compounded by questions of whether (or in which cases) information services pricing should be bundled in the access fee or charged separately in subscription or transaction charges.

Much of the information available in the village will be available in alternative formats, but with different time constraints. At least since the Library of Alexandria we have collected information in the form of artifacts—books and periodicals. It is now evident that in the "Information Age" the surrogate of a collection of artifacts is a poor measure of the quantity of information available. Access, delivery to the user's desktop, is the new measure of the availability of information. Modern communications networks have made the location of information unimportant and shifted our focus to access.

IN CONCLUSION

The shape eventually assumed by the Village is difficult to predict. The most successful services that will emerge are likely not in our current thinking. If the French experience with Minitel or the U.S. experience with 900 numbers is instructive, we should expect to see high popularity of multiparticipant games, various forms of directory services (white and yellow page equivalents), and discussion groups ranging from dogs and cats to extraterrestrials. Given our society's apparent desire to improve the K-12 education experience, we would expect that services ranging from communication between parents and teachers to multimedia learning materials will be found attractive.

As Mark Twain observed when he first heard Wagner, "that's better than it sounds." If the Village only provided alternative formats or more timely access to things we already had, it would fall far short of both our aspirations for it and its potential to significantly enhance the way we work, learn, and live. ■■

Accessing Multimedia Information In Virtual Libraries

Vinod Chachra

In the introduction to his book *The Work of Nations,* Robert B. Reich states, "We are living through a transformation that will rearrange the politics and the economics of the coming century. There will be no national products or technologies, no national corporations, no national industries . . . Each nation's primary assets will be the citizens' skills and insights. . . ."

Vinod Chachra is President of VTLS, Inc.

TRANSFORMATIONS

This is a period of transformations: transformation from the industrial age to the information age; from structured hierarchical organizations to unstructured, networked organizations; and from high tech to high value. Welcome to the information age. At the peak of the industrial age fully 65 percent of the U.S. labor force was employed in industrial production and less than 4 percent in agricultural production. By the mid-1980s, more than 50 percent of the U.S. labor force was involved in nonindustrial, information-related fields. This number has been growing as industrial and manufacturing jobs have declined to below 25 percent in the 1990s.

NETWORKED ORGANIZATIONS

In the industrial age we focused on muscle multiplication, the process of leveraging our physical capabilities. It was marked by high-volume production with repetitive tasks. Structured hierarchical organizations were best suited to manage this environment. In the information age we focus on idea multiplication, the process of leveraging intellectual capabilities. It is marked by creativity and innovation. The tasks are rarely repetitive. Structured hierarchical organizations are not suited for this environment. Knowledge workers need to interact with other individuals independent of rank or position. Sproull and Kiesler, in their book *Connections: New Ways of Working in the Networked Organization*, define a networked organization as "one in which all employees participate fully in the information life of the organization, independent of geographic, organizational or social location." According to Richard Nolan, combining new technology with old organizations always gives disappointing results. Organizations themselves must change to take advantage of new technologies. Library directors beware!

PAYING FOR INFORMATION

Pricing and costs are always sensitive issues. As we complete the transformation from the industrial age to the information age, we find increasingly that the cost of the product is based on its intellectual content and less on its physical content. Books, CDs, and computer software are examples of products that fall in this category. For example, the VTLS software you receive on tape has a reproduction cost of around ten dollars. The several thousand dollars that you pay for the software is for the intellectual content of the software; the physical content is indeed very, very small.

Let us carry this concept to the extreme. Imagine a situation where the physical content of the information we buy is down to zero. This is not an unrealistic situation, for this is exactly what happens when we copy software or information off a network. Suddenly, we have difficulties with this concept. Accountants get confused; no shipments are received; no packages are opened; hardly an audit trail exists; but the information is transferred, and a charge is levied. Some have difficulty in adjusting to this concept. But this is exactly what we will have to get accustomed to when we start accessing multimedia information from virtual libraries (or virtual bookstores) over national and international networks.

REAL BOOKS ARE NOT DEAD YET

Before we get into the topic of virtual libraries holding virtual books for real borrowers, one must raise the question of real books. Are real books really dead? We all know the answer to the question. Books are not dead—at least not yet. Your reasons for this conclusion may differ from mine. And you probably have good reasons. For me, there are two important reasons why books are not going to die soon. First, there are so many of them around. Even if no new books were printed, it would take decades to convert the existing books to electronic form. Further, it would not make economic sense to convert all the books. Second, and more importantly, in order for the new media to replace books it must cost like a book and act like a book. Cost issues are not insurmountable. On a unit cost basis, even for a small number of books (less than twenty-five), the cost of the electronic media is already lower than that of the printed book.

Acting like a book is another question. The reliability of a book is unbeatable. You can drop a book from five feet and it still shares its knowledge with you. Try that with your computer. Books are immune to power failures or weak batteries. Books are extremely portable, although some of the new electronic books are getting that way. The density of information on a printed page still exceeds that of electronic books. Books have a contrast ratio of 120:1, whereas that for electronic media is only up to 50:1 (but catching up fast). Electronic media still have a way to go, but by all estimates, the technology to completely surpass all the qualities of the book will be in place as early as 1995. Whether we will ever end our romance with the book is quite another question.

FOUR NEW VIRTUAL TERMS

Virtual books, virtual libraries, virtual bookstores, and *virtual collections* are four more or less fuzzy terms. The terms are fuzzy because they all try to label developing concepts that are not as yet concrete. One would think that virtual books make up virtual collections stored in virtual libraries or sold in virtual bookstores. This is not true. For instance, virtual libraries don't house virtual books. Virtual bookstores don't sell virtual books—real stores do. Let us define some of these terms. *Virtual book* is a term used for software, portable and hand-held hardware, and a data system all designed to be a physical replacement for a real book. A better term for a virtual book is an *electronic book*. It is a real product with a physical manifestation. A virtual library, on the other hand, is a collection of indexes (or bibliographic databases) and the electronic objects (text, sounds, images) that these indexes point to, both of which can be retrieved remotely over a network. A virtual bookstore is essentially a virtual library that sells its digitized electronic objects. During this transformation period, real libraries and virtual libraries must coexist with patrons' having access to real books and to electronic information. It is, therefore, better to think in terms of both virtual collections (materials that are accessible) and real collections (materials that are housed) that are available to patrons in a given library. A library may supplement its local collection with one or more virtual collections made available to its patrons. Whereas patrons may have to go to the library to access the real collection, they can access a virtual collection not only from the library, but from the home or office as well.

MULTIMEDIA INFORMATION AND MULTIMEDIA OPACs

Multimedia refers to the combination of two or more media—such as sound, images, text, sound-synchronized images, image-synchronized sound, and full-motion video. Multimedia products are available in three different types of applications—education and training, consumer electronics, and information access (OPACs). Multimedia in consumer electronics is characterized by the fact that general purpose computers are not used. There are VCR-like players that provide access to sound, images, and text, but they do not support full-motion video. Multimedia in education uses computers and additionally provides access to full-motion video. It also allows for the creation of hypermedia links between these multimedia objects. Multimedia in education is important in that it allows users to explore and discover as they follow their curiosity. Through sophisticated simulation methods, it also allows for learning by role playing and experimentation. Multimedia OPACs, on the other hand, use the same technology as multimedia in education, but their focus and implementation are different. Multimedia OPACs provide access to distributed multimedia databases over a network. Thus a multimedia OPAC is a physical implementation of a virtual multimedia library.

A FIVE-PART MODEL

The working model of the multimedia OPAC (or virtual library) consists of five parts: the display system, the search system, the storage system, the network, and the marketplace. The first three parts—the display system, the search system, and the storage system— need not be located near one another. Rather, they can be located anywhere in the nation or the world—it is the network that ties these three parts together.

The workstation, also called the scholar's workstation, consists of a computer system able to display high-resolution images and play CD-quality sound, and containing enough memory, storage, and appropriate interfaces to handle full-motion video. The workstation has the ability both to display and to edit the information. Thus it must support sound editors, text editors, and video editors and must also have the ability to capture and store the accessed information locally. Moreover, it must support the creation and maintenance of hypermedia links between these stored objects. The VTLS InfoStation is an example of such a workstation.

The index defines and provides access to the virtual collection. It is a system, like the VTLS system, that allows for the description and indexing of the multimedia data. A virtual library may have one or more of these virtual collections. The index to each collection points to the location on the network where the multimedia object described in the index is stored.

The file server can support standard protocols such as FTP (File Transfer Protocol) and NFS (Network File System) for accessing files in large databases. Stored on the file server are the electronic, digitized objects. These objects are sometimes called BLOBS (for Binary Large OBjectS).

The network ties the workstation, the index, and the file server together into a comprehensive system. The network has to be high speed and reliable; otherwise, the time required to display the information would be unacceptable.

The marketplace is the part that deals with issues such as costing, access control, authentication, copyrights, distribution and redistribution rights, and royalty payments. The marketplace ultimately defines who has access to what, when, and at what price. The marketplace has an important role to play in this model. It defines all the nontechnical issues related to making a concept like this a reality. Depending on how well we handle the legal, financial, management, access security, authentication, billing, and payment issues, we will either nurture or stifle the usefulness of a virtual library.

The operation of the model is very simple. The user sits in front of a workstation and issues a search query. The search query is directed toward a virtual collection that may be resident anywhere on the network. There are generally several hits for the specified query. Some of the hits have multimedia data associated with them. When one of these multimedia items is selected by the user, the multimedia button on the panel lights up. Clicking on the button allows multimedia data to be accessed directly. The workstation software reads the information on the index, determines where the BLOB is located, connects to it on the network, and retrieves the information for the user. The Multimedia OPAC workstation can play Beethoven's Fifth Symphony just as easily as it can display the full text of Hamlet or show your favorite movie.

Different implementations are possible for this five-part model. At one extreme lies the implementation where all five parts reside on a single machine. In the more general case, however, the workstation will have access to several virtual collections (indexes) residing in different locations. These virtual collections, in turn, point to a group of BLOBS that may be located on several computers across the network.

Let me close by sharing an anonymous quote with you. "We must go as far as we can see, for we will be able to see further when we get there." Skills and insights are developed through such a journey. ■■

Multimedia in the Classroom

Diana G. Oblinger

Diana Oblinger's presentation focused on multimedia in the classroom. The conference attendees learned through Oblinger's demonstration how professors use multimedia in teaching subjects such as French, chemistry, and American history. She stated that multimedia is important to libraries because "a lot of the work that is being done with multimedia assumes that the library is the central distribution point, training point, and conveyer of services. Many involved in the library community are going to find themselves very much involved in multimedia."

Oblinger began by illustrating the transition to multimedia technology over the past three to four years. She stated that in the past, chemistry was taught in a text-based mode. The text for cell division is "not terribly motivating," which makes it hard to teach and even harder to learn. But, beginning a few years ago, professors using a chemistry graphics software package could show the students the cell; how the nucleous begins to disappear; how the spindle fibers form; and how the chromatin continues to condense. Better yet, during the last year, full-motion video has been added to the course content. Oblinger demonstrated how by "taking the same piece of information about cell division, the student can now watch full motion video that explains about mitosis and cell division while the student watches the mitosis and cell division." The video can also be stopped at any point so that the student can examine the cell in different processes. This lesson can be kept at the library for future reference.

Before Oblinger showed the participants examples of the kinds of courses being taught using multimedia, she asked them to think about whether multimedia had an impact on the instructional effectiveness of each of her examples. The first example she chose was an introductory American history course, which, she said, was a good example because this course was taken by a large percentage of students attending college.

Oblinger explained that teaching the American civil rights movement to an average nineteen- or twenty-year-old student is hard because they don't remember it, and they don't know any of the people who were involved. Using the review, which is a standard teaching tool, a professor may talk about Martin Luther King's speech, or to get a better effect, the professor may have someone else read the speech. But, as Oblinger demonstrated, students who have seen Martin Luther King on the screen giving the speech himself will feel more emotionally involved in what was happening during the time period. Consequently, they will retain more of what they see and hear. Again, the lesson may be kept in the library and used as a resource tool for students.

In another example, Oblinger showed how the professor who teaches chemistry could use multimedia to teach students the periodic chart. Instead of copying the periodic chart and asking students to memorize it, the student could see a picture of bismuth and pictures of products that use bismuth. A further application shows the student the similarities and differences between chemicals and the reactions of combining different chemicals with acid, water, and other things. Oblinger said, "students will begin to take a tool like this and explore things and come back to the classroom with many questions for the professor." She indicated that America's higher education institutions have been criticized for not teaching students to think for themselves. Multimedia instruction will help students think for themselves because their dwell time is much longer. She said students don't want to study periodic charts, but they do like to experiment, and using multimedia instruction is the closest thing to experimenting in the laboratory. Keeping a laboratory fully stocked and accessible to the students is very expensive for

---

Diana G. Oblinger is Program Manager for IBM and the Institute for Academic Technology. This presentation is summarized by Barbara L. Scheid.

The Institute for Academic Technology was set up in 1989 as a collaboration among all institutions for higher education, not only institutions like the University of North Carolina at Chapel Hill, but also community colleges, private colleges, and IBM to work in a collaborative environment to develop the types of learning tools Oblinger demonstrated at the conference. IBM pursued this collaboration "because the needs of higher education are unique, and the only way to develop the right thing is to work with the people who have the needs."

any institution. Access to multimedia databases for courses such as chemistry will help cut down on laboratory expenses and, therefore, pay for the investment.

Oblinger then showed the attendees a lesson used for medical students. The lesson had case studies about what to do with a cyanotic and hypertonic newborn baby, and it explained and showed what the baby's symptoms were. The student was expected to choose which was the best solution for the baby's life threatening problems. Oblinger said, "This multimedia application helps nursing students build self-confidence because it builds critical thinking skills. This lesson also teaches the nursing student without endangering anyone's life."

Teaching methods over the course of approximately four to five years have gone from using text to graphics to full-motion video. Oblinger summarized that multimedia technology allows the professor to transform the classroom into a place for lively discussion, where students take an active role in the learning process and, therefore, remember more of what they learn.

In support of her observations that multimedia instruction provides a better learning atmosphere for the student, she cited the conclusion from a Department of Defense study: "Multimedia is generally more effective and less costly than conventional instruction." The survey report indicated that education results improved 50 percent for college students and 35 percent for K-12 students when multimedia instruction was used instead of conventional instruction.

In fact, according to Oblinger, "Studies taken show that multimedia instruction is more effective. You can teach more in less time; the time involved in teaching is about one-third less; and overall the cost of delivering instruction is lower than with conventional instruction."

Oblinger reported that institutions find multimedia instruction more effective because the senses get involved during the learning process. If a professor stands in front of the classroom and lectures to the students, they will only retain up to 20 percent of what they hear. Students who see and hear the information can retain up to 40 percent of what they see and hear. But, students who are actively involved in the learning process will retain approximately 75 percent of what they see, hear, and do. She said, "What you find in multimedia situations is that students are not merely passive recipients of information." The interaction between the students, teacher, and multimedia application lets the students become active participants in a multimedia learning process.

It is now easy to begin using multimedia instruction in the classroom, reports Oblinger. She said the hard thing to know is the subject matter and, obviously, the faculty already know the subject matter. The professor must also know what people typically understand or do not understand, and this knowledge comes from classroom teaching experience. Educators no longer need to know how to program. "If you can cut, color, count, and paste, you can do multimedia," said Oblinger.

Multimedia instruction can be used to teach any subject matter. Oblinger showed the participants how it was being used to teach French, chemistry, music, and American history. She reported that the tools necessary to produce the demonstration she gave were simple and affordable for most institutions. The multimedia lessons were developed using a system called the Advanced Academic System with Toolbook software from Asymetrix Corporation. During her demonstration, she used a computer, an analog source, a laser disk player, a video overlay board, and a LCD panel. She reported that the same equipment would be needed for the classroom.

Oblinger then briefly reviewed other applications for multimedia technology. She showed the audience how a science major accessing a multimedia database could research possible career opportunities at different companies. Salary information, a description of jobs available, and information about the companies could all be found in one database. A full-motion video of alumni discussing their experiences working in the science field could also be added to the database. Other applications where multimedia has proven effective are student records management and retrieval systems, class syllabuses and descriptions, and faculty directories.

Oblinger concluded, "The first thing I hope you learned today is that multimedia has value and promise for higher education and that we are in the forefront of delivering those services. The second is that it is easy to do." ■■

The Virtual Library: A Corporate Imperative

Eugenie Prime

In her presentation, Eugenie Prime defined the virtual library and reviewed her observations regarding its evolution and importance to users of information. She touched upon both the librarian's role as an information professional and the technological and corporate imperative for virtual libraries.

Prime began by citing Webster's definition of *imperative* as "unavoidable, necessitous, obligatory, mandatory, irrevocable, urgent, and driving." She said, "A time is coming when corporations, in order to survive, will have to move toward the virtual library because it is going to come whether they like it or not." She stated that she has seen her concept of the virtual library change from hard-wired desktop delivery to that of the library being wherever the user is. Again, she cited Webster's as defining *virtual* as "existing or resulting in effect though not in actual fact." She said, "What gives the virtual library its significant paradigm shift is not necessarily the new, dynamically changing technology. The technology is essentially the enabler. But the shift is a subtle, but nevertheless dynamic, shift in the power of the user."

Prime suggested that over the last three to four years, the requester of information or the researcher has become the accessor of information. Libraries and librarians have long ago made the paradigm shift from collections and custodians respectively. She said, "We have essentially been moving imperceptibly, but no less inexorably, toward the virtual library, so that the term *virtual library* is not so radical after all."

In the spring of 1992, *Time* magazine carried an article titled "The Office Goes Airborne." The article related US Air's plans to offer passengers at their seats the ability not only to phone their offices but to send and receive faxes as well. Prime contended that if the office goes airborne, there is no reason the library shouldn't go with it. "The library better go too, or it will go nowhere at all. A virtual library is available at any time, in any place, and to anyone," she said. She stated that this concept was critical to Hewlett-Packard because it is essentially an engineering organization. The research has shown that, given engineers' information-seeking behavior, it is important they be empowered as users. Research on engineers' information-seeking habits and practices indicated that

> The engineers act in a manner which is intended to minimize loss, not to maximize gain. The loss to be minimized is the cost in terms of effort, either physical or psychological, which must be expended in order to gain access to an information channel.[1]

Emphasizing the importance of empowering the user, Prime said the extent of library use is related inversely to the distance separating the user from the library. Engineers follow a "law of least effort. Users should be able to determine what they receive, how they receive it, how much they receive, when they receive it, and where it should be received." Prime explained that engineers like to solve their own problems, and they want specific answers to specific problems immediately. She said research also shows that accessibility is the single most important determinant of use of an information source and that perception of accessibility is influenced by experience.

Prime reviewed the technological imperative, describing how each technological advance has helped move the virtual library toward reality. (However, for lack of space, and because the technological imperative was mentioned earlier in the proceedings, a summary of her review has not been included here.)

In addressing user needs, Prime insisted that new user interface technologies be very easy to use and be as commonplace as kitchen appliances. She referred to an article called "Toward the Domestication of Microelectronics," written by Joel S. Bernbaum, vice-president of research and development at Hewlett-Packard, and published in 1985 in *Communications of the ACM.* In the article Bernbaum discussed the pervasiveness of the computer, noting that as technology becomes pervasive, you are more aware of its absence than its presence. Prime used electricity as a good example of a technology that had become pervasive, in that we do not even think

Eugenie Prime is Director of Corporate Libraries, Hewlett-Packard Company. This presentation is summarized by Barbara L. Scheid.

about it until we don't have it. Bernbaum also stated that technology should be mobile and miniaturized.

Prime believes that voice input and output will make significant contributions to the virtual library because the computer of the future should be used without a keyboard. She said, "We are going into the twenty-first century with an artifact from the nineteenth century, the keyboard, whose only present claim to fame is its ability to cause carpal tunnel syndrome. And the mouse is not much better; it causes tennis elbow." She said we have to get back to what comes naturally, and using one's senses is natural. She said she looks forward to a time when she has her choice of interfaces, such as light pens, touch-sensitive screens, and voice-operated computers. She said she heard Mark Weiser from Xerox Parc talk about information appliances. He told the audience that the time will come when we can go down to the corner store and pick up a "six-pack of computer" the same way we pick up a six-pack of beer.

Another area Prime touched on was human interface technologies, which are technologies that substantially improve the human-machine interface. She asserted that library professionals could make a definite contribution to this area of technology because they have been in the middle of the request-and-access-information process and know what patrons need. She insisted that computers cannot be considered truly user-friendly until one can go to K-Mart, pick one up, bring it home, plug it in, and start to use it. "That is what I do with my telephone and that is what must happen to the computer," Prime commented.

Prime observed that human interface technologies, or appliances, as some people call them, need drastic changes. She said people need intelligent interfaces that are more natural than what are now available. People need appliances that permit them to express their needs and formats. She reported that there is an area of emerging engineering called *Biocomputing*. Biocomputing holds much of the promise that artificial intelligence has lost. She said, "Biocomputing seems to be a combination of neuroscience, cognitives, psychology information science, information engineering, and electronics." This science is devoted to computers that acquire knowledge through learning but do not require programs. In essence, Biocomputing is a paradigm shift because it is patterned after the functional paradigm of living/learning systems. Prime challenged librarians to get involved in this technology. She said this is where librarians have a pivotal role to assume because they have long been involved in user training. She said, "We know as librarians that there has always been a high rate of recidivism because using the system is not as intuitive as we would like."

Another area of change is in the information itself. Prime said societies have placed barriers to information that are purely artificial. She said, "Somehow in the library world we have only felt comfortable dealing with information that originally came on an 8 1/2-by-11-inch sheet of paper. This has to change; information should be boundaryless." Accessing information should take only one step. It must be decided, however, how to store, index, and copyright information before it can be sent over networks.

Prime stated that time consumption has become a critical management and strategic parameter for corporations. "Corporations are becoming horizontal, moving decision making down," she noted. Organizations now have self-managed and self-directed teams in which each member of the team is empowered. The latest trend, like Total Quality Management, is the learning organization. Organizations whose employees can access information before their competition will come out ahead. Timely information is critical to the success of this type of organization and has driven organizations to make information more readily available to its employees. Prime said that Stan Davis' phrase "Time is money" is now giving way to "Money is time," where time has become the key resource and money the way to measure whether we are receiving its full value.[2] In order to maintain a competitive edge, information must be delivered under the just-in-time concept so that organizations can learn more quickly than their competitors. The "just-in-time" philosophy has as its goal maintaining just enough material, in the right place, at just the right time, to make just the right amount of product. Prime argues that librarians discovered the just-in-time concept long ago but just did not know it. She stated that the library philosophy of delivering information has always been the right or relevant information, in the right amount, in the right format, to the right individual, and at the right time.

Prime concluded by stating that she did not think the virtual library would happen overnight, in that we have still too great a romance with the printed page. She doesn't believe the traditional library will end one day and the virtual library will begin the next. "Our challenge is that there is no cutoff date. We have the challenge of stealing second base with half a foot on first," she said.

In order to manage the transition, the library profession must be organized and have an articulate plan for implementing a true virtual library. She said librarians must recruit buyers of their ideas so that they will be sponsored and funded. "Everyone knows the value of information until they are asked to pay for it," she noted. She maintained that this is a time like no other time for librarians, and the profession as a whole must be flexible, nimble, and prepared to take advantage of every opportunity that arises. She believes the challenge will continue to be *accessibility* and *delivery* and that librarians need to find themselves on the right teams in order to stay in the forefront. She concluded by quoting Stan Davis: "The only way organization leaders can get there from here is to lead from a place in time that assumes you are already there and that it is determined even though it hasn't happened yet."[3] "That," she added, "is our challenge. But that, I believe, is also our hope."

REFERENCES

1. Peter Gerstburger and Thomas J. Allen as quoted in Thomas E. Pinelli, John M. Kennedy, and Rebecca O. Barclay, "The Role of the Information Intermediary in the Diffusion of Aerospace Knowledge," *Science & Technology Libraries* 11 (Winter 1990): 59–76.
2. Stanley Davis, *Future Perfect* (Addison-Wesley, 1987), 18.
3. Ibid., 25.

The Changing Infrastructure for Information Distribution

Paul Evan Peters

In this brief talk I will define four moving targets that the industry is tracking, present some background for the evolving network infrastructure, discuss three opportunities and challenges that I believe are arising from the evolving networked information environment, and offer a few strategies for making progress in this area.

I believe that most all of us are now finally looking and perhaps moving in the same direction, and that more and more of us are encountering the turbulence at the barrier between a knowledge creation, distribution, and utilization system that is based predominantly upon print media and one that is based as much on network media.

My goal in this talk is to provide a frame of reference that will be helpful to those who want to get beyond that barrier as soon and as effectively as possible.

MOVING TARGETS THAT THE INDUSTRY IS TRACKING

The major target we had in 1980s was planning, installing, and operating integrated library information systems. In the 1990s we need to be tracking at least three other targets.

The first target is networks. We can no longer plan systems tacitly assuming that we will be able to hook them to networks at some later date. We need to be giving a lot more attention at a much earlier time to our networking capabilities and strategies. We also need to be procuring and updating our systems to meet the requirements of networks that broaden access to and build support for those systems and the services that they enable.

The second target is the delivery of information, not just access to information. This is the target that has caught most librarians off guard. We never really got much credit for modernizing access to information, because patrons want access to actual information, not just to information about information. We

Paul Evan Peters is Executive Director, Coalition for Networked Information.

need to press forward as soon as possible to link our modem access systems to modern delivery systems.

The third moving target, which we will be hearing more and more about as the 1990s wear on, is new types of networked information resources and services. I believe that librarians already have quite a lot of experience dealing with changes in information formats, as our assimilation of microforms and videocassette tapes and many other forms of information amply demonstrates. We are now being called upon to assimilate networked information, and we will soon be called upon to assimilate multimedia and other high-performance networked information resources and services.

THE EVOLVING NETWORK INFRASTRUCTURE

The first thing to understand and appreciate is that in networking, practice is ahead of policy. Networking is now being used to solve problems that, for the most part, were never imagined by the designers and builders of networks. Private, bypass networks have pushed networking technology and networked applications beyond the public offerings of the Bell operating companies, be they regional or local. This push has come from data communications, research, education, and commercial enterprises around the world, which have generally used private, bypass networks built on top of publicly available telecommunications circuits to engineer effective and affordable data communication networks.

The Internet is the premier example of such a private, bypass network in the research and education community. It has evolved since the mid-1970s to become both an operating entity and a testbed for the networking technologies and applications of people involved in research and education in well over fifty countries. It is the network in which the critical mass of people, resources, and performance that characterizes the "information society" has occurred for the first time in human history. It and other networking initiatives that are related to it (initiatives like the Blacksburg Electronic Village, the Big Sky Telegraph in Montana, and Freenet) are delivering real value to their communities on a daily basis and also generating lessons for the evolution of public telecommunication networks.

The second point is that the most important thing being networked is people and not computers. Networks are no longer conceived of as just computational matrices, they are also being conceived of as communication media. The National Research and Education Network (NREN) provisions of the High-Performance Computing and Communications Act of 1991 (PL102-194) make this very clear, and there are versions of the NREN story in numerous other countries around the world. It is encouraging to note that the word *library* appears twenty-four times in PL102-194; the word *publisher* appears only twice, but I am pleased that it appears at all! The public policy process that has brought us to this new conception of networks has been a fascinating one and it is far from over. It is a process that is being driven by a number of basic public interests.

I believe that the public policy constituency that has been the most active and successful to date is the one that conceives of the NREN as a way for increasing the returns on governmental investments in research and education activities by improving the quality of those activities and the productivity of people involved in those activities. Governments at all levels in the United States are the major funders of research and education of all types; therefore, this constituency has made the powerful and persuasive argument that it is good public policy to act to increase the returns on those investments by building the NREN.

I also believe that the vision of this public policy constituency, no matter how compelling and successful it has been to date, does not exhaust the national interest in networking. A second public policy constituency conceives of the national interest in networking to be one of providing a foundation for a post-industrial economy; a third constituency conceives of the national interest in networking to be one of creating a retail paradise for "couch potatoes"; and a fourth constituency conceives of the national interest in networking to be one of facilitating the emergence of a new social order. These other policy constituencies were activated and energized by the NREN public policy process. They represent important points of view that will continue to be brought to bear not only on the NREN but, perhaps even more strongly, on the broader public policy process by which the public telecommunications network will evolve into the national information infrastructure.

THE EVOLVING NETWORKED INFORMATION ENVIRONMENT

Experience with the Internet indicates that the evolving network infrastructure is being used to frame and address three major opportunities and challenges for how knowledge is created, distributed, and utilized in society.

The first opportunity and challenge is modernization, the process by which networks are being used to re-solve problems for which we already have solutions but for which networks offer superior solutions. Clear examples of types of modernization are "projection," as when we provide access to an online system to an entirely new user population via a network, and "substitution," as when we use electronic mail and file transfer instead paper mail and facsimile for correspondence and document exchange. A less clear but extremely interesting and developing third example of modernization is "reorganization," as when documents are printed, after being located in and retrieved from a network server, at the time they are needed using a printer and other equipment owned by a user instead of being printed by a publisher so that they can be acquired and organized by a library in anticipation of being needed by a user.

The second opportunity and challenge is innovation, the process by which the network is being used to solve problems that we have always known we have but for which we have never had workable or affordable solutions. I believe that the greatest innovative effects of network infrastructures and networked information environments will be felt, at least initially, in how we search for and retrieve information. I believe that we will quite rapidly move away from using strategies formulated using Boolean operators to retrieve descriptions of documents and other source materials from massive databases. Instead, I believe that we will use very sophisticated, even artificially intelligent, algorithms that act on our interests and preferences by getting and staying in touch with a large number of servers that provide access to the information that we routinely or occasionally need. I take comfort in the fact that such servers (WAIS, for instance) can already be found on the Internet and that work has already begun on such algorithms (Knowbots, for instance).

The third, and final, opportunity and challenge is "transformation," the process by which the network is being used to solve problems that we did not know we had because we were fully occupied with operating our previous solutions and worrying about our previous problems. The transformational potential of networks and networked information is extremely difficult to frame, let alone to address, until we are much farther along in realizing their modernizing and innovating potentials. Still, I believe that we already have enough experience with networked information resources and services that are highly interactive and that bring graphics, sound, motion, and color together with text and numbers to establish that these networked information environments greatly reenforce the creativity and problem-solving skills of their users. This means that suitably prepared and supported users can accomplish results in many more areas and in much greater numbers than has been possible with the knowledge creation, distribution, and utilization systems that are based upon the communication media that predate networks and networked information. In general, I believe that the greatest transformational effects of network infrastructures and networked information environments will be felt, at least initially, in how well and how widely we mobilize people and information to bring to bear on problems.

CONCLUSION

I would like to bring my remarks to a close by first of all encouraging all of you to start spending much more time thinking about community development, perhaps even as much time as you spend thinking about collection development. I believe that our future rewards and joys will flow to us in direct proportion to our understanding of the communities that we really serve and the needs that these communities really have. We simply cannot do what we need to do in the near- or long-term future with just a collection development policy; we have to go one step further and begin to articulate how our collections serve the constituencies that make up our communities.

Secondly, I need to make you aware of the rather sobering fact that many of the people who track the moving targets and the opportunities and challenges that I have identified believe that the breakthroughs ahead of us are going to occur on the "periphery" rather than in the "center." This is to say that we will have to look to unusual places engaged in what may appear to be risky or even reckless acts to find what works and what does not work in the new networked information environment. It is very important that we spend more time capturing and learning from the lessons that emerge from the periphery that we do dodging or criticizing those lessons.

Finally, I would like to leave you with the four-part leadership strategy of Charles DeGaulle: Never lose the initiative; embrace the inevitable; stay in with the outs; and never get between the dog and the lamppost. This strategy has offered me safe harbor from many a storm on the sea formed by the confluence of information technology and libraries. I hope that it may do the same for you. ■■

The Future Online Catalog: A Single View of Multiple Databases

Walt Crawford

In his presentation, Walt Crawford gave an overview of present and future trends in OPAC design, campuswide and remote access, the impact of standards, different needs of different libraries, and new ideas in searching and presentation.

APPARENT TRENDS IN OPAC DESIGN

According to Crawford, the trends in OPAC designs are a "mixed bag," but those that seem reasonably clear when comparing 1991–92 catalogs with those of 1984–85 are as follows:

- A move toward gutter-aligned, labeled "long" displays.
- Fewer abbreviations and less jargon.
- Less cluttered screens with better use of white space.
- Growing compliance with Common Command Language within command-based systems.
- A strong move away from dual-mode interfaces (with "expert" command-based and "new user" menu-based choices) toward single-mode interfaces, which are usually command-based with heavy assistance. According to Crawford, "users using novice systems stay novice users, . . . but users using command-driven systems tend to use the system to its capacity."
- More related-record searching.

Some questions regarding OPAC design that remain unresolved include:

- What type of labels should be used? What fields should be displayed and in what order?
- Should a single-interface system be menu driven or command based?
- How is user control over the interface assured? Crawford noted a "tendency for some systems to be very clever in making things better for users, but the users don't understand why what they did resulted in what is on the screen."
- What are the significance and worth of graphical user interface (GUI) systems?

GOING BEYOND THE LOCAL CATALOG

Why do libraries need to go beyond the local catalog? Crawford gives three reasons. First, libraries cannot buy everything locally. They need second-level access to the holdings of hundreds of institutions to maintain library service "in the face of problematic collection development decisions." Secondly, libraries need to provide access for scholars to unique information such as archival materials, the history of sound recording, and early printed books. Thirdly, remote access offers a better way to handle abstracting and indexing.

Crawford said that the first rule for remote providers is to be widely accessible. Their systems must also be "usable on the equipment available now, software available now, and networks available now." For Crawford, some of the realities of campuswide and remote access are as follows:

- Few campuses can assure a PC in every room—some can, but only a few.
- Availability of sophisticated end-user

Walt Crawford is President of LITA and Senior Analyst at The Research Libraries Group, Inc. This presentation is summarized by Gail Gulbenkian, Senior Information Officer, VTLS Inc.

computers is much less common than frequently supposed.

• Even where workstations are plentiful, the range of equipment is incredibly broad; there is no *de facto* standard for equipment or connectivity.

• Most libraries still have mostly dumb terminals, which they can't afford to replace.

• GUIs and menu-driven interfaces are slow and irritating if done remotely, unless the heart of the interface is in the client.

Given these realities, Crawford concluded that "the only reasonable short-term solution for remote service providers is VT100 emulation as the common denominator." He added, however, that unfortunately, "VT100 implies less than you might think," especially over the Internet where its use is very inefficient. Crawford's solution? "If efficiency is needed, VT100 emulation means use of full command lines, not tabs."

A proponent of command lines, Crawford believes that "you can prepare a clean, responsive, easy-to-use interface that relies entirely on the command line." Further, by using commands rather than menus, experienced users are not disadvantaged and are never slowed down.

THE IMPACT OF STANDARDS

"Z39.50 is happening right now," said Crawford, but he figures it will take some time to filter down and become stable and broad-based enough to handle all system needs. "Within a few years, we'll see many Z39.50-based systems, where the entire user interface is in the client." Crawford reported that Z39.50 is already moving beyond bibliographic data to serve as the heart of remote services for campuswide information systems.

He also told his audience that Unicode is in place, but unofficially. As Crawford explained it, Unicode is a two-byte universal character set that combines every known printable character set into a single character set. It will allow effective searching and display of CJK and other non-Roman material, which represents a significant amount of information.

LONG-TERM POSSIBILITIES

Crawford said that Z39.50 makes it feasible to provide a single view of multiple databases, and it may well expand to include digitized images and sound, "although one wonders about providing such access over remote links in the next few years." Crawford sees the next step as "multiple unified views of databases to meet different preferences and needs" such as those of blind users, the physically disabled, and non-English speakers. Moreover, as Crawford pointed out, multiple interfaces can also serve user preferences such as icon-based interfaces, graphical interfaces, and pure character systems. "Will future patrons carry their preferred interface with them on diskette?" Crawford asks. "Conceivable, if somewhat unlikely. Library use is not that central to most people."

DIFFERENT LIBRARIES, DIFFERENT NEEDS

Crawford offered a few words about why there is no ideal online catalog for everyone. Even though cheaper hardware is "lessening the power distinction" between small and large libraries, the needs of different types of libraries are different he said. For example, searching techniques should be different for different sizes and types of databases since recall may be more important than precision in one case and vice versa in another.

NEW IDEAS FOR SEARCHING AND PRESENTATION

Crawford pointed out that fancy search techniques go over the heads of most users. For example, most users can't use explicit Boolean searching effectively, he said. People want to know what is going on in a search, and if they don't know, it will be hard to sell them on an OPAC.

Crawford feels more investigation into search methodology is needed, but he cautions against what he calls "enthusiasts." "Be aware that assertion isn't proof, and don't discount the *expressed* needs of the users. It's incredibly arrogant to assume that what people want is irrelevant because we know what they need," said Crawford.

Presenting a case for conservatism, Crawford also cautioned OPAC designers with "radical" ideas to heed the opinions of librarians. "Librarians form the first line of explanation and the last line of clarity. Designers need to work with them, not against them." ■■

Legislative Review on NREN and Perspectives on Home Access

Rick Boucher

I am pleased to have been asked again to address the VTLS Library Directors' Conference. At the inaugural conference last September, I spoke about the goals and purposes of the High-Performance Computing Act, which was then nearing final passage by Congress. As you are all aware, the president signed the bill into law in December 1991, and the federal agencies participating in the High-Performance Computing Program are now implementing its various components.

I know that the National Research and Education Network, or NREN, which is authorized by the Act, is of particular interest to this group because of the enormous promise of high-speed digital communications for improving access to information and for transforming the way libraries function. This morning, I will review where we are and where we hope to be going in the process of establishing the NREN, and I will also address the broader question of what needs to be done to deploy the world's most advanced information network for the benefit of all segments of society.

Just as canals and railroads were the economic arteries of past centuries and interstate highways the arteries of this century, the major channels of commerce in the twenty-first century will be fiber-optic networks, digital switches, and associated software transmitting large volumes of data at gigabit speeds.

The future, ubiquitous network for voice, video, and data communications of all kinds will connect homes, schools, and workplaces. It will constitute an essential ingredient for our future economic competitiveness and will open new worlds of information and services for all of our citizens.

Using the analogy of the highway system, the challenge we face is developing the interstate highways of the information network and developing the exit ramps and secondary roads. These are very different tasks, the accomplishment of which will call upon different roles and require different means of financing on the part of the federal government and the nation's telecommunications industry.

We have taken a major step toward developing the interstate highways for information by enacting the High-Performance Computing Act. It will not be a network in the traditional sense. It will not be a connection of lines, wires, and cables. Rather, it will be a new generation of high-capacity switches and software capable of routing information over fiber-optic lines built and owned by the private sector. The federal roles in creating the information infrastructure will largely be to support development of the enabling technologies and to institute regulatory policies to shape its development.

We will be looking to the private sector to build the exit ramps, secondary roads, and municipal streets of the network, but federal policies must be altered to allow that to happen. I will address that point subsequently, but first will make some observations about policy issues associated with NREN development.

Following enactment of the legislation, this spring the Subcommittee on Science began its oversight activities for implementing the NREN with hearings on the present management of NSFNet. We started by reviewing the administration and operation of NSFNet, since current practices will strongly influence evolution of the NREN.

The members of our subcommittee were frankly impressed with the pace of growth in NSFNet use and the increasing diversity of its users. In the past five years, NSFNet has advanced from serving a few supercomputer centers in a narrow research mission to serving millions of scholars and researchers in scores of industrial labs and in most of the universities and federal labs. It also connects thousands of high schools and hundreds of libraries. Traffic on the NSFNet is growing at the rate of 11 percent per month. The hearing also revealed that the federal investment has leveraged by thirty to one or more the investment by states, industry, and universities in developing the network infrastructure.

One issue we addressed is the treatment of commercial network providers who use the NSFNet. There is now a lively competition in

---

Rick Boucher is Representative of the Ninth District, Virginia. This text was read by Peter Young, Director, National Commission on Library and Information Services.

the private sector for the provision of network connections. In this new environment, NSF will conduct a new competition for management of the backbone network. In the new agreement, various management functions will be separately awarded as distinct from the single contract under which the network is now managed. This approach by NSF, which is still being formulated, does not satisfy all potential providers of backbone services. But it appears to have been developed through consultation with the commercial network providers and the regional networks.

A second issue raised at the hearing was whether NSF should continue to distinguish between network traffic that has a commercial nature on the one hand and traffic that is purely for research or educational purposes on the other. Some witnesses characterized the NSF policy of acceptable use of the network, under which charges are imposed for commercial traffic but not for research or education-related use, as hindering the development of appropriate information services and unnecessarily restraining the volume of network traffic. NSF stated that is has no choice but to impose the policy due to restrictions of the foundation's enabling legislation.

The Subcommittee on Science has taken legislative action to modify the statute which NSF has interpreted as requiring the imposition of the current acceptable use policy. The modification will give NSF the necessary authority to remove the policy if the agency determines that removal will tend to improve the level of network service for research and education activities. The measure has been passed by the House and is awaiting action by the Senate.

The last issue raised at the hearing addressed management of the NREN. Concerns have arisen that interagency coordination through the auspices of the Office of Science and Technology Policy is inadequate to ensure steady progress toward the NREN. Management of the NREN is a concept that brings complexity with no good models. The constituencies that need to be represented in the governance of the network include higher education, federal agencies, industry, states, and communities. The structure that can best provide this management was not readily apparent.

It was evident to the subcommittee that there are many more questions than answers about the best way to achieve a successful transition to the NREN. In crafting the High-Performance Computing Act, Congress provided broad authority and a general template for the High-Performance Computing Program. The legislation left open many details for implementation of the program, with the expectation that they would be addressed by the agencies charged with carrying it out.

The legislation does impose a requirement to spur necessary planning activities. By early December, OSTP must report to Congress on topics related to establishment of the NREN. Among the topics to be addressed in the report are possible funding mechanisms for operation of the network and procedures for providing commercial services over the network. Also to be addressed are the means of protecting copyrighted material distributed over the NREN, and how to ensure the privacy of users. The network cannot grow without these policies and protections in place.

The deadline for this report is fast approaching and our committee expects the administration to give careful consideration to its preparation. We are expecting more than short and bureaucratic answers to the questions that must be addressed in covering the topics identified. The report should be developed in a broad, consultative process involving the many interested communities—as was the case in the development of the original report recommending the program five years ago. I know that efforts have been made by the various constituencies of the program, including library groups, to provide their views on implementation of the NREN. It remains to be seen whether the Administration's final plan will incorporate these views.

I would like to turn now from the interstate highways of the information network to the exit ramps and secondary roads, where an entirely different kind of legislative approach will be required. The importance of beginning to address this aspect of the national network is clear when we consider the plans of our economic competitors.

The Japanese have made a commitment that by the year 2015, there will be deployed throughout the nation of Japan fiber-optic cables into every home and business, school, and research laboratory in that country. And they have set aside the financing that is necessary in order to accomplish that fiber-optic deployment.

We call that deployment "deployment over the last mile," meaning the distance from the telephone company's last switch to the residence or business of the end user. And that is the most expensive part of network deployment. It is by far the greatest distance in terms of deployment when all of those various sectors are added together, and the most costly.

I am proposing that we do that in the United States, not by spending public dollars, as we have in developing the NREN, but by giving the private sector sufficient incentive to deploy the network on its own. Debate has been underway in Congress on major revisions to the Cable Act of 1984. That legislation broadly deregulated the television cable industry. It basically said that no level of government had the authority to set cable rates. It also contains some prohibitions that prevented logical competitors from getting into the business. Those prohibitions are known as Cross-Ownership Restrictions of the 1984 Cable Act and say that no telephone company may offer cable service within its telephone service areas and that no broadcaster may offer cable service within its broadcast service area. So that the most logical competitors for the cable industry are barred by federal law from getting involved in competition with the dominant cable provider.

The consequence of that is the industry operates as a monopoly throughout the country. We traditionally have sanctioned monopolies in those instances where there were economies of scale that required a single provider of the service. Electric utilities and telephone companies have been classic examples of that, but even in those industries today we are seeing competition. Independent power producers are generating electricity and cable companies are starting to offer telephone services in some localities across the country.

Yet we have this unnatural monopoly that is protected and guaranteed by federal law. I think the time has come to break up that monopoly. The time has come to give the telephone industry a fair opportunity to compete and to offer cable television service.

Now, as a side point, let me mention that there are some consumer benefits that would flow from that: lower cable television rates; better cable television service. I have constituents that say that every time it rains their cable television service goes out, and they say that often it does not come back on until it rains again. When they call the cable company, no one answers, and it takes days to get the repairs made.

Consumers are complaining about a lack of programming alternatives. I think if you have real competition, there would be a virtual explosion of programming with the effect that the number and quality of programs on all of the packages available in the market would be superior to what you have today. This is a way to get cable service out to rural America.

So there are a number of consumer benefits that would flow from passing legislation. But the major benefit the country would receive would be that telephone companies would then have the financial incentive to deploy fiber-optic technology over that last mile and get fiber-optic cables extending into homes, businesses, schools, and research laboratories throughout the country within a very short period of time.

How soon would it happen? In testifying on this bill about two years ago, Bell Atlantic, which is one the more forward-looking of the telephone companies in the country, indicated that if all they could provide over the telephone lines was what they call "plain old telephone service," or POTS, it would probably take forty years to get fiber-optics deployed over the last mile universally within the United States. But if they had the power to offer cable TV service, they could accomplish that deployment within about half the time, or twenty years.

Now, twenty years from today would allow us to beat the Japanese by about five years. We could get fiber-optics deployed into research laboratories, homes, and businesses across this country by about the year 2010. And that would give businesses in the United States access to high-speed data transmission capabilities at gigabit speed before it happens in Japan. That would be very important for us to do, and we could do it without spending a penny of public money simply by taking the brakes off the telephone industry today. So I am a strong advocate of doing that, and we are working daily in Congress to get that accomplished.

Much of the excitement associated with development of the modern national information network is the discovery of new and unanticipated uses for it. Major new technologies have often transcended the initial expectations held out for them. In his book *Perspectives on Technology*, Stanford economics professor Nathan Rosenberg reminds us that, "railroads were originally thought of as essentially feeders to canals and other forms of water transportation."

Many experiments are now under way to discover how to use the national network. This is certainly the case for activities supported under the High-Performance Computing Program. And we have just heard from Bob Heterick about an example from my own congressional district which I have followed with great interest. I think the Blacksburg Electronic Village may serve as a model for the information community of the twenty-first century.

The national network is an enabling technology that will unleash the creative energies of all segments of society. We must seek policies that will facilitate collaboration between public and private sectors for creation of the world's most advanced information network.

The expertise represented by the library community will be needed to assist in this undertaking. I appreciate having had this opportunity to speak to you and encourage you to make your views known to the Subcommittee on Science as we strive to put the national network in place. ■■

Creating the Future with the Initiators of Change: A Panel Discussion with Frank Bridge, Walt Crawford, Bob Heterick, and Peter Young, Moderated by Vinod Chachra

Summarized by Barbara L. Scheid

The following is a summary of what the panel members view as the most pressing challenges facing the information and library professions today. After the discussions the session was opened for questions from the participants to the panel members. Some of the questions were addressed by more than one panel member.

Bob Heterick's primary concern is for the remote, poor, underfunded public libraries. He stated that their problems extend well beyond access to the network. Specific to their access to the NREN, Heterick reported, "The NREN is not directed towards becoming a national information infrastructure, which is what is required to bring the information infrastructure to small communities."

Heterick argued that more efforts like the Blacksburg Electronic Village are needed in order to bring the importance of NREN to the attention of the legislature. Heterick reported that statistics show that approximately 300 universities leverage the federal investment for network technology by a factor between 30 and 100. To date, state legislatures have been absent from this network. "In order to bring the network to rural and/or small communities, and the whole world for that matter, we need to leverage state and local investments. In order to do this we need to keep government officials informed," said Heterick.

Peter Young reported that in the 1970s and 1980s we were concerned about automating the functions in the library. The library application of the software developed by companies like VTLS is proof that libraries and librarians have long been able to handle the technology and use it effectively. The crises, challenges, and opportunities the network will bring and how and where librarians will get the skills and education necessary to handle the new information resources and services are issues which concern Young.

Young stated that what "bothered" him was that the library profession, which he had been in for twenty-five years, is not struggling to get to the next step, but is primarily worrying about how it lost the opportunity for advancing the institution. The fact that librarians are now being urged to become more involed in the political atmosphere at both the national and regional levels is encouraging to Young. He recommended this political involvement come in the form of networking at technical and human resource levels.

The issue that concerns Walt Crawford is the tendency for people to accept one solution to very complex problems where many solutions are possible. He stated that there are no simple answers because there are no simple questions. He believes libraries must be able to continue to provide traditional services while they investigate future services.

Frank Bridge cited his research showing that 92 percent of the systems installed in 1991 were in libraries with between one and four terminals. The typical users of these systems are small school and public libraries, stated Bridge. The majority of the librarians that run these libraries are nondegreed librarians and may not even have a bachelor's degree. In general, these people are overwhelmed by automation and know little about connecting to the NREN. He stated that a large knowledge gap exists between the "information haves and the information have nots." His main concern is educating school and public librarians about automation and the NREN.

Question to Peter Young: We heard earlier from Frank Bridge and from Paul Peters that we must work closely with the state and local governments in order to get our small community libraries hooked to the NREN. If the state and local governments are not aware of the advantages of accessing the NREN, why will they expend funds to support it? What initiative is the federal government taking to educate the state and local governments?

The NREN legislation passed in December 1991 has a commitment in it to the research and education communities. The legislation does not address the economic, social, cultural, or equity issues we face at the local levels. Referring back to Representative Boucher's paper, Young noted that Congress is saddled with reducing the national debt. Therefore, support for the network at the state and local levels has been left up to the private sector. The good news reported by Young is that by investing in research and education projects the federal government is by default providing the backbone of the infrastructure for the network.

In envisioning the NREN we must decide what we want the network to do for our communities and what communities we serve. Allowing school children access to the entire database at the Library of Congress is an intriguing and common model of how local libraries would use the NREN. Although Young said he also had expected this model, he has lived it and does not recommend it.

Young argued that we need an agenda that formulates a national plan which addresses national, state and local needs for building a digital network over the next two decades. Where will the money come from which will amplify the federal investment? He suggests we consider shifting some of the national defense funds to this cause. This is a necessary transfer because national security issues are much more local than they have been in the last forty years. The investment to be made should certainly be made in the technical infrastructure, but more importantly we need to invest in the human resources coming out of our schools and universities, stated Young. These people are the future employees at our companies that must be competitive in the international arena. He said he was encouraged that the Congress had begun to address the legislative issues and was hopeful that future administrations would provide the necessary investment in human resources.

Question for Walt Crawford: How can the library deal effectively with electronic full text?

The models are really not in place for effective integration of full-text sources, replied Crawford. Based on what he has seen so far, "effective" is the key word. He said that scrolling endless lines of ASCII text is not going to serve the future patron very well. He also believes that printing a document at three cents a page is not very useful to the majority of patrons. Setting aside research and possibly chemistry libraries, the majority of materials patrons want access to can be found in their libraries, he said. He believes the current searching and retrieval methods are satisfactory for a large majority of users for the immediate future.

Followup question to Walt Crawford: Can we print ASCII text files to disk and then use the information for current or future research projects?

Crawford believes that while this is possible and one of the better solutions, this use still is not the norm.

Question to Bob Heterick: Is there a way to reduce line charges so that branch libraries can participate in consortia where there is one computer that serves six main libraries that may also have several branch libraries?

Heterick reported there is no short-term solution to this problem. It is a well-known fact that it is cheaper to call intrastate than it is to call interstate. Both FCC regulations and the lack of competition within the "lattice" need to be examined. Heterick observed an anomoly in that if we base the communication network structure on the idea of a voice telephone call, then it would be "prohibitively" expensive to carry data and video across these lines. But if we base the cost of telephony and the telephone company's existing infrastructure on the cost of video, then the telephone call becomes free. He said, "We have a very artificial pricing system that has no basis in bandwidth requirement and is propped up by state and federal regulations." We must begin to educate the people at state and national levels who work with the legislation that adopts the regulations regarding communication systems.

Question to Bob Heterick: Frank Bridge talked about the knowledge gap. Do you think ideas such as the Blacksburg Electronic Village will create a chasm between communities like Blacksburg, Virginia, that will have the knowledge and the access and communities like Princeton, Kentucky, that will not?

Heterick said the question was not easily answered. Inhibiting one community from the network will not help the other communities with no access. The reality is that there will always be some information-rich communities and some information-poor communities. This issue is one out of many social issues that we spend "three-quarters of our domestic budget researching and addressing." Heterick was glad to report that "at least network issues are more affordable to address than other social issues."

Walt Crawford added that the real emphasis coming from the passage of the NREN bill is that the information-rich get richer. It is true that preventing access to the NREN will not help anyone. Crawford said, "It only becomes a social disaster if the money gets siphoned away from poorer libraries that need the money in order to provide basic services." Crawford was quick to say that he did not see this happening, only that it should not happen.

Peter Young summarized that if we leverage the federal investment of the NREN by investing at the local level in experiments such as the Blacksburg Electronic Village, presumably we will be able to identify models and build upon our successes. Young believes that the NREN is not meant to better what we are currently doing. He agreed with Paul Peters' statement that the NREN will change the nature of communication between individuals. He reiterated Peters' point that at this level, NREN is not about the benefits it affords the citizens of Blacksburg, but about the communities that are going to be formed. He said what has begun to happen because of Internet and the NREN "is a loosening of the glue" that has been holding very traditional organizational structures together for decades. The traditional academic structure at our universities will begin to change when individuals from one college become interested in the information and projects happening at another college. The NREN will begin to open communication between all types of citizens so all types of communities will be created.

Heterick added to the complexity of this issue by reporting on the funding traditions in the state of Virginia. Since 1985, Virginia state funding has increased approximately 40 percent for libraries, 50 percent for higher education, 60 percent for K-12 education, 206 percent for Medicare and Medicaid, and 162 percent for prison facilities. Based on these figures, it is evident that any funding for public good will be under "extreme" budget pressures over the next two decades. Heterick believes that libraries will not be funded and the network infrastructure at the local level will not be funded by the federal or state governments. "The strategy must be that we take a fair piece of our income and devote it to trying to be a close follower of some of these activities." Unfortunately, "we have to be realistic and decide that we have to reduce current services which we all know are already too low," said Heterick. ■■

Tutorials

Simple and Inexpensive CD-ROM Networking: A Step-by-Step Approach

Mark Cain

Many librarians, especially those at smaller institutions, are intimidated by the prospect of installing a CD-ROM network. Networks are reputed to cost too much and to require someone on staff with considerable technical expertise to install one. Yet establishing a small network for CD-ROM access need not be a daunting task. The purpose of this article is to describe a low-cost, upgradable networking solution and then take the reader step-by-step through the installation process. The approach to the topic is elementary. No special technical knowledge or skill is assumed.

For purposes of illustration, this tutorial assumes the following situation: the library has three IBM-compatible microcomputers, each with its own CD-ROM drive and each running a different product. These three computers will be networked, allowing library users the ability to search any of the three databases from any of the three machines. Simultaneous searching of a single CD-ROM product will also be possible.

STEP ONE: TAKE AN INVENTORY OF ASSETS

If the library has a single CD-ROM product, there is probably at least one staff member who has a basic knowledge of the search software and of DOS. (A DOS-based network is assumed, since most library CD-ROM products run on IBM compatibles.) Such an individual is the logical candidate for installing and maintaining the network.

More than likely, the library has several CD-ROM products, or its staff would not be considering a network in the first place. If the library has at least two microcomputers and one CD-ROM drive, it may not actually have to buy any equipment, except the few adapter boards that will have to be installed in the back of the microcomputers.

STEP TWO: LEARN ABOUT THE COMPONENTS OF A NETWORK

What is a PC network? In its simplest form, it consists of microcomputers, cables, and an adapter board (also called an adapter or adapter card) in each machine to which the cables are attached. Together these form the data highway of the network, the physical medium that allows information to pass from one machine to another. Some of the computers hooked together are workstations. One (or more of them) is a server, providing access to network devices (printers, CD-ROM drives, hard disks, and so on) and/or software. Finally, there is the network operating system (NOS); the NOS allows all the components to work together.

In a CD-ROM network, the server provides the workstations on the network with access to one or more CD-ROM drives. These drives are physically attached to the server, but the workstations can access them as if the drives were directly and physically attached to each of them as well.

STEP THREE: SELECT A NETWORKING SOLUTION

For a simple network, Artisoft's LANtastic is a good choice. LANtastic 4.1 cost only $99 per network at the time of writing. (Version 5.0 has just been released.) If the adapter boards are bought from Artisoft as well, a network may be installed for about $350 (list) per workstation. With the usual discount from a computer dealer, this hypothetical three-station network will cost less than $1,000. Adapters may be purchased from a source other than Artisoft, but then LANtastic's price increases to $99 per node. Even with the higher software cost, a better overall price may be obtained if adapters are bought from a third party, though the neophyte may like the security of buying hardware and software

Mark Cain is Information Technologies Librarian, Xavier University, Cincinnati, Ohio.

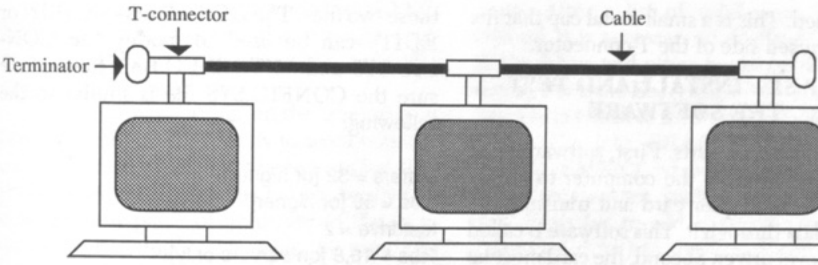


Figure 1. Cabling Configuration.

from the same vendor. LANtastic has built-in support for CD-ROM drives, while, as of this writing, other CD-ROM networking solutions require the purchase of network software (such as Novell NetWare) and a separate CD-ROM management program such as OPTI-NET or CD-Connection. In addition, the LANtastic software does not require a dedicated server; one or more workstations can perform double duty as servers. At least initially, the library will not have to invest in additional equipment. Every machine on the network that has a CD-ROM drive may be configured as a server, though this is not the most efficient network architecture.

In choosing adapter boards, look for products that are Novell compatible, so that if the library migrates from LANtastic to the popular Novell NOS, the same adapters can be used. For simplicity's sake, the boards used in the present example will be Ethernet boards, using coaxial cable (as opposed to twisted pair/10BaseT).

STEP FOUR: INSTALL THE ADAPTER BOARDS

Some people are intimidated by the thought of opening up a microcomputer and working on its internals. As long as computer and monitor do not share the same box (as they do with the Macintosh Plus, SE, and Classic series), there is nothing dangerous or subtle about the operation. Make sure the computer is unplugged before it is opened. Removing a few screws in the back is usually all that is necessary. The cover will then either slide or lift off.

With the top off, several empty slots near the back of the machine should be visible. The adapter card should be installed in the smallest slot into which the board will fit. Behind each slot is a silvery metal plate held in place by a screw and covering an opening in the back of the computer. Remove the screw and the plate. Take the adapter board, and insert it in the slot. There will be some resistance, but gentle pressure and a slight rocking motion will work it into place. Once the board is firmly seated in the slot, the screw can be used to secure it to the computer, and the casing can be reattached.

The computer generally knows how to find the newly installed adapter board. Sometimes, though, the address of the board (called the I/O address, or input/output address, which is preset at the factory) can conflict with some other peripheral that has been installed in the computer. The I/O address on most boards can be changed if necessary; the manual that comes with the board will provide instructions.

There are one or two other things that could go wrong, like having an interrupt (IRQ) conflict or a nonstandard bus in the PC, but usually this will not happen. As a general rule, maintain the factory-set defaults. Change them only if a problem occurs.

STEP FIVE: ATTACH THE CABLING

How the cabling is attached depends upon the type of topology, or system architecture, selected for the network. The present example employs the simplest: a bus topology. With this architecture, one machine is daisy-chained to the next. With each adapter comes a small T-shaped piece of hardware, called a T-connector. This attaches to the interface of the card, splitting the signal, allowing two cables to be plugged into the adapter card, one coming from the previous microcomputer, the other going to the next computer on the chain. Graphically, the configuration resembles the illustration in figure 1.

At each end of the chain, a terminator must be attached. This is a small metal cap that fits on the unused side of the T-connector.

STEP SIX: INSTALL AND TEST THE SOFTWARE

This step has four parts. First, software must be installed to allow the computer to recognize the new adapter card and transmit and receive data through it. This software is called the low-level driver. Second, the card must be configured to work with the NOS, in the present instance the LANtastic NOS. This is accomplished with a high-level driver. Third, the NOS must be installed and configured. Finally, each machine must be designated either a server or a workstation.

A LANtastic installation combines all four into a single process so transparent that the user is not even aware he or she is accomplishing them. For installation, the LANtastic software diskette is placed in the A drive. The individual performing the installation then types:

```
a:
install
```

The software guides the user through the installation process. It will ask for a unique name to assign to each machine.

After the machine is named, a screen of options comes up. Only one critical choice must be made at this point: Will this machine be a server or a workstation? Remember that a server will provide access to a CD-ROM drive, so if this machine has a drive attached to it, it must be designated a server. If not, it should be designated a workstation. Having made this choice, use the down arrow to move to the bottom of the page. Select SELECT TO INSTALL and press Enter. Go through this process with each machine on the network.

LANtastic is stored in a subdirectory called \lantasti. There is a batch file in this subdirectory, called STARTNET.BAT, for booting LANtastic. For a variety of reasons, however, it might be preferable to have the network software boot automatically on start-up. This can be accomplished by making changes to the CONFIG.SYS and AUTOEXEC.BAT files. These should be familiar to anyone with CD-ROM experience, because Microsoft's CD-ROM extensions—which are needed to access most CD-ROM databases—require these two files. The DOS editors—EDLIN or EDIT—can be used to modify the CONFIG.SYS and AUTOEXEC.BAT files. Make sure the CONFIG.SYS file is similar to the following:

```
buffers = 32 [or higher]
files = 50 [or higher]
lastdrive = z
fcbs = 16,8 [on servers only]
```

The AUTOEXEC.BAT file for each workstation should resemble the following:

```
path=c:\;c:\dos;c:\lantasti
[LANtastic needs to be in the path]
aex
ailanbio
redir workstation name logins = 5 [or something similar]
net show
```

For servers, the AUTOEXEC.BAT file should have lines in it that look like the following:

```
path = c:\;c:\dos;c:\lantasti
aex
ailanbio
redir servername logins = 5
\mscdex.exe /d:mscd000 /e /m:8 [/e is optional]
server
```

It is important that the Microsoft CD-ROM extensions (the line that has "mscdex" in it) come between the lines that begin with "redir" and "server." If they do not, edit the file (using the DOS line editor EDLIN or a word processor) so that the lines are in the proper order.

STEP 7: INITIAL NETWORK TESTING

Reboot the machine and watch the screen carefully for any error messages. Make note of these messages, because they may help you in diagnosing problems. This article cannot anticipate all that might go wrong in any particular installation process, but some of the most common problems are:

- The cables, T-connectors, or terminators are not attached securely.
- The adapter card is not properly seated in its slot within the computer.

• The CONFIG.SYS and/or AUTOEXEC.BAT files do not contain the requisite commands, or they are not in the proper order. Try reinstalling the software.

• The jumper settings on the adapter card are not configured properly to work with the computer into which it has been installed. The problem might be that the card is conflicting with another piece of hardware in the computer—for example, it is using the same I/O address or IRQ, or the problem might be that the computer has a nonstandard bus. These are both fixed by changing some jumpers on the card. Consult the manual that came with the adapter.

• The software that drives the card is not configured properly. Try reinstalling the software.

Other problems may require a call to the technical support staff at Artisoft or at the company from which the adapter card was purchased.

STEP EIGHT: CONFIGURE THE SERVER(S)

Configure the server(s) to recognize valid users and to make CD-ROMs available to them. For each machine designated as a server, a list of valid users must be created. Each CD-ROM drive must be designated as a shareable resource. After typing "net_mgr" <Enter> and displaying a menu, select INDIVIDUAL ACCOUNT MANAGEMENT. The network administrator creates a list of users with access rights to the server at this point. Note that the names of the users do not necessarily have to be the unique names given to the other servers/workstations. One approach is to make the user names the same as the names of the databases that will be accessed over the network. For example, if the server has ERIC mounted on a drive, create a user named "Eric." Set the maximum number of concurrent logins (simultaneous uses) equal to the number allowed by the network license. (Negotiating licenses is an important detail. Databases obtained through a single-use subscription should not be networked until a new agreement is obtained from the database provider.) Two benefits accrue with this approach. One, as implied above, the server can help enforce the network license. Two, the system will record each login to ERIC in its log file, thus maintaining important data that can be used for statistical purposes. After a list of valid users has been created, it is necessary to designate the CD-ROM drives and other resources for shared access. Press the Esc (escape) key one or more times to get back to the MAIN FUNCTIONS menu. Go into SHARED RESOURCES MANAGEMENT, where there appears a list of resource names, such as the server's hard disk. Press Ins to add a resource (such as a CD-ROM drive) to the list. Name the item and give its local name or "true path"; the first CD-ROM drive is typically the D drive. Provide this information, then press Enter. Press Enter again to display the detailed information for this resource. Make sure the line that says "CD-ROM drive: No" is changed to "CD-ROM drive: Yes" by using the down arrow to highlight the line and pressing Enter.

STEP NINE: INSTALL THE SEARCH SOFTWARE

The CD-ROM search software probably is already installed on the server. Installation across the network may be as simple as logging into the server from another machine, making the server's hard disk accessible to the workstation (see next section for the necessary commands), and then copying the software from the server to a subdirectory on the machine. This option works fine for Information Access Corporation (IAC) databases. For SilverPlatter and Dialog databases, it may be necessary to log into the server, make the CD-ROM accessible to the workstation, and then follow conventional installation procedures. UMI Proquest installations are more complicated, but their technical support staff can help.

Note that the search software is being installed on each workstation. It is possible to store the software on each server's hard disk and have each station load the program from the server. Because this example uses nondedicated servers, these machines already will be taxed by providing access to the CD-ROM drive at the same time they are functioning as workstations. While installing the search software on each workstation is more trouble, this approach makes fewer demands of the servers and should help with overall performance.

STEP TEN: INSTALL A MENU AND WRITE BATCH FILES

LANtastic provides a menu-driven method for connecting to servers and their resources, but to make network use effortless install a menu from which library users can select their databases. There are many products from which to choose, e.g., Saber Menu, Direct Access, and Automenu.

When an option from the menu is selected, a DOS batch file is run. The purpose of this batch file is to log into the server, make the CD-ROM available to the workstation, boot the search software, and, when the search session is finished and the user quits the software, redisplay the menu. Here are the key lines of a sample batch file for the SilverPlatter databases, as they might run under Automenu:

```
net login \\SERVER ERIC
(Remark: Log into the server named "Server"
    in order to search ERIC.)
net use d: \\SERVER\D-DRIVE
(Remark: Where "D-drive" is the name given
    the CD-ROM drive designated in the Net
    Manager software as a shareable
    resource)
cd\SPIRS
(Remark: Changes to the SPIRS subdirectory)
call SPIRS
(Remark: Boots SilverPlatter software. The
    "call" will have the computer return to this
    batch file when it is done searching.)
net logout \\SERVER
(Remark: Logs out the workstation)
cd\menu
(Remark: Returns the workstation to the
    menu subdirectory)
auto
(Remark: Redisplays the menu)
```

The command to run the menu may be added to the end of the autoexec.bat file so that the menu displays each time a workstation is powered up.

STEP ELEVEN: LEARN SYSTEM ADMINISTRATION PROCEDURES

Learn procedures that will enhance the functionality of the CD-ROM network and provide some security.

- Create a user named "Supervisor" and give that user access to server hard disks. This will make things much easier for the network administrator. Access to a server's hard disk will let the administrator connect in from a workstation and copy files across the network.
- Limit the access other users have to server hard disks. This is easy to do and provides some network security.
- Run DOS 5.0 or higher on workstations and servers. With DOS 5.0, it is possible to load the operating system, device drivers such as the Microsoft CD-ROM extensions, and TSRs (terminate and stay resident programs) into memory above 640K, leaving more system base memory available for search programs.
- Use the menu on each workstation to block access to DOS. While hardly foolproof, passwording access to DOS will prevent most users from deleting files from or copying files to the station's hard disk.

STEP TWELVE: START PLANNING ENHANCEMENTS

The configuration described above, where servers perform double duty as workstations, will indeed work satisfactorily for small networks, but it is not the fastest configuration. As additional resources become available, consider making the following improvements:

- Install a dedicated server: Buy a fast microcomputer with a large amount of RAM, attach all networked CD-ROM drives to it, and move the machine out of the public area. Then reconfigure the other computers as workstations instead of servers.
- Replace older CD-ROM drives with newer, faster ones.
- Bring in separate CD-ROM management software, such as OPTI-NET, to manage CD-ROM access, leaving LANtastic to manage only the network. Performance will improve considerably.
- Add more workstations.
- Add more databases and drives.
- Consider dial-in access. There are three or four basic ways of providing this service, but what dial-in access essentially requires is a microcomputer to be dedicated for each remote session. An inexpensive approach might be to buy a computer, install a modem in it, attach a phone line to it, and place it on the network. The machine would need to be taken over by the computer dialing into it, so software like pc-ANYWHERE would be needed. More elaborate (and expensive) dial-in solutions are available from companies such as Virtual Microsystems (V-Server) and Logicraft (OmniWare, 386Ware, 486Ware). ■■

SALBIN: PC Software for Accessing Internet Resources

Richard Gartner

JANET, INTERNET, AND SALBIN

November 1991 was an important time for libraries in the United Kingdom. For the first time it became possible to gain direct access to the vast information community known as the Internet. During this month the X.25-based academic network JANET was opened up to the TCP/IP protocol, allowing any institution with the appropriate hardware an open link to Internet services. Although the Bodleian Library at Oxford University had been using JANET for some time to provide readers with access to remote catalogs and data services within the United Kingdom, it soon became clear that the Internet would dwarf the resources the library had been able to make available previously. How to handle this influx of new information sources and how to make them easily available to readers became major challenges.

At the Bodleian, readers have had access to JANET-based resources with the help of a personal-computer software package known as SALBIN. Edinburgh University Library designed this package to provide a menu-driven interface to remote data services. SALBIN is a very user-friendly system; users select the information service they want from a menu, and the software handles all connection procedures. The user disconnects from a remote catalog or database by pressing a single function key. SALBIN also incorporates routines to handle occasions when it does not prove possible to make a connection, informing the user of the problem and returning to the opening menu. SALBIN is configured almost entirely by simple text files, similar to those employed by Kermit, and the language used is very straightforward and takes little time to learn.

With the advent of Internet access, it was logical to extend the SALBIN service already offered to include the new range of services that had suddenly become available. SALBIN allows the library to present the same interface to the newly expanded reservoir of information, integrating Internet services with those on JANET and the European X.25 network IXI into a seamless whole. The new service—known by the acronym BARD (Bodleian Access to Remote Databases)—provides access to 360 online catalogs and forty databases and has proved instantly popular with users.

SALBIN does not run the TCP/IP protocol directly but requires an asynchronous link to an IP service, such as a mainframe with a TELNET access facility, and this limitation may be a drawback for some libraries. At Oxford, it is possible to use the local Gandalf data network, which has access to a Nova server from which TELNET calls can be made. Using the Gandalf node proved especially useful, as the same system provides X.25 access to JANET and IXI. Note, however, that access via this link, which has a speed of 9,600 baud, is slower than running TCP/IP directly. The SALBIN software costs only fifteen pounds sterling and will run on very low grade PC compatibles (256K RAM suffices), potentially providing a way of recycling otherwise obsolete equipment.

SETTING UP SALBIN ON LIBRARY MICROCOMPUTERS

Obtain the SALBIN Software

The only ftp site currently offering SALBIN is the HENSA public-domain software service operated by Lancaster University, United Kingdom. One can obtain a copy by anonymous ftp from MICROS.HENSA.AC.UK. The file name is MICROS/IBMPC/DOS/F/F144/F144SALBIN.BOO. The archiving utilities DEBOO.EXE and PKUNPAK.EXE or ARCE.COM are necessary to unpack SALBIN, which should be stored in a separate directory. DEBOO.EXE and PKUNPAK.EXE are available on HENSA, in the directory MICROS/IBMPC/DOS/TOOLS, under the file names DEBOO.MSG, DEBOO.BAS, DEBOO.BOO, and PKPAK.BOO. The ASCII file DEBOO.MSG gives full extracting instructions. All of the above files with the exception of DEBOO.MSG should be downloaded in binary format. The ARCE package

---

Richard Gartner is Pearson New Media Librarian at the Bodleian Library, Oxford University, United Kingdom. Questions regarding SALBIN may be directed to the author by e-mail at the Internet address GARTNER@VAX.OX.AC.UK.

is widely available on software archives, including HENSA (under the file name MICROS/IBMPC/DOS/E/E638/E638ARCE.BOO).

Having retrieved a copy of SALBIN, the user should register it to receive a full manual. To register, send fifteen pounds to SALBIN Office, Edinburgh University Library, George Square, Edinburgh, EH8 9LJ, United Kingdom.

North American users should request delivery of documentation by air mail, as it is much faster than surface mail.

Obtain Information on Available Internet Services

Several good sources of information are on databases available on the Internet. The TSR utility HYTELNET, a well-known hypertext system, provides an extensive list of library catalogs and information services available on the Internet. HYTELNET is available via anonymous ftp from ACCESS.USASK.CA in the directory PUB/HYTELNET/PC/. The file name of the current version (6.3) is HYTELN63.ZIP. A copy of PKUNZIP.EXE is required to extract this utility (which currently extends to more than a megabyte in size). Copies of PKUNZIP.EXE are widely available on software archives, including the HENSA archive listed earlier. Both files should be downloaded in binary mode.

Explore the SALBIN System

The next stage is to study the SALBIN manual and to explore the files that make up the complete package. All the files configurable by the user are simple text files that can be altered with any standard text editor. The two most important types are those with .MNU extensions and .CNT extensions. The former define the menus presented to the user; the latter are the script files handling the connection to and the disconnection from each service. It is, of course, necessary to construct a separate .CNT file for each service, since each remote site has different logon and logout procedures. SALBIN also allows one to construct a separate help screen for every service by editing the text files with .HLP extensions.

A few files control the central operations of the system and set parameters such as baud rates, com ports used, screen colors, etc. The most important is DIRECT.PAD, which includes the initial sequence of commands necessary to access the TCP/IP service and to disconnect from it. These commands are run when starting up and closing down the system. This file also contains subroutines necessary to disconnect remote sessions, should normal logout sequences not function as expected. Getting to know the system is quite straightforward, and the brief manual explains everything in detail.

Put the System Together

The Bodleian system was constructed in three basic stages.

Configuring the Central System Parameters

The first stage was the configuration of the central system files, described earlier, that set basic, systemwide parameters. The most important of these are the connect/disconnect sequences in DIRECT.PAD, which need to be constructed with care and tested thoroughly.

Designing a Menu Structure

One can structure the information service menus by country, by type of system (e.g., catalogs, bulletin boards, databases), or in various other ways. For its large number of offerings, the Bodleian decided to make an initial distinction between library catalogs (which are then listed by country) and other information services (listed alphabetically). Some countries have so many catalogs that they have to be further divided into submenus. Usually this was done alphabetically, although in the case of the United States a division into individual states was most convenient.

Writing the Script Files

A separate .CNT file, with specific logon and logout procedures, must be created for each information service. The SALBIN package comes with a number of sample script files, although these were designed specifically for the University of Edinburgh's system, so they cannot be run directly. These sample files, along with the manual, are useful guides to the scripting process. Figure 1 shows a sample file.

The CONNECT and DISCONNECT scripts are structured the same way in all files: The SALBIN package sends a command to the remote system and expects a response (with a timeout specified). If the expected response is forthcoming the script proceeds

```
SYSTEM:     "University of Canterbury"
HELP:    canterb.hlp
CONNECT:        send "tcp\r"
        expect "telnet" timeout 15 else skip "ABORT TELNET"
        send "open cantva.canterbury.ac.nz\r"
        expect "Username: " timeout 60 else skip "ABORT TELNET"
        send "OPAC\r"
        expect "Connected." timeout 30 else skip "ABORT TELNET"
        send "\r"
        succeed
        end
DISCONNECT:  send "QUIT\r"
        expect "Off?" timeout 20 else skip "ABORT TELNET"
        send "y\r"
        expect "telnet" timeout 30 else skip "ABORT TELNET"
        send "logout\r"
        succeed
        end
```

Figure 1. Sample File.

to send the next command; if not it branches (or "skips") to a subroutine located in the central file DIRECT.PAD to disconnect from the remote system. Both sequences end with the "succeed" command. When "succeed" is reached during the CONNECT sequence, control is passed to the user, who can then operate the remote system directly. During the logout sequence, this command returns the user to the opening menu, from which he or she can choose another service or exit from the SALBIN system.

Constructing these files is quite simple, and with some practice, it is possible to produce them quickly. The most efficient way to proceed is to connect to each system manually using Kermit or a similar communications package. Following the instructions in Hytelnet, one can log on to the system and log the session to a disk file, using the communications software's logging facility. A sample .CNT file such as the one listed earlier can be copied to serve as a template. It is simple to cut and paste the necessary commands and system responses from the log file into this template. Then test the system out by running SALBIN, making any corrections necessary.

CONCLUSION

After completing the script files, the system is ready to offer readers a painless gateway to hundreds of remote data services. Putting the whole package together at the Bodleian took a number of months, although work on it was part-time only, so this time scale can be reduced. The rewards in terms of reader satisfaction and a raised information technology profile for the United Kingdom's oldest library have been considerable. SALBIN has proved a very satisfactory way of putting electronic information sources at the center stage of an institution more often associated with the historical than the technological. Other libraries may find it just as useful. ■■

Filtering Out Noise Lines from OPAC Downloads with *sed*

Brian Sealy

Internet access to other libraries' online catalogs has provided patrons and librarians alike with rich bibliographic data beyond the home institution's online public access catalog (OPAC). Users who know that a particular institution has strong collections in certain subject areas often can simply dial into that institution's OPAC and search through the holdings. Most communications software allow the user to download retrieved records to a file for adding to a bibliography or bibliographic database at a later time.

Of course, saving OPAC screens to a file will result in capturing many "noise lines," which form part of the source OPAC's user interface, in addition to the desired bibliographic information (as seen in the sample downloaded records shown in figure 1). One could import the file containing these and other records into a word processor and edit out the noise lines by hand or perhaps with the help of some macros. One could also write a program to strip out noise lines at the point of downloading, much as Hagee and Boewe have done for their institution's OPAC.[1]

Another solution would be to run the file of downloads through a text stream editor that can filter out noise lines. The stream editor *sed* performs this function well, and it is available to UNIX users on most workstations as well as to DOS users in the form of a public-domain program or as part of the UNIX utility package for DOS, called MKS Toolkit.[2,3]

THE BASICS

The way *sed* works is simple: It matches on a text pattern and then performs some action on the pattern or on the line that contains the pattern. Suppose that in a text file titled "libcomp" that has used the term "DOS," the user wants to change the term "DOS" to "MS-DOS." The *sed* editor can do this easily with the following simple script:

Brian Sealy is Assistant Librarian at the University of Michigan, Ann Arbor.

```
s/DOS/MS-DOS/
```

Example 1

The "s" at the beginning of the script stands for the "substitution" command. Patterns are always enclosed in "forward slashes" (/), with the pattern to be matched coming first and the pattern to substitute coming second. The syntax of this statement is then '*s/<pattern to be matched >/<pattern to be substituted>/*'. An excerpt from the text file libcomp reads

The Library uses PCs running DOS as terminals for its online catalog. It is important, then, for Library staff to have a working knowledge of DOS. . .

After going through the *sed* script the text reads

The Library uses PCs running MS-DOS as terminals for its online catalog. It is important, then, for Library staff to have a working knowledge of MS-DOS. . .

Note that pattern matching in *sed* is always case-sensitive. The command 's/dos/MS-DOS/' would not result in any matches or substitutions in the passage cited above.

Many word processors have an easy-to-use search-and-replace feature, but even at this elementary level *sed* offers the advantage over the word processor in that the *sed* script operates on and outputs ASCII text; there is no need to import the text into a word processor, make the change, and then possibly have to export the text back into ASCII format.

How does the user execute a command like the one in example 1? As an executable program, the tool *sed* is always invoked on the command line in the UNIX or DOS shell, but the command script to be executed may be either on the command line or in a file that the *sed* command calls. The examples in this article assume that scripts will be placed in files, since this is the most practical method for anything other than a very short script. Saving sequences of commands as a script in a file also allows the user to go back and easily edit the script if it does not give the desired result. To issue example 1 on the command line the user would enter

```
sed 's/DOS/MS-DOS/' libcomp
```

Example 2

The syntax of executing a *sed* command on the command line is then

sed '<command>'<filename>

The command must always be enclosed in single quotes so that sed recognizes it as a command string and not as a filename.

To execute a script by calling up a file, the user would save the script in a file called, say, "dosex" and then on the command line type

sed -f dosex libcomp

Example 3

The "-f " flag tells *sed* to look for the file "dosex" containing a script. Once the file dosex is found, then the script in the file is executed on the text file libcomp.

Sed applies each line of a *sed* script sequentially from the first line to the last to each line of the text file starting with the first line and moving to the last line and takes any appropriate action if there is a pattern match. In the case of example 1, *sed* alters a line of the text stream only if the term "DOS" appears in that line. The script terminates execution after the line in this script has been applied to the last line of text. It is important to keep in mind the order in which script lines operate on text file lines when writing longer scripts that may make multiple changes to a single line.

REGULAR EXPRESSIONS

Patterns in *sed* scripts are written as regular expressions. Example 1 illustrates a simple substitution of one string for another, but regular expressions allow an expansion of pattern matching with wild-card characters and line positions. Regular expressions achieve this increased flexibility in specifying patterns of text strings by assigning special meanings to certain characters. The special characters, or "metacharacters," are what give flexibility and power in formulating patterns. The following is a list of metacharacters used in regular expressions:

• A period (.) matches any one character, including a space, but not including "newline," i.e., the hidden character "\n" that ends each line of text.

• An asterisk (°) matches zero or more occurrences of the character immediately preceding. Using the character '☐' to denote a space, the regular expression '/☐☐°/' matches a pattern space that has at least one space (the first '☐'). The '☐°' indicates there may be additional spaces up to any number or there may be no spaces in addition to the first one.

• Brackets ([ ]) are used to denote one character out of a "class" of characters to be matched. A hyphen '-' can be used to denote certain standard ranges of characters as in the expression '/[a-z]/', which matches on any one lowercase letter or '/[A-Z]/' on any one uppercase letter or '/[0-9]/' on any one single digit. Likewise, brackets may enclose a quite heterogeneous class as in '/[c3%@]/'. This class matches on any one of the characters listed between the brackets. If a circumflex is used as the first character in the class, then the pattern matches on any character except those enclosed between brackets. So the expression '/[^wlis85@]/' matches on any one character except the seven listed after the circumflex between the brackets. Metacharacters appearing between brackets are interpreted by *sed* in their literal—not their special—meanings, except for '^', as noted above. '/[.°]/', for instance, matches on an occurrence of either a period or an asterisk.

• Used as the first character of a regular expression (except when enclosed in brackets, as noted above), the circumflex (^) has the meaning "beginning of line." The expression '/^747/' means "the string '747' appearing at the beginning of a line." Specifying line position with "^" (and the following metacharacter "$") allows more specificity in pattern definition.

• Used as the last character in a regular expression, the dollar sign ($) means "end of line;" so the expression '/747$/' translates as "the string '747' at the end of a line."

• When a backslash (\) precedes a metacharacter, then *sed* interprets the character literally. The expression '/\°\.\\/' will match on the character string "°.\".

DELETING LINES

To delete an entire line matching a specified regular expression the command format is

/<pattern>/d

The "d" tells *sed* to delete any line that matches the pattern in the '/*<pattern>*/'. To delete all lines with the pattern '/^volumes/', i.e., with the text string "volumes" at the beginning of a line, the *sed* script should contain the line

/^volumes/d

Example 4

To delete lines in a text file beginning with the word "volumes" or "Volumes," one could write two commands: */^volumes/d* and */^Volumes/d*. Regular expressions, however, make it possible to combine these two commands into one:

/^[Vv]olumes/d

Example 5

The above statement translates as "delete all lines that begin with 'volumes' or 'Volumes'." In other words, *sed* looks for either a 'V' or a 'v' to begin the specified string.

DELETING STRINGS

While "d" deletes an entire line matching the pattern in the specified regular expression, to delete a character or a character string in a line without deleting the entire line, it is necessary to use the substitution command ("s"). To delete the strings "volumes" or "Volumes" coming at the beginning of a line one would include in the *sed* script the command

s/^[Vv]olumes//

Example 6

What this script does is "substitute nothing (//) for the string 'volumes' or the string 'Volumes' coming at the beginning of a line." This command does not substitute spaces for the specified pattern; it simply takes out all characters, including spaces, which the pattern occupied, leaving nothing in its stead. The command in example 6 would change the line "Volumes 1–3, 5, 7–10, 12" to read "1–3, 5, 7–10, 12".

GLOBAL SUBSTITUTION

The "s" command operates only on the first occurrence of a pattern in a line, but it is also possible to delete more than one occurrence of a pattern in a line by appending 'g', the global specifier, to a substitution ("s") command, as in

s/DOS/MS-DOS/g

Example 7

This command will substitute "MS-DOS" for every occurrence of "DOS" in a single line, so if a text file contained the line "Staff should use DOS bootup diskettes to start up DOS machines without hard disks," example 7 would alter this line to read "Staff should use MS-DOS bootup diskettes to start up MS-DOS machines without hard disks." The earlier command "s/DOS/MS-DOS/" in example 1 would have performed the substitution only on the first occurrence of "DOS," so the altered line would have read "Staff should use MS-DOS bootup diskettes to start up DOS machines without hard disks."

WRITING A FILTER SCRIPT

The first step in developing a script to filter out noise lines from the file of OPAC screens shown in figure 1 is to identify those lines that do not contain bibliographic information and should be removed from the file. The most obvious candidates are

HU SHORT DISPLAY.....
FIND SU ...
———— ... HU HOLLIS...
.
.
.
———— ...
OPTIONS:
LOCATION ... PREVIOUS ...
HELP COMMANDS ...
COMMAND?

Furthermore, the scripter does not want local locations in the final list of citations, so these should be removed as well. The delete command ("d") will make short order of all these unwanted lines in a script that matches on the text string patterns and then removes the lines that contain them.

sed script hvd.sed

/^☐*HU☐SHORT/d
/^☐*FIND☐/d
/——/d
/^☐*LOCATION:/,/^☐*COMMAND?/d
Note: '☐' indicates a space in the script.

Example 8

The first line of the script is typical of the entire script. Translated into ordinary language, it gives the instruction "If the line

```
 HU SHORT DISPLAY  page 1 of 1    Item 7 of 87 retrieved by your search:
FIND SU MOTION PICTURES --INDIA
-------------------------------------------------------------HU HOLLIS# AHE9565 /bks
       TITLE: Cinema and cultural identity : reflections on films from Japan,
                India, and China / edited by Wimal Dissanayake.
    PUB. INFO: Lanham, MD : University Press of America, c1988.
  DESCRIPTION: vii, 214 p. ; 24 cm.

     SUBJECTS: *S1 Motion pictures--Asia.
               *S2 Motion pictures--Japan.
               *S3 Motion pictures--India.
               *S4 Motion pictures--China.
               *S5 Japan--Civilization--20th century.

     LOCATION: Harvard-Yenching: W 6838.4 173
               Widener: WID-LC PN1993.5.A75 C56 1988
                 C1 - Enter DISPLAY C1 for circulation information

-------------------------------------------------------------------------------------
OPTIONS: DISPLAY LONG                              NEXT - next item      HELP
         LOCATION                                  PREVIOUS - prev item  INDEX

OPTIONS: DISPLAY LONG                              NEXT - next item      HELP
         LOCATION                                  PREVIOUS - prev item  INDEX
         HELP COMMANDS       TRACE *S1 (etc)       QUIT - exit database  REDO
COMMAND?

 HU SHORT DISPLAY  page 1 of 1    Item 11 of 87 retrieved by your search:
FIND SU MOTION PICTURES --INDIA
-------------------------------------------------------------HU HOLLIS# ACB4435 /bks
       TITLE: Fifty years of Indian talkies, 1931-1981 : a commemorative
                volume.
    PUB. INFO: Bombay : Indian Academy of Motion Picture Arts & Sciences, 1981.
  DESCRIPTION: [166] p. : ill. (some col.) ; 29 cm.

     SUBJECTS: *S1 Motion pictures--India.
               *S2 Motion picture actors and actresses--India.

     LOCATION: Fine Arts: FA11446.3
  NON-HARVARD
     LOCATION: CRL (Chicago):

-------------------------------------------------------------------------------------
OPTIONS: DISPLAY LONG                              NEXT - next item      HELP
         LOCATION                                  PREVIOUS - prev item  INDEX
         HELP COMMANDS       TRACE *S1 (etc)       QUIT - exit database  REDO
COMMAND?
```

Figure 1. File of OPAC Screens.

matches the pattern 'zero or more spaces at the beginning of the line followed by the string 'HU', followed by a space and then the string 'SHORT', then delete the entire line." Remembering how *sed* acts on a text file (see figure 2), the script considers every line in the download file and applies each line of the script to the line. If the script were at the "HU SHORT DISPLAY ..." line in the download file, the first line of the script would apply, so it would perform the indicated action, i.e., delete the entire line.

Why are the "beginning of line" (^) and "zero or more spaces" (□*) necessary? Why not just delete any line with the string "HU SHORT"? There is the remote possibility that the string "HU SHORT" could appear in the text of a bibliographic citation, but the chance that it will be the first text string on a line in a citation is extremely remote, so by specifying "beginning of line with zero or more preceding spaces" the script still specifically targets the noise line for deletion while reducing the possibility of removing any lines of a citation to almost nil. A regular expression in a *sed* script should achieve a balance between being general enough to match the target lines, with any small variations they may contain, while not being so general that desired text lines are altered inadvertently.

```
      TITLE: Cinema and cultural identity : reflections on films from Japan,
               India, and China / edited by Wimal Dissanayake.
  PUB. INFO: Lanham, MD : University Press of America, c1988.
DESCRIPTION: vii, 214 p. ; 24 cm.

   SUBJECTS: *S1 Motion pictures--Asia.
             *S2 Motion pictures--Japan.
             *S3 Motion pictures--India.
             *S4 Motion pictures--China.
             *S5 Japan--Civilization--20th century.


      TITLE: Fifty years of Indian talkies, 1931-1981 : a commemorative
               volume.
  PUB. INFO: Bombay : Indian Academy of Motion Picture Arts & Sciences, 1981.
DESCRIPTION: [166] p. : ill. (some col.) ; 29 cm.

   SUBJECTS: *S1 Motion pictures--India.
             *S2 Motion picture actors and actresses--India.


      TITLE: Indian cinema superbazaar / edited by Aruna Vasudev, Philippe
               Lenglet.
  PUB. INFO: New Delhi : Vikas, c1983.
DESCRIPTION: xvii, 384 p., [21] p. of plates : ill. ; 25 cm.

   SUBJECTS: *S1 Motion pictures--India.
             *S2 Motion picture industry--India.
             *S3 Motion picture producers and directors--India--Interviews.
             *S4 Motion picture actors and actresses--India--Interviews.


      TITLE: Cinema in India.
  PUB. INFO: [Vol. 1, no. 1] (Jan. 1987)-
             [Bombay : National Film Development Corp., 1987-
DESCRIPTION: v. : ill. ; 28 cm.

   SUBJECTS: *S1 Motion pictures--India--Periodicals.


      TITLE: Deep focus.
  PUB. INFO: Vol. 1, no. 1 (Dec. 1987)-
             Bangalore : Deep Focus, [1987-
DESCRIPTION: v. : ill. ; 24 cm.

   SUBJECTS: *S1 Motion pictures--India--Periodicals.
             *S2 Motion pictures--Periodicals.
```

Figure 2. Text File.

DELETING RANGES OF LINES

The last line of the script deserves special mention because it deletes a range of lines rather than a single line. The command matches on a line with the first of the two patterns and then marks that line. As the *sed* script processes succeeding lines it looks for a match on the second pattern. If the script finds a match on the second pattern in a succeeding line, then it deletes all the lines from the line with the first pattern to the line with the second pattern. The syntax of the command is as follows:

/<pattern 1>/, /<pattern 2>/d

In fact, looking again at example 8, the first three lines of the script could all be combined into one command to delete a range, since the patterns in these commands always appear in successive lines in figure 1. Combining the first three commands into one results in a two-line script:

revised sed script hvd.sed

/ ^ □ *HU □ SHORT/,/——/d

/^ □ *LOCATION:/,/^ □ *COMMAND?/d

Example 9

ADDING REFINEMENTS TO THE SCRIPT

When *sed* is applied to the file of screendumps called "dwnld.tmp" with the command "*sed* -f hvd.sed dwnld.tmp", the citations in figure 1 appear as in figure 2. Reviewing the output

```
      TITLE: Cinema and cultural identity : reflections on films from Japan,
               India, and China / edited by Wimal Dissanayake.
  PUB. INFO: Lanham, MD : University Press of America, c1988.
DESCRIPTION: vii, 214 p. ; 24 cm.
   SUBJECTS:  Motion pictures--Asia.
              Motion pictures--Japan.
              Motion pictures--India.
              Motion pictures--China.
              Japan--Civilization--20th century.

      TITLE: Fifty years of Indian talkies, 1931-1981 : a commemorative
               volume.
  PUB. INFO: Bombay : Indian Academy of Motion Picture Arts & Sciences, 1981.
DESCRIPTION: [166] p. : ill. (some col.) ; 29 cm.
   SUBJECTS:  Motion pictures--India.
              Motion picture actors and actresses--India.

      TITLE: Indian cinema superbazaar / edited by Aruna Vasudev, Philippe
               Lenglet.
  PUB. INFO: New Delhi : Vikas, c1983.
DESCRIPTION: xvii, 384 p., [21] p. of plates : ill. ; 25 cm.
   SUBJECTS:  Motion pictures--India.
              Motion picture industry--India.

   SUBJECTS:  Motion pictures--India.
              Motion picture industry--India.
              Motion picture producers and directors--India--Interviews.
              Motion picture actors and actresses--India--Interviews.

      TITLE: Cinema in India.
  PUB. INFO: [Vol. 1, no. 1] (Jan. 1987)-
             [Bombay : National Film Development Corp., 1987-
DESCRIPTION: v. : ill. ; 28 cm.
   SUBJECTS:  Motion pictures--India--Periodicals.

      TITLE: Deep focus.
  PUB. INFO: Vol. 1, no. 1 (Dec. 1987)-
             Bangalore : Deep Focus, [1987-
DESCRIPTION: v. : ill. ; 24 cm.
   SUBJECTS:  Motion pictures--India--Periodicals.
              Motion pictures--Periodicals.
```

Figure 3. Sample of Output.

shows that "hvd.sed" has cleaned up the download file a great deal, but it could be still tighter and cleaner. There are still blank lines within the citation, while it would be better to have each citation display as a continuous block of text with a blank line between citations. Removing blank lines is an easy task with the command "/^$/d", which translates as "delete all lines that have nothing between the beginning of the line and the end of the line." This will be the first addition to a revised script.

Since subject heading numbers such as "*S1" clutter the display, a line should be added to the script to delete just these numbers (but not the subject headings). The subject heading numbers contain an asterisk (*), which is a special metacharacter, so for the script to interpret the asterisk as a literal character in a pattern, the asterisk needs to be escaped with a backslash (\). Since there could be more than nine subject headings, the command to delete the subject numbers should allow for the possibility of more than one digit following the "S". The resulting command would be

s/ *S[0-9][0-9]* □ //

The first asterisk is treated as a literal character because it is escaped by the metacharacter "\", while *sed* reads the second instance of an asterisk as a metacharacter referring to the preceding class of characters.

For readability it is desirable to leave one blank line between each citation. The command "/^$/d", as discussed above, would remove *all* blank lines, and it would leave no lines separating the citations. Examining the input in figure 1 shows that each screendump ends with the line "COMMAND?" Instead of deleting this line, as was done with the first version of the script in example 8, it is possible to replace this line with a blank line using the command "s/^COMMAND?//", which simply removes the text from this line and leaves it blank. The command "/^$/d" to delete blank lines should, then, come before "s/^COMMAND?//" in the script; otherwise the blank line deleter will also delete the blank line the script just created to separate citations.

The revised script, taking into account all the changes discussed, appears as follows:

```
revised sed script hvd.sed

/^□*HU.□SHORT/, /——/d
s/\*S[0-9][0-9]* □ //
/^□*LOCATION:/,/^□*HELP□COM-
   MANDS/d
/^$/d
s/^□*COMMAND?//
```

Example 10

After running the file "dwnld.tmp" through this *sed* script, a sample of the output looks like that shown in figure 3.

PRACTICAL IMPLEMENTATION

Needless to say, no single universal script will work with any and all OPAC screens. One should also keep in mind that institutions change their OPAC screens, so periodic revisions to the filter scripts may be necessary. It is hoped that the preceding discussion has shown that once the basics of *sed* are mastered, writing a filter script is no more difficult than writing a macro for an applications program. It is, then, quite possible to develop several *sed* scripts, one for each remote OPAC most often consulted at, say, the reference desk. Distribution disks with these scripts as well as the public domain version of *sed* for DOS might also be made available to patrons.

Once screens are downloaded to a file from the remote OPAC, a batch file, in the case of DOS, or a shell script or command alias, in the case of UNIX, could make running the downloads through a *sed* script a simple matter of issuing one command. Following is one implementation for DOS:

1. Place the filter scripts in a directory called "filters."

2. Always save the downloaded OPAC screens to a temporary file with a standard name such as "dwnld.tmp."

3. Develop batch files that call a *sed* script and run the dwnld.tmp file through it and then write the filtered output to a file. For example, the batch file "hvd.bat" runs dwnld.tmp through the script hvd.*sed* in example 10 and writes the filtered ouput to a file called "hvd.fil":

```
batch file hvd.bat

sed -f hvd.sed dwnld.tmp  hvd.fil
```

Example 11

4. Once records are downloaded, the user would simply leave the communications software and issue the command "hvd" on the DOS command line. The resulting file hvd.fil should then be saved to another file so that it is not overwritten the next time hvd.bat is run.[4]

REFERENCES AND NOTES

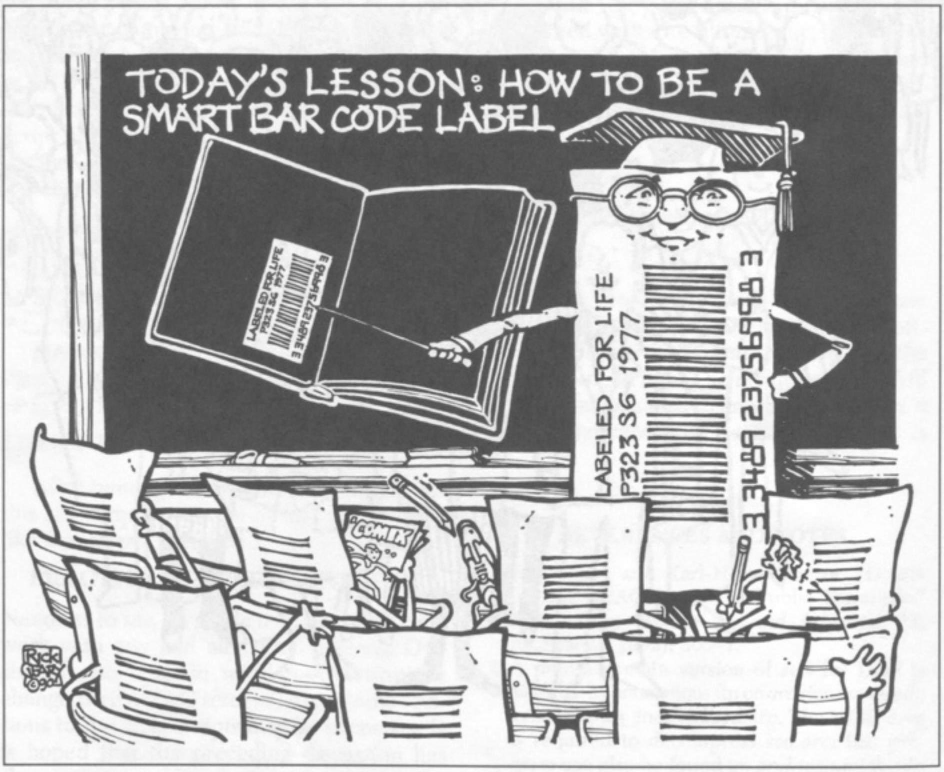
1. Jon Hagee and Karl-Heinz Boewe, "Downloading OPAC Screens at Public Terminals," *Information Technology and Libraries* 11, no.3 (Sept. 1992): 305–7.
2. A public domain version of *sed* for DOS is available by anonymous ftp on *archive.umich.edu* as */msdos/un_indexed/sed.arc*. The utility *arce* is required to decompress *sed.arc*; this program can also be found on *archive.umich.edu* as *msdos/compression/arc/ arce.com*. Searching an archive for *sed.arc* and *arce.com* will retrieve a listing of other anonymous ftp sites for these programs.
3. MKS Toolkit, a package of 160 UNIX utilities as well as a UNIX shell for DOS machines, is made by Mortice Kern Systems of Waterloo, Ontario. For the author's review of this software package, see *Information Technology and Libraries* 11, no.4 (Dec. 1992): 448.
4. Questions about *sed* or this article may be directed to the author on the Internet: BRIAN.SEALY@UM.CC.UMICH.EDU.

BIBLIOGRAPHY

Abrahams, Paul W., and Bruce R. Larson. *UNIX for the Impatient*. Reading, Mass.: Addison-Wesley, 1992.

Dougherty, Dale. *sed & awk*. Sebastopol, Calif.: O'Reilly & Associates, 1991.

Dougherty, Dale, and Tim O'Reilly. *UNIX Text Processing*. Carmel, Ind.: Sams, 1989. ■■


TODAY'S LESSON: HOW TO BE A SMART BAR CODE LABEL
LABELED FOR LIFE
P323 S6 1977
3 3489 23756998 3
COMIX
Rick Geary ©90

Recent Publications

Book Reviews

Advances in Library Resource Sharing. Vol. 3. Ed. by Jennifer Cargill and Diane J. Graves. Westport, Conn.: Meckler. 1992. 153p. $55 (ISBN 0-88763-737-9).

This annual compendium of papers addressing resource sharing issues covers a wide variety of topics grouped into five sections: perspectives on resource sharing, collection development models, technology and resource sharing, copyright, and other areas. Within these sections is a mix of information, both theoretical and applied, intended to address contemporary resource sharing issues for all kinds of libraries.

On the applied side, seven case studies are reported, including the effect of technology on resource sharing in a school media center, document delivery via courier in Oregon, and automated collection analysis in Boston. Of particular note is the article concerning the Illinois Library Computer Systems Organization, which outlines issues and challenges in the operation and governance of ILLINET Online.

On the theoretical side, there are articles concerning resource dependence theory and its applications to interlibrary lending patterns and decisions, and the history and value of resource sharing. The latter, offered by Robert Miller, Director of Libraries at Notre Dame, focuses on the increasing importance of resource sharing for effective delivery of information and the need for consideration of public financial support for resource providers. In addition, there are technical updates addressing general copyright trends, the place of electronic mail in interlibrary loan, and library fax server technology as well as survey results concerning patterns of usage among library fee-based information services. This survey is notable in its findings that library fee-based services need not be based in major research institutions in order to be successful.

This collection contains information of value to both interlibrary loan librarians and library administrators. The articles are timely and accessible, and they reflect the impact of both enhanced technology and decreased funding for comprehensive collection building. As increased end user access through the Internet and NREN evolves, even more rapid changes in resource sharing will occur; the library community is monitoring the effects of these changes and will, it is hoped, report on them in future editions of this publication.—*John Hammond, North Country 3R's Council, Canton, New York.* ■■

Beiser, Karl. *DOS 5.0 for Libraries.* Westport, Conn.: Meckler. 1992. 185p. $39.95 (ISBN 0-88736-835-2).

In existence since 1981, DOS remains the world's most commonly used microcomputer operating system. There is every indication that it will continue to hold the lead for some time. Therefore, it behooves all but the most casual user of personal computers to have at least a passing acquaintance with DOS.

Karl Beiser's new book serves two purposes: to introduce beginners to the mysteries of DOS and to provide experienced users with a moderately detailed description of features that entered DOS with version 5.0, which appeared in 1991. In order to facilitate the first goal, the author has placed discussions of the most important commands (such as FORMAT, COPY, and MKDIR), as well as subdirectories and the DOS shell, at the front of the book. People who have used DOS for a while, however, will be more interested in the book's last half, where they will find information on the new DOS editor, memory management, batch files, advanced commands, and DOS supplements.

Since the arrangement of DOS commands in this book is not alphabetical, the index is critical to the volume's success as a reference tool. The index passes muster, but unfortunately, it rarely goes beyond DOS commands; most other information in the book is simply left unindexed. Also very important for the book is the glossary, which proves to be clear and straightforward.

Although the title implies that the book's intended audience is librarians, the text makes few references to them or, indeed, to libraries. Only where it treats the special DOS needs of machines that operate CD-ROM drives does the book address an issue of peculiar relevance to the library profession.

Beiser's information is accurate, but his writing is sometimes awkward. In general, though, his style is better than that found in most standard DOS manuals. Better proofreading could have prevented some annoying typographical mistakes, and Beiser would have been better served with different typefaces and page layouts, which could have made this a more appealing and useful resource. A helpful feature of the book is Beiser's occasional evaluation of third-party software that improves on or supplements aspects of DOS.

Beiser has written several other books on computer software, including an earlier volume on DOS. These books have found their way into numerous libraries. Many medium-sized to large libraries, as well as all library schools, would do well to purchase this book, too, for their collections.—*Douglas Koschik, Birmingham, Michigan.* ■■

Buckland, Michael. *Redesigning Library Services: A Manifesto*. Chicago: American Library Assn., 1992. 83p. paper, $18 (ALA order code 0590-0-0010) (ISBN 0-8389-0590-0).

With its provocative title and enticing foreword (by Michael Gorman), this manifesto, which promises to be both "visionary and practical" (p. v), aroused my strong interest. Indeed, ALA's publicity release about this book noted that "no one interested in the future of library service should miss this uncommon, thoughtful analysis of where we've been, where we are, and where we may be going." Unfortunately, *Redesigning Library Services* does not live up to these high expectations. It is a slender volume (just over eighty pages in length) and thin in content, presenting little new about the changes libraries are undergoing as a result of the introduction of electronic information technology.

It is not that Michael Buckland is wrong in his description of the present or in his predictions about the near future of library services. His strength is his simplicity in sticking to basic principles and broad concepts: the underlying role and mission of library service, the important distinction between technological means and service ends, and the general characteristics of the information revolution in libraries. His weakness, however, is his failure in this book to move beyond an outline of abstract observations into the details and complications of the change facing libraries. Maybe it is unfair to expect more from a "manifesto," but I was disappointed in this work as a scholarly monograph.

There are serious flaws of omission and underdevelopment of critical issues. For example, chapter 8, titled "Organization and Implementation," begins with the understandable point that political and economic issues related to governance and organizational structure are critical to a new scheme for library service. But then Buckland claims, "Nevertheless, for three reasons, this area will not be examined in any detail beyond the noting of a few general guidelines" (p. 69). For reasons Buckland describes as situational and strategic, important organizational and governance issues are dismissed in a three-page chapter. The appendix of sources that ends the book too is skimpy and superficial. Buckland lists a few source books and journal titles, but says "a systematic approach to published literature is not very practical because little of what is published addresses in any direct way the sorts of strategic issues examined in this book and because what is published is not concentrated in any one place" (p. 77). Buckland is simply wrong here. It may not be concentrated in one place, but there is a wealth of good, and I might add better, literature on the short-term and long-term future of libraries.

Buckland's general observations about the basic changes in library services as the technology moves through phases he calls "print," "automation," and "electronic" hold some useful insight for bibliographic control and collection management planning. He sees in the near future an integrated information system in which bibliography, cataloging, inventory control, and full text and full images are available over an electronic network. His clarity and logic, unencumbered with messy details, does have an effective "manifesto" quality at times in the book. But this vision is not new, and he tells us very little in practical terms about how to get there.—*Joseph J. Branin, University of Minnesota Libraries.* ■■

CD-ROM Market Place 1992: An International Guide to the CD-ROM, CD-I, CDTV & Electronic Book Industry. Westport, Conn.: Meckler, 1992. 273p. paper $30 (ISBN 0-88736-851-4).

While the road to a comprehensive source of corporate information on the CD-ROM industry has been a long one, the journey is not yet ended. Back in 1987, Learned Information's *Optical Publishing Directory* listed companies under three categories. The current edition of this directory, now renamed *CD-ROM Finder*, simply provides a company index, where about 600 firms are listed. Some 1,700 companies and organizations are included in Gale's *Directory of Databases Vol. 2: CD-ROM, Diskette, Magnetic Tape, Handheld, and Batch Access Database Products* (formerly the *Directory of Portable Databases*), while TFPL's *CD-ROM Directory 1993* has a company information section covering almost 2,800 firms in the industry. In all three cases, the large majority of companies described are database distributors, information providers, publishers, and software developers.

CD-ROM Market Place 1991: An International Guide was the first attempt at a separate corporate directory for the industry. The entries for that edition were based on the company listings in Meckler's *CD-ROMs in Print*. This new edition carries the expanded title of *CD-ROM Market Place 1992: An International Guide to the CD-ROM, CD-I, CDTV & Electronic Book Industry* but fails to live up to its billing. There is no way to identify the CD-I, CDTV, or electronic book publishers.

The bulk of the book is composed of the company directory, an annotated list of companies and institutions involved in every stage of the business of CD-ROM. Along with the four categories referred to above, this directory lists conference organizers, consulting firms, data preparation/premastering facilities, hardware producers, mastering/replication plants, and publishers of information about CD-ROMs.

The coverage has grown from about 1,000 organizations in 1991 to nearly 3,000 in the 1992 edition. However, many entries include only partial information, a fact acknowledged in the book's introduction. While the inclusion of partial entries adds to the directory's comprehensiveness, surprisingly, some significant firms and organizations are not represented. These include Micro Design International, developers of SCSI Express software; Pemberton Press, publisher of *CD-ROM Professional*; and SIGCAT, the world's largest CD-ROM user's group and publisher of the SIGCAT Software Showcase CD-ROM.

A typical entry includes the following fields: name, address, telephone and fax numbers, key personnel, area of CD-ROM publishing activity, CD-ROM program, databases published, and year of first CD-ROM title. The CD-ROM program field is used as a note field for company and product descriptions and sometimes for small ads.

There are four indexes: subject, geographic, activity, and name. The subject index covers topics of CD-ROM titles published by the companies; the geographic index is by state and city in the United States, and by country and city otherwise. The activity index relies on information in the "area of CD-ROM publishing activity" field in the directory portion of the book. The names and numbers index attempts to be the "white pages" for the industry and lists the key contacts from the directory entries with their organization name and telephone and fax numbers.

Errors are probably inevitable in compiling a directory like this: the entry for CAB International, based in the United Kingdom, notes the company merely as a data provider in the activity field even though its publishing activities are described in the program field. Instead, a branch office in Malaysia is acknowledged as the publisher of CAB Abstracts.

More serious is the misuse of the CD-ROM program field (rather than the activity field) in profiling a company, resulting in the company's being omitted from the appropriate index. An example is Merisel, which is noted as a distributor in the program field but is not in the activity index under "distributors."

One useful feature for future editions would be a way to identify the firms involved with CD-I, CDTV, and other CD-ROM-based formats. Another would be some indication of the CD-ROM titles to which a given information provider contributed. It would also be handy to have financial information, at least for the industry if not for individual companies. Granted, the publisher may have a hard time keeping the price as low as it is now. There are more steps to take on the road to an ideal source of information on the organizations involved in the CD-ROM industry.—*Ka-Neng Au, Rutgers University Libraries.* ■■

Classification Research for Knowledge Representation and Organization: Proceedings of the 5th International Study Conference on Classification Research, Toronto, Canada, June 24–28, 1991. Ed. by Nancy J. Williamson and Michèle Hudon. New York: Elsevier, 1992. 427p. $128.50 (ISBN 0-444-89343-1).

The function of classification is "to order documents . . . in a sequence which can be understood by users and which actively helps to promote their thinking about their subjects" (Foskett, quoted by Wilson, p. 389). Yet the contents of this book about classification are not classified. The papers in the three sections—keynote address (Elaine Svenonius), plenary sessions (nine papers arranged by author), and contributed papers (thirty papers by author)—lend themselves to cross-classification. In sections 2 and 3 there are philosophical papers (Francis Miksa) and general overviews (Eric de Grolier); many of the contributed papers are far more substantive than the plenary report on a planned thesaurus for records management. There are fortuitous juxtapositions of related papers (e.g., Rebecca Green's and Susanne M. Humphrey's on frames, and William Stiles' and Elaine Svenonius' on chain indexing), but there are mainly jarring transitions from papers on traditional classification schemes to thesauri and expert systems.

Some of the papers have promising beginnings but disappointing endings—e.g., Pauline A. Cochrane's "Indexing and Searching Thesauri" poses interesting questions but concludes with musings about the need for further research. A paper that delivers more than the title promises is Linda Smith's "UNISIST Revisited"—a well-documented overview of the compatibility of indexing languages.

The editors contributed an introduction and a summary, each one page long. The introduction states that camera-ready copy was used in order not to delay the availability of the proceedings. The book appeared almost a year after the conference was held, however, and much could have been done to improve the quality of the papers. In many of the papers in the volume, the type is very small. Timothy C. Craven's "Concept Relation Structures and Their Graphic Display" lacks figures; the figures in Susanne M. Humphrey's paper follow her references, rather than being juxtaposed with the text. Raya Fidel's "Thesaurus Requirements for an Intermediary Expert System" lacks references. The editors did not proofread their introduction, and many of the papers are marred by gross misspellings. Fugmann should have asked a native English speaker to extricate his German syntax and to correct his spelling.

Elsevier spokespersons tout the "value-added services" of publishers as justification for their high prices. The publisher provided bibliographic data at the head of each paper, but it lacks main entry—the name of the conference—as indicated by the CIP. Running heads would have facilitated fan-searching or browsing through the volume, but there are none.

The two-page subject index contributed by the coeditor is structurally sound, well formatted, and somewhat analytical. Some of the see references could have been converted to double entries. There are no see also references; these would have created a hidden classification that could have compensated to some extent for the lack of logical arrangement in the text.

Many experts in classification theory participated in the conference; most have published similar papers elsewhere. There are some golden nuggets in this book, but they require a better setting to shine through.—*Bella Hass Weinberg, St. John's University.* ■■

Clayton, Marlene, with Chris Batt. *Managing Library Automation*. 2d ed. Brookfield, Vt.: Ashgate, 1992. 218p. $59.95 (ISBN 1-85742-003-9).

This work, an update of a book published in 1987, is intended to provide a guide to the most important issues in library automation for librarians and information workers. Chapter titles are "Planning for Automation"; "Hardware and the Library Manager"; "Software"; "Technical Support Functions"; "Established Standards, Techniques, and Compatibility Issues"; "Communications and Networking"; "Choosing a Library Management System"; "Implementing the System";

"File Creation"; and "Managing Working Systems." There is a well-rounded bibliography and an index.

United States readers should be aware that this book was written and published in the United Kingdom and that the examples and terminology are very specific to that country. This does not invalidate any of the concepts discussed in the book, but it may hinder U.S. readers who are new to library automation and need a familiar frame of reference.

I found this book to be frustrating for two reasons. First, far too much space is devoted to the description and discussion of the technical aspects of information technology and too little space is given over to management issues. Second, Clayton's impenetrable writing style renders inaccessible much of the useful information that does exist in the book. However, the book does contain helpful overviews of the management issues associated with the steps in automating library processes.

In the introduction, Clayton promises that "no attempt has been made to describe or explain computing techniques except where an understanding of these assists the effective use of systems by librarians" (p. vii). It turns out that about 40 percent of the book is devoted to "computing techniques," and a good deal of this information is not relevant to automation in libraries. Its purpose in this book, therefore, is unclear. Worse, very little of this information, relevant or not, is actually useful to the library manager attempting to organize an automation project. It is insufficiently clear and detailed for the novice and is redundant and irrelevant to the knowledgeable reader.

By contrast, the chapters on planning for, choosing, implementing, and managing an automated system and on file creation are relevant and helpful—up to a point. In particular I appreciated Clayton's emphasis on the persons affected by the introduction of automation into an organization and on human factors issues. Her discussion of publicity, education, and training is detailed and knowledgeable. Unfortunately, when discussing other important management issues, Clayton fails to deal with suggested techniques in sufficient depth, being satisfied with identifying the ideas as important.

The updating performed by Chris Batt on the 1987 edition is, in general, well integrated and useful. Given the high price of this book and the large number of books and articles on this subject more suitable to U.S. readers with respect to terminology and examples, I do not recommend its purchase.—*George Rickerson, University of Missouri System.* ■ ■

Crawford, Walt. *The Online Catalog Book: Essays and Examples*. Boston: G.K. Hall, 1992. 560p. $50 (ISBN 0-8161-1996-1); paper, $35 (ISBN 0-8161-1995-3).

Crawford's third book about online catalogs is an excellent resource for anyone involved in the design, implementation, or selection of an OPAC. Part 1 contains Crawford's own essays about issues in catalog design. Part 2 consists of contributions from users or vendors of thirty-two different online catalogs. Actual screen displays, captured from each system using screen capture software, are accompanied by commentary about the user interface and the system's search and display capabilities.

The design essays are informal, based mostly on Crawford's experience and opinions rather than on scholarly research. They are clearly written, well organized, and contain sprinklings of humor, such as the use of the "mutter test" as an indicator of system coherence—"If I can get all the way through a half-hour searching session without muttering, the system passes" (p. 63). Crawford goes beyond the current state of the art of OPACs to discuss emerging trends such as library-defined catalogs, graphical user interfaces, distributed systems, the Z39.50 protocol (used for searching outside systems using the local system's user interface), and integrated systems achieved by seamlessly linking together components from different vendors.

Some of Crawford's views are controversial, even "heretical," as he puts it. While agreeing that libraries should offer access beyond their own collections, he also sees the ongoing need for physical as well as "virtual" collections of material. He says that catalogs should meet the immediate needs of "quick-check searchers," who are the vast majority of catalog users, instead of forcing them into more comprehensive, research-oriented searching. He argues strongly for including all textual fields in the MARC record in a full bibliographic display, including added entries and notes which are often excluded from patron displays.

Crawford warns against using the included OPAC examples as a tool for making purchase decisions. Not every system on the market is represented, the information may already be out of date due to ongoing system development, and software modules other than the OPAC are not discussed. Despite all the caveats, the range of systems and types of libraries represented is impressive, and the examples at least provide a good starting point for system selection. The systems covered include not only the larger, better-known systems but also PC- and Macintosh-based systems (such as the Assistant, Catalog Plus, and MacLAP), CD-ROM OPACs (e.g., IMPACT, MARCIVE), and special library systems (e.g., BiblioTech, TECHLIBplus). Several locally developed systems, such as those at Dartmouth, Rensselaer Polytechnic, and Purdue, are also included. In several cases, the contributors are vendor representatives rather than users.

The screen displays make up the bulk of each contribution, accompanied by brief explanatory comments. Guided by a set of questions from Crawford, each contributor steps the reader through a variety of searches, different types of bibliographic displays, help screens, and responses to user errors. Specific searches suggested by Crawford are used if they apply to the particular database. Seeing all of these systems in such close proximity, the reader is struck more by their similarities than by their differences. Most are menu-driven, with a few offering a command mode in addition. A few offer command interfaces only. Most ask the user to choose a type of search, such as author, title, subject, call number, or other access point; a few offer global or cross-index searching as one of the choices. Keyword and Boolean searching are supported to varying degrees. Almost all the systems offer brief, full, and MARC bibliographic displays.

The screen displays are clear and legible, although generally not as easy to read as they would be at a terminal. Most show highlighting as they would on the actual system. In some chapters the displays inexplicably vary between small, three-to-a-page reproductions and larger, easier-to-read, two-to-a-page screens. In each chapter, the complete commentary appears first, followed by all of the screen displays, forcing the reader to do a lot of page flipping. In a few chapters the displays are out of synch with the numbers used to reference them in the text. Integrating the screens into the text would have made the book easier to read, although production would have been much more difficult and costly.

The entire book would have benefitted from closer integration of its two parts. The essays in Part 1 would have been more effective had they incorporated some of the screens as examples or at least referenced them more specifically. Nonetheless, the book is a valuable reference for anyone involved with OPACs.

This book does not supersede Crawford's two earlier works on the subject of online catalogs. *Bibliographic Displays in the Online Catalog* (1986) focuses more narrowly on displays of bibliographic records and does not discuss the other issues. *Patron Access: Issues for Online Catalogs* (1987) covers some of the same issues, some in even more detail, but obviously is not as up to date and does not include the screen examples. For those who want even more screen examples, Crawford offers *The Catalog Collection*, to be published annually by LITA. The collection includes all of the contributed screens, not just those selected to appear in the book, and promises to be updated annually to reflect new and enhanced catalogs.—*Fae K. Hamilton, Information Technology Consulting, Carlisle, Massachusetts.* ■■

Ensor, Pat, and Hardin, Steve. *CD-ROM Periodical Index: A Guide to Abstracted, Indexed, and Fulltext Periodicals on CD-ROM*. Westport, Conn.: Meckler, 1992. 420p. $65 (ISBN 0-88736-803-4).

Ensor and Hardin set out to list "all the periodicals indexed, abstracted, or carried fulltext by certain research-oriented CD-ROMS" (p.vii). More than 30,000 periodicals are listed from seventy-nine major CD-ROM databases. This book expands and supplements Pat Ensor's 1991 *CD-ROM Research Collections*, which provides descriptions and evaluations of 114 CD-ROM products but does not list the individual periodical titles indexed or abstracted on them.

The first section of *CD-ROM Periodical Index* is an alphabetic list of the CD-ROM databases, with each entry listing the title of the database, publisher, price, years covered, type of coverage (abstracting, indexing, fulltext), search software and capabilities, and the titles of periodicals covered in the database. The second section of the book, the Periodical Title Index, is an alphabetic listing of the periodicals covered and the CD-ROM databases in which they are indexed. Whether the journal is available fulltext is noted here as well. The Publisher Index, last in the volume, is a list of company names, addresses, telephone numbers, and the page numbers on which products appear in the volume.

This reasonably priced index would be a handy addition to any reference collection. It would be useful in maximizing use of CD-ROM databases present in the library's collection, in making referrals, or in the decision-making process for CD-ROM or serials acquisition.

In future editions, a one- or two-sentence database description following the CD-ROM title would be welcome in the Periodical Listings by CD-ROM Database section. Consistency within the periodical lists of each CD-ROM database would also be desirable. In about one-eighth of the CD-ROM periodicals listed there are notations following periodical titles as to years of coverage. For example, most titles in *INSPEC ONDISC* appear with a year notation following: "Applied Spectroscopy '68–, Applied Spectroscopy Reviews '71–," etc. Confusion arises when one compares these notations with the descriptive material at the beginning of the entry for *INSPEC ONDISC*. One field is titled Years Covered and lists 1989 to present. There is no explanation of this discrepancy. For other titles, a note explains the difference, as in *General Periodical Index-Public:* "Current database coverage begins in 1988 unless otherwise indicated." The majority of the CD-ROM periodicals lists have no such indication of years covered following the periodical title.

Another area that appears to need stylistic editing is the Periodical Title Index, in which some titles appear fully capitalized while others do not. Because no explanation for these differences is given, and no particular function can be discerned, one suspects that this is due to the work being a collaborative effort.

CD-ROM Periodical Index is a valuable reference work, since nothing else available covers periodicals indexed by CD-ROM databases in this depth. The book holds further promise as well, since the authors, in the preface, reveal plans for future editions that will include more CD-ROMs.—*Julie Pinnell, Johnson County Community College, Overland Park, Kansas.* ■ ■

ERIC Identifier Authority List (IAL) 1992. Ed. by Carolyn R. Weller and James E. Houston. Phoenix, Ariz.: Oryx, 1992. 497p. $55 (ISBN 0-89774-738-0).

The ERIC Identifier Authority List (IAL) 1992 continues the *IAL* first published in 1980 as an effort to add depth and expansion to ERIC searching and indexing. The 1980 *IAL* contained 25,000 identifiers, while the 1992 *IAL* contains more than 44,000. The volume is published as a companion or supplement to the *Thesaurus of ERIC Descriptors*, 12th ed. (1990).

It is necessary to compare ERIC descriptors with ERIC identifiers in order to provide the basis for use of the *IAL*. ERIC descriptors are controlled vocabulary terms and are immediately displayed—for example, when one performs an "Easy Menu" search in an ERIC CD-ROM database. Identifiers are more specific than descriptors. Often the name of a specific entity—e.g., person, place, organization, test, computer program, legislation—identifiers may be transitory and may come and go from versions of the *IAL*. Identifiers also may be candidates for descriptor status. Terms are first introduced as identifiers and exist as such until their scope, definition, and long-term use are determined. Format standards, such as geographic locations, organization names, and ethnic groups, are imposed on identifiers to eliminate variant forms of the same concept.

The 1992 *IAL* presents two displays. The "Alphabetical Display" lists all identifiers, plus a category number and postings count for both *Current Index to Journals in Education* (*CIJE*) and *Resource in Education* (*RIE*) (formerly *Research in Education*, an ERIC monthly publication). This list also includes cross-references to approved identifiers. The second part of the volume contains the "Category Display," twenty lists of identifiers arranged alphabetically within each subject category. These lists enable the searcher to scan lists of terms relating topics or to find specific indexing terms that might be difficult to locate alphabetically. Both displays are useful for novice and experienced ERIC searchers, providing a depth of indexing and retrieval not possible with descriptors alone. The *IAL* also includes the Used For (UF) and Scope Notes (SN) familiar to users of the *Thesaurus of ERIC Descriptors*.

As an example of the use of identifiers in ERIC CD-ROM databases, a search was performed on the phrase "course evaluation questionnaire," first using Easy Menu Search's Word/Phrase Index. The search displayed the phrase in the index, resulting in one retrieved record. When this same phrase was entered (selected from the *IAL*), using Words & Phrases as "Another Option" selection from the Easy Search Menu, the search retrieved fifteen records (including the one found previously), all of which contained "course evaluation questionnaire" in the identifier field. Several additional searches were performed using both the "Word/Phrase Index" and then the "Words & Phrases" Option to enter identifiers chosen from the *IAL*; all searches retrieved a larger number of pertinent records when identifiers were entered as a "Words & Phrases" search.

The *IAL 1992* is highly recommended as a resource or search tool for any library using ERIC in any format, whether computerized or in the *CIJE* or *RIE* print. Identifiers serve not only to broaden the intended search but also to pinpoint terms through the Category Display that might not otherwise come to mind. This volume is also necessary for anyone writing in the field of education and wishing to submit indexed materials to *CIJE* or *RIE* for inclusion in the database.—*Linda Gunter, The Claremont Colleges.* ■■

Gilmartin, Jacqueline, and Anne Beavan. *DYNIX: A Guide for Librarians and System Managers*. Brookfield, Vt.: Ashgate, 1992. 392p. $79.95 (ISBN 1-85742-010-1).

The title of the book is perfectly descriptive: It is a comprehensive overview of the Dynix integrated library system, with copious illustrations of command options, menus, and specimen record displays. Every significant Dynix command is illustrated, and many of the local options available to the Dynix client are presented. Unfortunately, the book was outdated by the time it reached copyright.

Gilmartin and Beavan have done an excellent job of assembling and organizing a great deal of information, but the book's utility is problematic, since it is based on Dynix Release 120, which was superseded in the U.S. marketplace in 1991. Release 140, two generations newer, is now in beta testing (April 1993) and is scheduled for general distribution by the end of the summer. It is distressing to contemplate the obvious work and care that has gone into this book for which there is either no market at all or at best a highly circumscribed one. The information presented in this book is of good quality, but the book's utility is negligible, its market limited, and its cost ($79.95) high.—*Harold D. Neikirk, Western Maryland College.* ■■

Information Sources in Patents. Ed. by C. Peter Auger. London: Bowker-Saur. 1992. 187p. $75 (ISBN 0-862981-906-0).

Ever since a 1474 Venetian statute established the importance and necessity of protecting inventors' rights and thus articulated the concept of intellectual property, debate over the law's usefulness has occurred across both ends of the spectrum. Competition in the current global village is shifting the focus of those debates to areas not even envisioned in 1474 but still in need of some form of protection.

Patents are a form of intellectual property. Because of its technical nature, patent documentation can be quite challenging in terms of approachability and can be daunting to the neophyte. Therefore, any new effort, like this work, that allows patents to be understood more easily or enhances accessibility of important information is always welcome.

After giving a brief history of patents, from a solidly British and European perspective, this work provides an overview of patents in Europe, the United States, and the United Kingdom in three chapters of, respectively, eleven, six, and twenty pages. Another chapter of six pages covers Germany, Japan, and Australia "chosen as representatives of the world's national patent systems." The chapter on the international aspects of patents is particularly useful. It addresses the Patent Cooperation Treaty, discusses the International Patent Classification, and explains the activities performed by the World Intellectual Property Organisation (WIPO), providing a listing of the unions and treaties it administers.

Where does one obtain information about patents? This work describes sufficiently well the British Patents Information Network and the American Patent and Trademark Depository Libraries system. Brief paragraphs are devoted to French and German patent and trademark registration systems. With regard to sources of information, all basic publications are listed and described for Europe, the United States, the United Kingdom, and the three other "representative" countries. Also described are other abstracting and indexing services such as the International Patent Documentation Centre (INPADOC) and the various Derwent publications.

One of the best chapters of this work covers all important online services (with the exception of Mead Data's LEXPAT). Three other chapters deal with "information needs in the engineering industries," "the importance of chemical patents," and "patenting in the life sciences." In these areas the relationship with published sources, limitations of current patent laws, and difficulties in standardizing descriptions are emphasized. A final chapter very briefly contemplates future developments and at the same time provides a short selection of materials for further reading.

Libraries wishing more detailed coverage should still rely on Rimmer's *International Guide to Official Industrial Property Publications*, Kase's *Foreign Patents* (which, incidentally, needs to be revised and updated), and Ardis' *Introduction to U.S. Patent Searching.*

Well written and concise, although obviously intended more for British readers, this work is valuable for those libraries in need of a well-packaged but general overview of patents. This is an unforbidding and quite helpful work.—*Erminio D'Onofrio, The New York Public Library.* ■■

Machalow, Robert. *101 Uses of LOTUS in Libraries*. Westport, Conn.: Meckler, 1992. 350p. $42.50 (ISBN 0-88736-791-7).

This volume updates and expands the author's previous book on Lotus, *Using Lotus 1-2-3: A How-to-Do-It Manual For Library Applications.* In this new book, Machalow details 101 ways Lotus can be used to handle library functions such as budget, reference, circulation, ILL, periodicals, acquisitions, cataloging, database services, and bibliographic instruction.

For each library application, the author begins with a concept, then adds step-by-step explanations with menu commands, followed by discussion of alternative macro language. Machalow further assists by suggesting Lotus' advanced built-in functions. For each section, every feature of Lotus is explored, and innovative ideas related to specific library functions demonstrated.

A large part of the book covers bibliographic instruction. In one example, a full-screen menu is designed on a Lotus worksheet to permit the user to choose bibliographic instruction lessons, which include a description of features of library catalog cards and user quizzes on the cards. Macros associated with the full-screen menu are shown. Worksheets containing the lessons have been created and are incorporated as subroutines into the menu.

As the author states in the introduction, the book's targeted audience is the experienced user; thus it focuses heavily on built-in functions and the macro command language. For intermediate users who have some understanding of Lotus, the book will be most valuable when used together with other Lotus reference books.

However, although the author provides many creative uses for Lotus, he fails to indicate in each application the limitations of Lotus, possible PC hardware constraints, or the limitations of database records. The spreadsheet function is the major feature of Lotus; other functions—e.g., database management and graphics—are less powerful. Some applications indicated in the book might be more effectively managed by other software or might be easily accomplished by using Lotus together with another package.

The book provides valuable illustrations and specific examples of finished products. The author explains the library applications fully and makes effective use of illustrations and sample screens. More references to the built-in functions and macros available in later releases of Lotus would have been helpful. Overall, however, this is an excellent book for Lotus users searching for myriad ways to use the program in libraries.—*Birong Ho, Wayne State University Libraries.* ■■

Macintoshed Libraries 5. Ed. by Edward J. Valauskas and Bill Vaccaro. Cupertino, Calif.: Apple Library Users Group, 1992. 97p., paper, free (No ISBN). Also available as a Hypercard stack from Apple Computer, Inc., 10381 Bandley Dr. MS:8C, Cupertino, CA 95014.

This book is the fifth edition of a popular series of books that describes innovative library applications using the Macintosh computer. Editor Edward Valauskas characterizes the sixteen case studies as adventures "where imagination triumphs" in the use of technology to meet library patrons' information needs. Of the sixteen articles, nine describe university library projects, four describe school libraries, two present creative programs in special libraries, and one discusses a public library system. Some of the programs appear to be transferrable to different libraries, but these are not projects for the faint-hearted. The ability to use HyperCard and understand scanning equipment—and the willingness to put in some long days—may be just some of the requirements for undertaking the projects described.

The subtitle of this book might have been *HyperCard in the Library*. Several articles describe the use of this software, mostly at university libraries, to develop training modules that introduce library staff and patrons to the online catalog and other library resources. Two articles describe workstation interfaces called Gateway and the Chemist's Crystal Ball; HyperCard is the program used to facilitate searching in different databases. The CORE (Chemistry Online Retrieval Experiment) project uses HyperCard to meet a unique challenge, that of converting electronic typesetting languages used in print journals (containing special characters, signs, etc.) into compatible formats for electronic display. Still another article describes digitized document transmission using HyperCard.

Immigrant is a HyperCard stack program that has been written about elsewhere but has a special place in this collection. High school students not only have access to historical archives for their own research, they also participated in the collection and scanning of text and graphics using portable computer equipment. Similarly, another article describes how a Virginia school district, with a technology-intensive high school known as the "Smart School," is building a collection of scanned images and historical and science-related data using a variety of software in addition to HyperCard. Singing Light, one of several programs funded by an ALOT (Apple Library of Tomorrow) grant, incorporates HyperCard into the development of an innovative multimedia database of Indian history on CD-ROM.

The remaining articles in this collection cover general hardware considerations and some special library applications of software such as File Maker Pro. There is a general index and a special index of hardware and software vendors. The total effect of this book is one of inspiration. *Macintoshed Libraries 5* meets its goal as stated in the introduction—to stretch the technology to meet the challenge of providing information where, how, and when it's needed. Many examples of successful applications are packed into this small (free!) book.—*Catherine Murphy, Three Village Schools, Stony Brook, New York.* ■■

Morris, Anne. *The Application of Expert Systems in Libraries and Information Systems Centres*. London: Bowker-Saur, 1992. 241p. $65 (ISBN 0-86291-276-8).

Expert Systems (ES) seem a marvelous panacea for the overworked librarian, but the fact is few are in operation today, and these seem to work best when developed and used for a specific purpose. The technology is still at an evolutionary stage and is far from widely utilized.

Do we need yet another book on ES? Yes. This text makes a valuable contribution to our knowledge and understanding of ES. It was written to review the progress made thus far in applying ES technology to library and information work. We have seen *Expert Systems in Reference Work* (Roysdon and White, 1988), *Expert Systems in Libraries* (Aluri and Riggs, 1990), *Expert Systems in ARL Libraries* (Bailey and Myers, 1991), and numerous journal articles. The Morris text enhances the extant literature.

The text is divided into seven chapters, each with a different author, which lead the reader logically and clearly to an understanding of expert systems and their relevance to today's libraries. Anne Morris provides both an introduction and an overview of expert systems generally, listing their characteristics, importance, applications, and components. She then discusses ES tools—artificial intelligence programming languages, expert system shells, knowledge engineering environments—and the selection of these to develop ES. Her explanations are easily understood by the thinking librarian.

Forbes Gibbs successfully tackles the problem of explaining knowledge-based indexing. This is the toughest chapter to read but is well worth the attempt because the concepts are necessary to our understanding of ES work. Bill Black follows with an explanation of the linguistic aspects of the information accessing process in his chapter describing the links between rule-based systems, natural language processing, and abstracting.

The next three chapters deal with more practical examples of the use of ES. Roy Davies, Alastair Smith, and Morris discuss possible applications of expert systems in the reference process in relation to several models of that process. They describe design issues and discuss systems already implemented or being developed. Davies writes about the use of ES in cataloging, in which the process of creating and maintaining the catalog is surprisingly complex. Gwyneth Tseng describes application of ES to simplify online information retrieval.

The final chapter, Ralph Alberico's essay on the future of ES, should be read by everyone interested in the near-future development of existing library technologies. It includes a scenario for a physics knowledge base which exemplifies that needed for any discipline. He cautions that the next generation will likely bear little resemblance to current technology.

Each writer explores his or her area very well, from historical development to present possibilities and future needs, and presents many examples. Each chapter has its own pertinent bibliography. Overall the book is an honest appraisal of the history and capabilities of ES today, with some glimpses into the near future. The text is very readable, particularly to those who regularly read *ITAL*, and contains a brief glossary of terminology. Several of the writers are not Americans, which adds a welcome international flavor to the writing and a few quirks in spelling.

The most glaring flaw appears in the preface, where Morris introduces the development of the chapters. The chapter order was obviously changed later into a more logical flow of thought and one chapter was added.

Readership for this book is widespread: it can serve as a valuable introduction for the novice and will be useful for those curious about the ways in which present and future technologies might affect the provision of library services—in other words, all of us! It is most valuable for its clear explanations and context-based examples.—*Valerie Jackson Feinman, Adelphi University Libraries.* ■■

Retrospective Conversion: History, Approaches, Considerations. Ed. by Brian Schottlaender. New York: Haworth, 1992. 167p. $24.95 (ISBN 1-56024-328-7).

This book, also published as volume 14, numbers 3 and 4, of *Cataloging and Classification Quarterly*, consists of a collection of papers edited and introduced by Brian Schottlaender. The work is divided into four sections: History, Case Studies, Coordination and Innovation, and Control Issues. Each of these sections is composed of two articles. The History section contains a bibliography and review of the retrospective conversion literature published between 1980 and 1990 compiled by Daphne C. Hsueh and a history and status report on the Library of Congress PREMARC file by Susan H. Vita.

Case studies are often a way to discover what others have done in dealing with issues. Mary K. Bolin and Harley B. Wright share their experience in the retrospective conversion of the University of Idaho's catalog. Different approaches were used at different times to convert this medium-sized academic library, and Bolin and Wright provide comparisons of the approaches used. Ruth Christ and Selina Lin share the University of Iowa's experience in converting its serial titles. They include a helpful list of questions they apply to serial conversion projects but that could also be used for any retrospective conversion project.

The first article in the Coordination and Innovation section compares and contrasts the handling of music scores in retrospective conversion by seven different libraries. Michelle Koth and Laura Gayle Green summarize the workflow and give the advantages and disadvantages of the process for each library.

Beth Sandore reviews the procedures used in converting a reading room collection of a computer science department. Programming was done to allow most of the work to be done at a personal computer.

The Control Issues section begins with a selective review of the literature on editing recon records. Edward Adrian Lentz surveys the literature to determine the standards to which theorists and practitioners adhere and the impact these standards have on the cost of recon projects. James Tilio Maccaferri discusses the various ways authority control can be incorporated in a conversion process. He covers conducting authority work in-house, sending work to a vendor for cleanup, and combining the two approaches.

The articles are all written by people who have undergone a conversion project and who have information of interest to those involved in their own projects. Anyone involved in retrospective conversion will find these articles useful, thought-provoking, and well worth their time.—Susan Moore, University of Arizona Library. ■■

Stieg, Margaret F. *Change and Challenge in Library and Information Science Education*. Chicago: American Library Assn., 1992. 206p. paper $37 (ISBN 0-8389-0576-5).

This book addresses the continuing debate about the scope and quality of education for schools with ALA-accredited programs. Librarianship, information science, and archival administration are the focus of this overview of issues. One of the author's purposes is to foster understanding among the specialists in this information profession triad. In order to do this, she presents the historical context of the debate and the past and current direction of professional education.

All the major works that have contributed to our understanding of education for traditional librarianship are examined, and the roots of archival education and information science education explored. Stieg relies on the C. C. Williamson report on library education, published in 1923, to help us focus on how we got into the university (and what it has cost us), and she notes the contributions of the other professions that helped form the service ideal of the American university system. She strikes a balance between the demands of the practice and those of the university setting. The link between goals and objectives of the schools and the influence of the standards on these is firmly established. Stieg comments on the various administrative possibilities, the role of the dean, and the critical need to define leadership.

Stieg doesn't deny the importance of technology but questions its role in the schools and voices the fear that it is an end in itself. A significant thrust of the work is the call for more regional and national planning, but the difficulties of achieving this are not oversimplified. Librarians, archivists, and information scientists have different ways of defending their professional status, and Stieg offers a rich variety of sources from which to draw new conclusions. She addresses the qualifications needed by faculty in research institutions and the need for careful planning for a curriculum that must have foundation courses and electives for specialization. Distance education and continuing education are seen as viable options for course delivery because of the increasing number of part-time students in the schools.

The process of accreditation is carefully explained. The need to understand the dilemma that pleasing several accrediting agencies creates is examined, the acceptance by ALA of NCATE (National Council for the Accreditation of Teachers' Education) being one example of this. Stieg also raises the question of certification. The recent development of certification within the SAA (Society of American Archivists) is briefly mentioned.

This book can be the stimulus for open discussion between educators and practitioners as we prepare for incipient changes. It is a call to action. Each reader can use it as a catalyst to determine the overarching question asked by Stieg: "What is in the best interests of the professions and the clienteles they serve?" The work is in eleven chapters, has an excellent bibliography, and an index of sufficient depth.—*Kathleen J. M. Haynes, The University of Oklahoma.* ■■

Virtual Worlds, Real Challenges: Papers From SRI's 1991 Conference on Virtual Reality. Ed. by Teresa Middleton. Westport, Conn.: Meckler, 1992. 119p. paper, $35 (ISBN 0-88736-870-0).

For most people, virtual reality begins and ends with war games, where teams use high technology to do battle. There is another, more serious side though, one presented in *Virtual Worlds*, a collection of papers delivered at a conference sponsored by SRI International (formerly Stanford Research Institute) and VPL Research. A few of the papers may be too far removed from the library world and too technical in nature to warrant more than a quick glance. However, the book as a whole should be read by anyone wanting a greater understanding of virtual reality and its potential.

Virtual reality is still in its infancy, and researchers are intent on developing and refining new and better tools. The excitement of this pioneering group is communicated in work that challenges its own assumptions and strives to imagine a different reality. Developers of virtual reality gathered in June 1991 to discuss systems development, applications, and human interface issues. Some of the applications discussed were assistive technology for the disabled, surgical simulators, and genetic imaging systems. Readers may never have need to use a stereoscopic viewer like the one being developed at the Fake Space Labs (now Fakespace, Inc., of Menlo Park, California), but it is instructive to learn about the design considerations and philosophy of this and other tools.

Virtual reality has the potential to take us beyond our physical limitations and guide us to new insights and knowledge. Complex computer technologies are proliferating and becoming more difficult to manage. Virtual reality uses technology to create human-machine interfaces that permit control of multiple peripheral devices with relative ease.

Librarians, particularly, may be interested in the work being done on virtual reality for the disabled. VPL Research's DataGlove uses fiber-optic sensors to replicate hand movements. Other interface devices simulate visual or auditory experiences. Using one of these tools may give a disabled person an entirely different view of the world, and the world a much fuller appreciation for the disabled.

Even though it dates from two years ago, the research reported at this conference is far ahead of most current library applications. If you think you know what virtual reality is about, read this and increase you knowledge. Even if you know nothing about virtual reality, scan this book. Reading selected papers will help you appreciate a new and exciting field and create new approaches to delivering information and services. Librarians are beginning to envision the virtual library; this book may serve as a catalyst for thought.—*Mary Alice Ball, Loyola University.* ■■

Software Reviews

Pro-Cite for the Macintosh. Personal Bibliographic Software, P.O. Box 4250, Ann Arbor, MI 48106; (313) 996-1580. Hardware and software requirements: Mac Plus or larger, using Macintosh system software version 6.0.2 or higher. Hard disk highly recommended. Price: Single user, $395; multiuser packs available.

Pro-Cite, now in version 2.1, is probably the most powerful personal bibliographic database management system available for the Macintosh. Others can be found that are cheaper and simpler to use, but these are generally called reference managers, rather than database managers, and thus provide a smaller set of functions than Pro-Cite.

Pro-Cite is oriented toward the building of databases and the production of bibliographies. One builds and maintains a database either by entering data by hand or by importing data from another file. Databases or subsets may be sorted by a variety of sort keys, and editing operations may be performed globally on multiple records. The database can be searched to create subsets. Selected records from the database may be formatted and printed or saved to disk as bibliographies or exported in formats readable by other database systems. It is possible to use a Pro-Cite database to generate a bibliography from in-line citations in a word-processing document, but Pro-Cite does not offer the functionality of generating in-line footnotes. Version 2.1 is network aware, allowing multiple users on networked Macintoshes to work with the same database. Only one user may access the database for editing, although several users may be using it in read-only mode. A reduced-price read-only version of the software, allowing a user to read existing

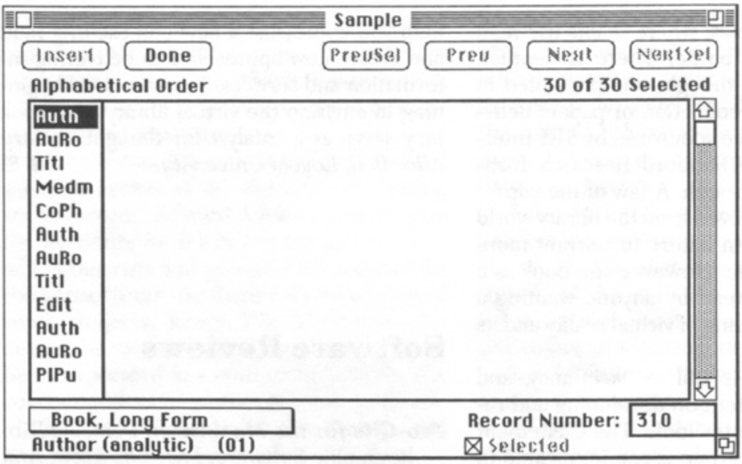


Figure 1. The Database Window.

databases but not to update them, is in progress.

According to the product documentation, Pro-Cite will manage up to 100,000 records in a single database, with a maximum size of 32,000 characters per record.

To build or edit the content of a database, one works with the database window (see figure 1).

The field labels are predefined by Pro-Cite and appear in a column at the left of the data. Inexperienced users may find difficulty with the field name abbreviations, especially since more than one field may have the same abbreviation. There are twenty-six predefined work forms (in figure 1, "Book, Long Form"; a few other work forms are "Journal, Short Form," "Dissertation," "Letter"), which contain the likely fields for each type of citation. New work forms can be created, and old ones modified, but the user needs to beware, because each field is linked to an internal field number with certain immutable qualities that do not change with a change in field label. By using the appropriate menu choices and clicking on the correct buttons, one may add new records or edit existing ones.

It is also possible to import records en masse from another database system. Importing a file saved from an online search service session or a CD-ROM requires a separate program called Biblio-Link, purchased individually for each service. A Biblio-Link program can also be purchased for importing raw MARC records. Pro-Cite can directly import data that has been formatted in tab-delimited or comma-delimited format, requiring that all defined fields exist in the proper order. Several OPAC vendors now provide the ability to download records in Pro-Cite's import format.

The output of a database is a bibliography. The key to formatting a bibliography is the punctuation file, which specifies the formatting of each field for output. A number of predefined punctuation files are provided, including all the predominant citation formats (MLA, Chicago, Science, Turabian, and so on). By choosing a punctuation file and then selecting the PRINT BIBLIOGRAPHY or SELECTED RECORDS TO DISK commands, one can print or save the formatted bibliography. New punctuation files can be created for a personal style or for a template that can be used for compiling load-ready files for other database systems.

Pro-Cite offers powerful searching features. Full-text searching, although slow, allows you to find any character string anywhere in any record. Limited searching, also a slow sequential scan of the database, searches only specified fields. Quick searching, by far the fastest, is a true index search. The user specifies which fields will be indexed for quick searching. Boolean searches are supported. Search strategies and result sets can be saved for later use. The result of a search is that certain records in the database are selected. When the search is complete, one returns to the database to review the selected records. It is possible to deselect unwanted records at this point. Rather than a summary display of brief records, only the complete record is available for display and only one record can be viewed at a time (three in the browse mode). The search window can be opened again, and the results further refined by adding more search criteria. When the search is complete to the user's satisfaction, the selected records may be sorted, globally edited, or formatted for a bibliography.

Pro-Cite offers the capability of creating and maintaining authority lists. An authority list is an independent list of terms that can be used to control the vocabulary in a database. Any list that has been stored on the Macintosh disk can be called up for use in any field in any database. Authority lists may be made in a variety of ways. One can add terms to a list while using Pro-Cite, create a list in a word processor, or generate a list from all the terms found in an existing database (this would be a good way to review all the terms that have been used in a database), or one can use one of the many lists that are supplied in the Pro-Cite package. When building a database, one uses the authority lists by moving the cursor to the place where the new term is to be added, opening the desired authority list, highlighting the term or terms to be used, and clicking on the OK button in the authority list. Unfortunately this tedious procedure must be repeated for each field being edited; it is not possible to enter data into the database work form while an authority list is open, and only one authority list can be open at a time. One of the features of authority lists is the ability to add alternate text to authority entries. At the point of generating a bibliography, one can specify whether to use the primary or alternate text in the bibliography, thus achieving to a small extent the function of *see* references. Unfortunately the length of an authority list is limited to 32,000 characters, making extensive lists cumbersome.

Pro-Cite's user documentation is thorough and easy to follow, well indexed, and with a good table of contents. The instructions are neatly divided into tabbed sections. A single folded quick-reference card summarizes all the operations. Online help is available while you are using the program.

With all its powerful features, the only real flaw in this package is its lack of observance of the Macintosh user interface guidelines. Many of the operations are needlessly modal; that is, one must enter a function, perform an operation, and then exit the function before going on to anything else. At many points only one window can be viewed, while it would be useful to be able to switch between several open windows; only one database, one record (although three may be viewed in browse mode, they are still in a single window and can't be opened or closed one by one), one authority list, and so on may be viewed at a time. Even the help window must be closed before returning to work with the database. Now that many users have large screens, the ability to keep multiple windows open would be an advantage. The dialogue boxes do not tend to make full use of the Macintosh's graphical and WYSIWYG capabilities; for example, when building a punctuation file, one types the prepunctuation in a box, then clicks OK to move it into another box, rather than simply typing it into the destination box. In general, Pro-Cite feels more like a converted IBM program than one that was designed using the Macintosh mental model, and Macintosh aficionados will find this frustrating because they cannot rely on their Macintosh mind-set to figure out how to use the program.

Personal Bibliographic Software (PBS) provides a variety of customer-service options; the user manual gives a phone number, a fax number, a U.S. Postal Service address, and e-mail addresses on four different mail systems. This reviewer experimented with electronic mail and received a response virtually instantaneously. There is also a Bitnet listserv (discussion group) for Pro-Cite users.

If you are interested in managing footnotes or one-time bibliographies for specific papers, Pro-Cite is probably needlessly complicated. If you are building a database with a long lifetime and a variety of uses, then constructing a Pro-Cite database is well worth the effort.—*Katherina Klemperer.* ■■

UNIX for (Almost) Any Macintosh. MachTen Kernel System Software. Tenon Intersystems, 1123 Chapala Street, Santa Barbara, CA 93101; 1-800-6-MACH-10; Internet: INFO@TENON.COM. Hardware and software requirements: Macintosh Classic or higher (MachTen will also run on a Mac Plus or a Mac SE); 4MB RAM for System 6.0.x, 5MB RAM for System 7.0.x; 40MB hard disk (Kernel

System Software will occupy just short of 20MB of disk space). Price: Workstation license, $595; server license, $895; optional documentation, $175; academic pricing and site licensing available.

MachTen Kernel System Software, by Tenon Intersystems, offers a port of the BSD 4.3 Reno version of the UNIX operating system for Macintoshes ranging from the Plus to the Quadra. Instead of taking over as the operating system of the Macintosh, as does Apple's AUX version of UNIX, MachTen runs as an application on top of the Macintosh operating system. MachTen is not a UNIX emulator or merely a set of UNIX-like tools, however; it is real multiuser and multitasking UNIX.[1] This architecture yields a very satisfactory integration of Macintosh and UNIX. For instance, under MachTen one can use the Macintosh MultiFinder or the usual UNIX file-handling commands to access other files or programs. The user can also copy and paste between MachTen and Macintosh applications. Macintosh applications can work on UNIX documents created in MachTen, just as MachTen can work on documents prepared with Macintosh applications. Most Macintosh printers are supported, with the exception of the Stylewriter, which Tenon says it will support in an upgrade due out in 1993. (It may be available by the time this review is published.)

APPLETALK TO ETHERNET

The most salient feature of UNIX is its ability to support multiple users and multiple tasks, and MachTen offers a full range of choices. MachTen will run on an AppleTalk network, though the slowness of AppleTalk translates into slow response times for UNIX commands executed over this network configuration. For institutions that do not have the luxury of a speedier LAN configuration, the ability to run over AppleTalk is a plus, however. MachTen also supports Ethernet, as well as serial communications.

SHARING FILES WITH NFS

MachTen has implemented Sun Microsystems' Network File System (NFS) to enable Macintoshes to share files, not only with other Macintoshes, but also with other machines that support the NFS standard. The ability to set up a Macintosh as an NFS client or as a server are included in the MachTen basic package. Using a distributed file system, a user can access files on a server in the same way as if the files were on her local machine. If the user has "mounted" directories from the server, the directories on the remote machine will appear in the user's home directory as if they were resident on a local hard disk, though they remain stored on the server.

LINKING TO THE OUTSIDE WORLD

For those sites without direct connections to the Internet, MachTen offers some serial communications solutions. MachTen supports SLIP (Serial Line Internet Protocol) connections that allow a Mac running MachTen to dial up a cooperating remote Internet host that also supports SLIP and establish the local Mac as a sort of virtual Internet host. UUCP (UNIX-to-UNIX copy) also numbers among the features included with the MachTen basic package. UUCP offers a way of linking a Mac running MachTen to a global network of UNIX machines that run UUCP. Access to USENET and the ability to send mail to colleagues with USENET addresses as well as to exchange files are also part of what UUCP offers the user. For those sites fortunate enough to have a direct connection to the Internet, MachTen also offers a complete set of Internet protocols, including Berkeley TCP/IP, FTP, Telnet, UDP, and SMTP.

SHELLS AND UTILITIES

MachTen has implemented two of the three most popular UNIX shells: the Bourne Shell and the C Shell. Those who are familiar only with the DOS shell may be pleasantly surprised by the flexibility and programmability that these shells offer when compared with the MS-DOS shell. Of course, there is a sharper learning curve in mastering these more powerful tools. As with any UNIX system, MachTen comes with a large set of utility programs, ones that are found on nearly every UNIX system. Text-handling tools such as sed, awk, grep, sort, and so on have great potential for use in libraries for manipulating text files and extracting data from them.[2] In the case of awk, which has undergone several revisions over the years, MachTen contains an up-to-date version of this text-oriented programming language. Similar to C, awk has built-in text-handling features, so it is easier to use on text files. It is a good beginning language for the novice programmer mainly interested in text manipulation.

DOCUMENTATION

The documentation that comes with the basic package is rather spare. The sixty-five page *User's Guide* introduces many of MachTen's features but does not go into much detail. *MachTen UNIX Basics* and the manual *MachTen Text Processing* contain reprints of old but classic articles on using UNIX. The *System Architecture* booklet discusses the relationship between MachTen and the Macintosh operating system and MachTen's capabilities in a networking environment. Installation instructions as well as some other documentation are supplied as offprints. One wishing to learn UNIX basics would do well to purchase one of the many excellent introductions to UNIX now on the market. On the other hand, a veteran UNIX user who needs some technical information will not find it in the material included with the basic package. One will find the standard MAN pages available from the UNIX command line or as a HyperCard stack. Oddly enough, the MAN stack does not take effective advantage of HyperCard by including links among MAN entries.

It is frustrating to discover that there is not sufficient documentation in the basic package to use the many exciting networking and communications tools included in it. Fuller documentation is available for a separate charge, as noted in "Beyond the Basic Package," below. Users might also consult BSD UNIX manuals and other UNIX works in lieu of or in temporary place of the documentation supplied by Tenon.

INSTALLATION

Installing the ten 800K MachTen diskettes on a Macintosh LC was fast (twenty minutes) and painless. The time to set up users and configure MachTen for a networked environment will vary from site to site.

BEYOND THE BASIC PACKAGE

Veteran UNIX users will wonder about the C language-development tools and X-Windows capabilities. Though the Kernel System Software does not include any of the C tools normally expected on a UNIX system, these may be purchased separately as the Development System. Also available separately for an additional charge is X Window software. Libraries will probably be interested in the Development System because its make and compile tools allow the user to download and compile UNIX shareware and freeware. The Development System's list price is $395, and it includes the GNU C and C++ compilers as well as such programming tools as make, yacc, lex, and lint.

CONCLUSION

Libraries that want to slip into the UNIX world in an inexpensive way would do well to consider MachTen. Many libraries already have Macintoshes capable of running this implementation of UNIX. The cost per machine of the software could be quite reasonable given the various pricing options. In times when equipment budgets are tight, requests for new software that will put existing computers to new uses may well have a greater chance of finding funding than requests for new workstations and new software. Besides its potential economies, MachTen also allows Macintosh sites to continue employing the user-friendly system they have come to like in addition to enjoying the powerful capabilities of UNIX.—*Brian Sealy, University of Michigan, Ann Arbor*.

NOTES

1. See the reviewer's review of MKS Toolkit, a set of UNIX tools for MS-DOS computers, in the December 1992 issue of *Information Technology and Libraries*, p. 448–51.
2. See the author's tutorial article on p. 270–76 of this issue on using sed to edit downloaded OPAC files. ■■

Other Recent Receipts

Agenbroad, James Edward. *Nonromanization: Prospects for Improving Automated Cataloging of Items in Other Writing Systems*. Washington, D.C.: Cataloging Forum, Library of Congress, 1992. 16p. free (No ISBN), LC Card No. 92-36934.

Advances in Online Public Access Catalogs. Vol. 1, 1992. Ed. by Marsha Ra. Westport, Conn.: Meckler, 1992. 190p. $55 (ISBN 0-88736-775-5).

Advances in Preservation and Access. Vol. 1, 1992. Ed. by Barbara Buckner Higginbotham and Mary E. Jackson. Westport, Conn.: Meckler, 1992. 296p. $55 (ISBN 0-88736-787-9).

Baker, Kim, and Sunny Baker. *Color Publishing on the Macintosh: From Desktop to Print Shop*.

New York: Random House Electronic Publishing, 1992. 479p. paper, $45 (ISBN 0679-73977-7).

Books & Periodicals Online: The Guide to Business and Legal Information on Databases and CD-ROMs. 1993 ed. Ed. by Nuchine Nobari. New York: Library Alliance, 1992. 1632p. paper, $249 (ISBN 0-9630277-1-9).

Cohn, John M., and others. *Planning for Automation: A How-To-Do-It Manual for Librarians*. New York: Neal-Schuman, 1992. 116p. paper, $39.95 (ISBN 1-55570-120-5).

Cumulative Index to ONLINE, DATABASE & CD-ROM Professional 1986–1991. Comp. by Linda Fetters. Wilton, Conn.: Eight Bit Books, 1992. 224p. paper, $29.95 (ISBN 0-910965-06-4).

Directory of Library Automation Software, Systems and Services. Ed. by Pamela Cibbarelli. Medford, N.J.: Learned Information, Inc., 1992. 350p. $79 (ISBN 0-938734-65-2).

European Directory of Software for Libraries & Information Services. Ed. by Library Information Technology Center. Brookfield, Vt.: Ashgate Publishing Co., 1992. 180p. $58.95 (ISBN 1-85742-092-6).

Everything You Wanted to Know About the Mac. Ed. by the Hayden Development Group. Carmel, Ind.: Hayden, 1992. 1,000p. paper, $29.95 (ISBN 0-672-30142-3).

FAX: Digital Facsimile Technology & Applications. 2d ed. By Kenneth McConnell and others. Norwood, Mass.: Artech House, 1992. 400p. $78 (ISBN 0-89006-495-4).

Gersitz, Lorraine, and Margie Epple. *Managing the Circulation Process: A How-To-Do-It Manual for Librarians*. New York: Neal-Schuman, 1992. 150p. $35 (ISBN 1-55570-138-8).

House, David. *Information Technology in Polytechnic Libraries*. Brookfield, Vt.: Ashgate Publishing Co., 1992. 200p. $50 (ISBN 0-566-05210-5).

Information Management and Organizational Change in Higher Education: The Impact on Academic Libraries. Ed. by Gary Pitkin. Westport, Conn.: Meckler, 1992. 152p. $42.50 (ISBN 0-88736-842-5).

Information Management: From Strategies to Action 2. Ed. by Blaise Cronin. Medford, N.J.: Learned Information, 1992. $88 (ISBN 0-85142-281-0).

Information Technology—It's for Everyone! Proceedings of the LITA Third National Conference, 1992. Library and Information Technology Association, Denver, September 13–16, 1992. Ed. by Thomas W. Leonhardt. Chicago: American Library Assn., 1992. 206p. paper, $22 ($19.80 to LITA members) (ISBN 0-8389-7634-4).

Information Technology in Special Libraries. Ed. by Margaret Brittin. New York: Routledge, 1993. 224p. $49.95 (ISBN 0-415-03824-3).

Intner, Sheila S., and William E. Studwell with Simone E. Blake and David P. Miller. *Subject Access to Films and Videos*. Lake Crystal, Minn.: Soldier Creek, 1992. 133p. paper, $25 (ISBN 0-936996-60-9).

Integrating Technology into the Curriculum. Ed. by Les Lloyd. Supplement to Computers in Libraries Series, no. 62. Westport, Conn.: Meckler, 1993. 200p. $39.50 (ISBN 0-88736-867-0).

Jukebox and Robotic Libraries for Computer Mass Storage. Ed. by Sanjay Ranade. Westport, Conn.: Meckler, 1992. 128p. $49.50 (ISBN 0-88736-571-X).

Lane, Elizabeth, and Craig Summerhill. *An Internet Primer for Librarians & Educators: A Basic Guide to Internet Networking Technology*. Westport, Conn.: Meckler, 1992. 175p. $37.50 (ISBN 0-88736-831-X).

Library Technology Consortia: Case Studies in Design and Cooperation. Ed. by Bruce Flanders. Supplement to Computers in Libraries Series, no. 70. Westport, Conn.: Meckler, 1993. 165p. $42.50 (ISBN 0-88736-886-7).

McClure, Charles M., and others. *Public Libraries and the INTERNET/NREN: New Challenges, New Opportunities*. Syracuse, N.Y.: School of Information Studies, July, 1992 Report. 34p. $15 (no ISBN).

Meadow, Charles T. *Text Information Retrieval Systems*. New York: Academic, 1992. 303p. $45 (ISBN 0-12-48710-X).

MeckJournal: An Electronic Monthly from Meckler Publishing. Monthly, 1991– . Westport, Conn.: Meckler, 1991. Free. Available from the Internet at: meckler@jvnc.net. (ISSN: 1058-692X).

Michael, James J. *From A to Z39.50: A Networking Primer*. Westport, Conn.: Meckler, 1992. 150p. paper, $29.50 (ISBN 0-88736-766-6).

Miller, Heather S. *Managing Acquisitions and Vendor Relations: A How-To-Do-It Manual*. New York: Neal-Schuman, 1992. 191p. paper, $35 (ISBN 1-55570-111-6).

Networks for Networkers II: Critical Issues for Libraries in the National Network Environment. Ed. by Barbara Evans Markuson and Elaine W. Woods. New York: Neal-Schuman, 1992. 200p. $45 (ISBN 1-55570-128-0).

Pennington, Catherine A. *Microcomputer Software Selection for the Law Library, Pt. Two: Library & Legal Software*. Law Library Information Reports, vol. 13. Dobbs Ferry, N.Y.: Glanville Pubs., 1992. 152p. paper, $100 (ISBN 0-87802-088-8).

Performance Appraisal of Collection Development Librarians. Systems and Procedures Exchange Center Kit 181. Comp. by Jack Siggins. Washington, D.C.: Assn. of Research Libraries (Office of Management Services, 1527 New Hampshire Ave., NW, Washington, DC 20036), 1992. 109p. paper, $33 ($22 for ARL members) (no ISBN).

Retrospective Conversion: History, Approaches, Considerations. Ed. by Brian Schottlaender. New

York: Haworth, 1992. 167p. $24.95 (ISBN 1-56024-328-7).

Rowley, Jennifer, and Shelagh Fisher. *Bookshelf: A Guide for Librarians and Systems Managers*. Brookfield, Vt.: Ashgate, 1992. 283p. $79.95 (ISBN 1-85742-008-X).

The Software Encyclopedia 1992: A Guide for Personal, Professional, and Business Users. 2 vols. New York: Bowker, 1992. 2,200p. $209.95 (ISBN 0-8352-3180-1).

Stankus, Tony. *Making Sense of Journals in the Physical Sciences: From Specialty Origins to Contemporary Assortment*. New York: Haworth, 1992. 239p. $39.95 (ISBN 1-56024-180-2).

Strengthening Technological Capabilities: A Challenge for the Nineties. 3d ed., revised. Geneva: International Labour Office, 1992. 113p. paper, $12 (ISBN 92-2-108165-6).

Studies in Multimedia: State-of-the-Art Solution in Multimedia and Hypertext: Proceedings of the 1991 Mid-year Meeting of the American Society for Information Science, San Jose, California, April, 1991. Ed. by Susan Stone and Michael Buckland. Medford, N.J.: Learned Information, 1992. 263p. $39.50 (ISBN 0-938734-59-1).

Thinking Robots, An Aware Internet, and Cyberpunk Librarians. Ed. by R. Bruce Miller and Milton T. Wolf. Chicago: Library and Information Technology Assn., 1992. 206p. $22 ($19.80 to LITA members) (ISBN 8389-7625-5).

Urbanski, Verna, with Vao Chu Chang and Bernard L. Karon. *Cataloging Unpublished Nonprint Materials: A Manual of Suggestions, Comments, and Examples*. Lake Crystal, Minn.: Soldier Creek Pr., 1992. 144p. paper, $22.50 (ISBN 0-936996-61-7).

Virtual Reality Market Place 1993. Ed. by Sandra K. Helsel and Susan Denoble Doherty. Westport, Conn.: Meckler, 1992. 180p. $55 (ISBN 0-88736-795-X).

Watters, Carolyn. *Dictionary of Information Science & Technology*. San Diego, Calif.: Academic Pr., 1992. 300p. $29.95 (ISBN 0-12-738510-X).

Westlake, Duncan, and John Clarke. *Geac: A Guide for Librarians and System Managers, Second Edition*. Brookfield, Vt.: Ashgate, 1992. 328p. $79.95 (ISBN 1-85742-000-4).

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

Wilkins, Peter. *Managing New Technological Change: Case Studies in the Re-Organization of Work*. Brookfield, Vt.: Ashgate Publishing Co., 1992. 235p. $58.95 (ISBN 1-85628-336-4).

Working with Faculty in the New Electronic Library. Ed. by Linda Shirato. Ann Arbor, Mich.: Pierian, 1992. 192p. $35 (ISBN 0-87650-291-5).

Wolverton, Van, and Michael Boom. *Van Wolverton's Guide to Windows 3.1*. New York: Random, 1992. 220p. paper, $20 (ISBN 0-679-73912-2).

Wolverton, Van, and Steve Lambert. *Van Wolverton's Guide to DOS 5*. New York: Random, 1992. 194p. paper, $20 (ISBN 0-679-73913-0).

Woodsworth, Anne, and James Williams. *Managing the Economics of Owning, Leasing and Contracting Out of Information Services*. Brookfield, Vt.: Ashgate Publishing Co., 1992. 200p. $49.95 (ISBN 1-85742-018-7). ■■

INDEX TO ADVERTISERS

News and Announcements

New Booklet from NOTIS Explains Buzzwords

Open systems, client/server architecture, and *Z39.50* all have become buzzwords in library management. But not everybody has a clear understanding of what they mean and what their impact on library services will be. A new booklet, *Open Systems and Your Library*, written by Gary Lee Phillips, provides a brief, thorough overview of these emerging technologies and cites specific examples of how they can expand patron usage and streamline library activities. To obtain a copy of *Open Systems and Your Library*, contact Linda Sullivan at NOTIS, 1007 Church St., Evanston, IL 60201; (708) 866-0150.

GUIDON Interface Improves Internet Access to Electronic Journal

With the release of GUIDON 1.1 software, subscribers to the *Online Journal of Current Clinical Trials* now have access to the electronic journal's complete text and graphics through the Internet.

GUIDON is a graphic user interface developed by OCLC for electronic journals. With GUIDON, subscribers view the full text of articles with graphics and hypertext links to references, figures, equations, and tables. GUIDON also supports sophisticated Boolean searching and a document-ordering capability.

OCLC plans future releases of GUIDON that will support its searching and graphic capabilities on the X-Windows and Macintosh platforms.

Two interfaces are available for viewing, displaying, and printing *Online Journal of Current Clinical Trials* documents. GUIDON, which is sent to all subscribers, runs in the Microsoft Windows environment on an 80286 or faster PC and provides typeset-quality displays and graphics. EPS, the OCLC Electronic Publishing Service, is a command-oriented ASCII user interface that runs on a terminal or PC with software emulating a VT100 terminal. Both interfaces are now available via the Internet, the OCLC Network, and CompuServ.

The *Online Journal of Current Clinical Trials*, introduced last July, is a peer-reviewed medical journal published by the American Association for the Advancement of Science and distributed electronically by OCLC.

For more information call Andrea Keyhani at (614) 764-6474 or Nita Dean at (614) 761-5002.

Second IFLA Satellite Meeting

The second IFLA Satellite Meeting on Automated Systems for Access to Multilingual and Multiscript Library Materials will be held in Madrid on August 18 and 19, 1993. The meeting was organized jointly by IFLA's Section on Library Services to Multicultural Populations and its Section on Information Technology in cooperation with the Section on Cataloguing and the Biblioteca Nacional, Madrid.

Prominent specialists from Europe, North America, and Asia will speak on topics such as new developments in automation of various scripts, ISO standards, design of multilingual and multiscript catalogs, graphic displays for OPACs, cataloging and filing rules for universal character sets, and issues in telecommunications.

Registration is U.S. $100 and accommodation is at the Residencia de Estudiantes at U.S. $85 per night, tax and breakfast included. The registration deadline is July 15, 1993.

For additional information and registration forms, contact Marie Zielinska, Multilingual Biblioservice, National Library of Canada, 395 Wellington St., Ottawa, Ontario, Canada, K1A 0N4; (819) 997-9931; fax (819) 953-6984; or Belen Altuna, Unidad de Coordinacion Informatica, Biblioteca Nacional, Paseo de Recoletos, 20, 28071 Madrid, Spain; 580-7872; fax 580-7873.

University of South Carolina College of Library and Information Science Celebrating its 20th Year

The College of Library and Information Science at the University of South Carolina is celebrating twenty years of leadership and service to the profession. In September 1972 the first class enrolled in the first professional seminar, and in 1974 the college won ALA accreditation. Founding dean Wayne Yenawine was succeeded in 1976 by F. William Summers (now dean of the School of Library and Information Studies at Florida State University), in 1982 by acting dean John Olsgaard, now associate provost of the university; and in 1986 by current dean Fred Roper.

Several major events marked the second half of the college's year-long celebration. In March, Elfreda A. Chatman spoke at the Women's History Month event, and storyteller-in-residence Augusta Baker invited Barbara Bryant to speak at the Black History Month program.

On April 2 and 3, Carol Tenopir and Charles McClure spoke at the college's Information Access Conference, and ALA president Marilyn Miller gave the Seventh Annual Dean's Lecture on April 23. Also in April the college sponsored its Anniversary Extravaganza, which included a homecoming celebration; a visual history program prepared by LISSA, the student organization; the Beta Phi Mu chapter initiation; and a dance in the university's Russell House Ballroom.

In May, Augusta Baker was again featured in "(Augusta) Baker's Dozen: A Celebration of Stories," a two-day event on May 7 and 8. Pat Feehan was in charge of planning. ■■

Letters

To the Editor:

In response to the very interesting paper by Scott Seaman (*ITAL,* September 1992, p. 229–35) on unfilled OCLC requests, I should like to draw readers' attention to some of the categories identified by David King of ILLINET in 1987 as "accounting for an usually large portion of failed requests."

The categories included conference proceedings, technical reports, and microforms. The British Library Document Supply Centre has very large stocks of such material: over 300,000 conferences, to which around 15,000 new items are added each year; over 3 1/2 million reports in microform, plus many more in hard copy and over 1,700 miles (2,700 km) of roll microfilm.

All of this material is available to registered DSC customers via the OCLC system, although at present, the only DSC holdings listed on CCLC are around 30,000 of our current intake of 53,000 serials. These form only part of our total holdings of over 228,000 serial titles.

Other "hard to find" materials with which we should be able to help are UK Official Publications, particularly HMSO, and publications of the European Community. We are a depository library for the latter.

Our present fill rate for OCLC requests is 65%–70%, and we hope to improve on this when internal procedures allow us to pass some of the unfilled photocopy requests to other parts of the British Library.

For further information about BLDSC, please see BRI in your OCLC NAD, or contact Richard Walker, Customer Service, +44937546049, fax +44937546333, e-mail BL 14 @ UK.AC.RUTHERFORD.GEC-B.—*Janet Martin, Head, Special Materials Section, The British Library Document Supply Center, Boston Spa, United Kingdom.*

To the Editor:

This might be an idea for future topics in the Journal [*Information Technology and Libraries*]. The article was in the Raleigh, N.C. newspaper. It is called "The Raleigh News and Observer." I decided to join the Information Technology Section after reading the article.—*Nancy Sosnik, American Library Association member, Raleigh, North Carolina.*