A REVIEW OF CURRENT TECHNIQUES IN INFORMATION TECHNOLOGY

Sandra C. Dow, Associate Librarian,
Alfred Taubman Medical Library, The University of Michigan, Ann Arbor, Michigan (U.S.A.)

INTRODUCTION

Nearly 7,000 scientific articles are written each day. Every five years, the body of scientific information doubles. Fortunately, exciting new technologies are evolving to help deal with this torrent of information. Being information specialist and not a technology expert, I’m here today to discuss current techniques that are available to meet your information needs.

Information technology exists today that would allow an individual to access all of the latest information in his/her field without ever setting foot in a library. Imagine the following hypothetical scenario:

A graduate student is working on a paper dealing with a new drug treatment for aids that she plans to submit for publication in a journal. She begins by searching 3 sources for bibliographic information. The 3 sources are 1) MEDLINE on CD-ROM player next to the personal computer, 2) The BIOSIS and Chemical Abstracts databases on the Dialog Information System and 3) the local university’s online catalog. From these 3 sources, the student assembles a collection of references on the topic under study. the combined records are downloaded and converted to a database int he workstation using a pre-configured data management software package such as Pro-Cite. Duplicate records are eliminated, and Pro-Cite is used to produce a bibliography for the paper in the format required by the journal to which the paper will be submitted.

From the Pro-Cite database the student selects the documents she wants to examine. this selected set is then sent by modem to a workstation in the library. The resulting documents are physically taken from the shelves and the relevant pages are scanned into the library’s workstation. The images are then sent via FAX modem to the student’s PC. The student stores the document images on the hard disk on her workstation. Using optical character recognition software, the articles are converted into ASCII characters. The student then uses her wordprocessor to begin work on the paper. Quotes form the scanned documents are inserted into the document and citations are appended to the quotations used.

When the paper is nearing completion, a bibliography is generated automatically and appended to the paper. When the paper is completed the student sends it via modem to her professor’s computer where it can be examined. A copy will also be printed using a laser printer.

The technology exists today that will allow everything I have just described to take place. However, I know of no one who is performing at this level of expertise other than on an experimental basis. Various aspects of this scenario, however, are routinely performed daily throughout the world. Lets start at the begging.
ONLINE INFORMATION RETRIEVAL

The evolution of computers combined with communications technology has provided us with the power to expand our capabilities in ways we are just beginning to imagine. It is now possible for anyone with access to a home or office computer to have simple, rapid, and inexpensive access to the literature with minimal training. Two to three thousand dollars will purchase the necessary hardware which includes a microcomputer, keyboard, printer, modem and telecommunications software which provides a connection to mainframe computers.

The next step is to determine which of the over 3,000 available electronic database are likely to contain the information of interest to you. An electronic database is a collection of information in computer-readable form, organized so it can be expanded, updated, and retrieved very rapidly. All related information, or data that completely describes one physical item, such as a journal article, book or person, are treated as a unit and called a record. For example, in a database covering graduate school applicants, all information about one applicant is a record. Each item of information within that records called a field, such as a individual applicant’s name or GRE scores.

Within any database, the records are all structured the same way and contain the same parts or fields to make it possible for a user to access the desired information easily, accurately and quickly. Database records may contain directory-type information, as in the previous example about graduate school applicants, or they may include the full-text of an article, or encyclopedia, or they may contain bibliographic information.

Bibliographic databases describe documents by giving information about them, such as a title, the author’s name, and the publication source. Each record in a bibliography database stands for another full document, such as a journal or book. Output from bibliographic databases is not meant to be an end in itself. Rather, it helps the user locate relevant articles or books which must then be obtained form the library or publisher. Abstracts often help the user decide whether to read the full article.

Some examples of databases that might be of interest to attendees of this conference are MEDLINE, CORK, NTIS, ETOH, Compendex Plus, Lexis, Westlaw, TRIS, Toxline, Drug Information Full Text and so on.

MEDLINE is produced by the National Library of Medicine and indexes over 3,500 journals published around the world on literature relating to all aspects of clinical and experimental medicine as well as allied health fields.

CORK was developed in 1978 by a grant from the KROC Foundation and is intended to serve health care providers in study of alcohol use and abuse.

NTIS (National Technical Information Service) contains abstracts of reports from major government agencies. It is the central source for Government sponsored research, development, and engineering reports, as well as foreign technical reports. major areas of coverage include transportation and Biomedicine.

ETOH, the Alcohol and Alcohol problems Science Database consists of
bibliographic records with abstracts of alcohol related scientific references from U.S. and foreign sources. It covers all aspects of alcoholism research including: psychology, psychiatry, physiology, biochemistry, accidents, safety, legislation and public policy.

The TRIS (Transportation Research Information Service) database is a composite file whose records are either abstracts of documents or resumes of research projects that are relevant to the planning, development, operation and performance of transportation systems and their components. The information in this database is supplied through the cooperation of information services and centers that specialize in transportation information and whose combined holdings cover much of the transportation field, including alcohol and its relation to traffic safety.

Compendex Plus is the single best online source of engineering literature. Produced by Engineering Information, Inc., and corresponding to the printed Engineering Index, Compendex Plus provides authoritative coverage of the international literature of engineering and technology.

TOXLINE, produced by the National Library of Medicine, is a comprehensive resource for the health and environmental effects of chemicals, drugs, and other substances.

Westlaw offers a comprehensive legal library featuring the full text of cases from the federal and appellate courts of the fifty states.

Lexis contains Federal and State court reported cases, Federal and State constitutions, codes, rules and regulations, decisions from certain Government agencies and material related to specific subject matters.

In order to gain access to one of these databases you must obtain a password by registering with the database Vendor. Vendors are companies who buy or lease the tapes containing the electronic database from the producer. The tapes are loaded or mounted on mainframe computers and are searched by individual subscribers who pay a fee for use of the system.

When selecting a vendor, it is important to select the database of greatest interest to you first and then shop for the vendor that offers the majority of the databases needed. In some cases, it may be necessary to contract with two or more vendors in order to gain access to all databases of interest. The second most important consideration is the way the cost is configured. Determine if there is a sign-up fee and/or monthly minimum. These can become excessive when searches are infrequent. Another feature to consider is the amount of training required to master the search technique. The biggest change we are observing in information technology today, after the reduction of costs, is the increasing "friendliness" of the systems. Until now, access to the various information retrieval systems required the special skills of a searcher. Now we see more and more systems like PaperChase, BRS/After Dark and Dialog's Knowledge Index, which have simplified the searching procedures making it possible to conduct a search of the literature with a minimum of training.

In order to gain a competitive edge, vendors have added a vast array of special features to their search systems, such as full text of certain journals and books, electronic bulletin boards which can be used to communicate with
colleagues, and the capability to search hundreds of nonmedical databases.

BRS is a vendor which offers access through its general search service or through 2 user-friendly systems: BRS/After Dark or BRS/Colleague. With the general search service, you pay a sign up fee depending on the type of subscription plan which varies according to the amount of use. There is monthly minimum but you pay 14 cents for each citation printed. There are approximately 150 databases available. BRS/After-Dark provides access to 90 databases during the evening hours after 6 p.m. and on weekends. The sign up fee is $95 and there is $12.00 per month minimum. Each citation costs 3 cents. BRS/Colleague charges $95 to sign up and $20.00 per month minimum. Each citation costs 9 cents and there are approximately 130 databases plus 9\80 full-text medical journals and 25 full text medical books.

Dialog Information Services provides access to approximately 340 databases through its general search service which requires a $25.00 annual fee but no monthly minimum. The user must be fairly sophisticated to use the expert mode software. For less sophisticated searchers, Dialog/Knowledge Index charges $35.00 to sign up and no monthly minimum. There is no daytime searching allowed except on weekends and there is also a limited number of databases (70) from which to choose. Dialog Medical Connection is available for a sign up fee of $95.00. Approximately 22 medical databases are available and may be searched in either user-friendly mode or command mode.

Doing one's own literature searching has obvious benefits in terms of the ability to find precisely what one seeks, once the search technique is mastered. Professionals who do not want to perform their own searches, or who do not have access to the systems described earlier, can call libraries and information centers such as the University of Michigan Transportation Research Institute in Ann Arbor, Michigan and ask an information specialist there to conduct the searches. Depending on the amount of time available for learning a system, users can either rely on "user friendly" or menu driven systems or they can expend their skills and learn to search the expert systems. The goal is to obtain information, not to become a computer expert.

CD-ROM ACCESS TO INFORMATION

There is currently an alternative to using online databases. Compact Disk Read-Only Memory, or DC-ROM, is a powerful resource which combines the best features of both online and print access to the literature, including multiple access points and leisurely browsing. Identical to audio CDs the approximately five-inch discs are coated with a reflective metal layer. Data are burned-on, so to speak, by lasers in a series of lands and pits. These "hills and valleys" of data are then read by lasers in a Cd drive, which translates them into electronic impulses that a computer can process. Unlike floppy disks, CDs are virtually indestructible and unreasonable and they provide massive storage capacity. Translated into concrete terms, one CD can contain over 250,000 pages of single-spaced text f the equivalent of 1,500 floppies.

The hardware needed for accessing a CD-Database is a CD-ROM drive, an interface card and cable, a microcomputer, two floppy drives or one floppy drive and one hard disk, keyboard, monitor and a printer for producing hard copies. DOS (Disk Operating System) software support is also required.
Once a master disk is created, CD-ROM discs are relatively inexpensive to reproduce: in the range of $15.00 per copy. Mastering the discs still costs around $10,000 to $20.00 per disc, but the cost is dropping. The present cost of a CD drive ranges from $1,000 to $1,150, but the price is expected to drop even lower in the near future. Many distributors will loan or lease the drive and CD-interface card needed for a trial period (usually one to three months).

But storage of information is only one facet of this new development. In order to access a CD-ROM disc from your personal computer, software must be created. The software is stored on a diskette (floppy) that directs your PC. This software system permits you to access a huge file via instructions stored on your floppy disk or hard disk. It permits you to work with the data without even dialing up a central computer or vendor. All the information you are using is on the compact disk. Retrieval software to extract the information from the CD database is usually included with the sale of the CD. The retrieval software enables searches to the database to be made using keywords, and boolean logic just as when using an online system.

With CD-ROM installations, once the initial costs for the hardware and software are met, running costs are minimal. They have fixed, budgetable costs and are generally in depending of telecommunications. You can access the data from your PC without a modem, without telecommunications, online connect, or per article print charges, and at your convenience, regardless of mainframe up or down-time. It combines the advantages of multiple access points associated with online searching and the browsing capability of the printed indexes. Since costs are independent of usage, end users are less concerned about the time spend on searches and are more likely to engage in experimentation and self-instruction. CD-ROM disk are also portable. A large database can be carried easily or mailed from one location to another. In addition, the stand-alone CD-ROMs are immune to computer viruses.

Currently, more than 500 titles are on the market or in planning stages. MEDICINE may be one of the most popular databases available on CD-ROM. At the present time there are 8 versions of the product available each with its own search software. On September 23, 1988, the National Library of Medicine presented an "Evaluation Forum: MEDLINE on CD-ROM. The purpose of the forum was to present the evaluation findings of a project involving field tests of seven MEDLINE CD-ROM products at 21 sites throughout the country. One of the findings was that the best CD-ROM product will depend on the setting. What is important is to assess who is going to use the information, what their needs are, and then match those needs with the appropriate product. Some products are easier to use than others. In some systems, for example, search steps can be completed with one keystroke, as in a menu system. There are also complex command-based systems that demand from users knowledge of Boolean logic and very specific commands. Some MEDLINE products have two tiers - both a novice menu mode and an experienced user mode.

There are some problems associated with the use of CD-ROM's, however. Equipment costs may be higher than with other end-user services because CD-ROM players are required and most systems need PCs with hard disks. Many systems have one year of the database per disk, which is inconvenient to users wanting to search retrospectively several years. One user at a time in fixed location has been the rule, although some vendors are now coming out with equipment that allows multiple users. Also, equipment has recently been developed that will

63
require less disc swapping—similar to a multi-disc jukebox. This of course adds to equipment costs. CD-ROM technology is alluring but generally overpriced a this stage of its development. Online promoters see multiple problems with CD-ROM and vice versa.

It is not currently realistic to consider a personal subscription for databases on CD-ROM due to cost. Subscriptions cost vary, ranging from $900 to $10,000 for a single disc. However, this could change. If producers were to change their marketing strategies from selling few institutional subscriptions at high prices to selling many personal subscriptions at low prices, I believe that they could realize the same revenues.

The question of the quality to backup service for the hardware is also of concern. Most of the problems with CD-ROM use are encountered in the installation and the initial use of the CDs. Local CD drive suppliers may not be able to provide satisfactory installation, maintenance and repair service for CD drives since they are new on the market and individuals ma find it prohibitive at present to have private installations. The high initial cost of Cd software may limit it use to institutions that are able to provide the capital cost of several hundred to a few thousand dollars.

How is one to choose between CD-RM and Online? Not all institution or health centers have the facilities needed for online access to information. Specifically some researchers in developing countries may lack the necessary equipment for a mainframe to microcomputer lineup, or have poor telephone connections with waiting time for access to the database. Another problem may be that the database is not operating at the time a search is required.

REPRINT FILE MANAGEMENT

Whether one uses an Online or a CD-ROM system for retrieving information, a collection of reprints usually results. The purpose of collecting reprints is to have a readily available resource of information at hand. In practice, however, prints files frequently become a burden and are neglected. For continued long term use, reprints must be filed, maintained, and retrievable in a single, rapid, and accurate way. The reprint file must be organized before the collection becomes unmanageable. The same inexpensive microcomputers used to retrieve the information from online and CD-ROM databases can also be used with Database Management Software to improve the usefulness and organization of a reprint file. An ideal file management software program should allow maximum flexibility in searching, similar to that used with the online or CD-ROM system. Common computer-bases techniques for searching involve Boolean logic (combining search terms with And, Or, and Not operators) and wildcard-character or truncation search terms—using asterisks or question marks—to find words with similar spelling, prefixes, or suffixes. In addition it should: index all authors' names, index an unlimited number of keywords; print an alphabetical list of keywords, automatically count the number of citations containing each keyword; include fields of variable length rather than fixed length; edit data on-screen; transfer or import the bibliographic data from other databases and have easy-to-follow menus. There are currently several products on the market that will allow one to search an online database, download the retrieved citations to a disk, and then transfer the citations directly into a personal database. Work is underway to provide similar software for users of some of the more popular CD databases as well.
Many of these products will also produce a bibliography for a paper in the format required by the journal to which the paper will be submitted.

CONCLUSION

Computer users who are confounded by the bewildering and seemingly endless advances in both hardware and software may take a little solace in the thought that leaders in the industry often feel the same way. The following paragraph is from a speech delivered at the Software Publishers Association’s spring symposium by Jim Manzi, chief executive officer of the Lotus Development Corp: “We are asking our customers to put up with quantum jumps in release technology that render their prior database unteachable, their prior word processing files incompatible, their prior investments in knowledge work, staff training and hardware unrecoverable. And then to add insult to injury, we give them documentation which is not more meant for the average end user then science fiction is for scientists and ghost stories are for ghosts.”

Buying a computer is becoming more and more like buying a new car in August: You know there’s going to be a new model out soon, and it’ll probably be niftier. When do you do? Buy or Wait? The important thing to remember about the microcomputer is that it is not obsolete just because there’s suddenly something better. Even without upgrading, a computer can give you years of service long after it’s been superceded by newer, faster cheaper ones. Why wait? The information is there waiting to be retrieved. The technology, although not perfect is waiting to be used. The time to get started is now.

A Review of Current Techniques in Information Technology

Access to information has been remarkably enhanced by the rapid advances in computer technology. This paper will review techniques that the scientist can use with his or her personal computer that only a few years ago could only be done by a professional librarian. For example, today a scientist can run his own database search of the literature on Alcohol, Drugs and Traffic Safety using a variety of online and CD-ROM (Compact Disc - Read Only Memory) systems. The available resources will be reviewed, giving the advantages and limitations of each. A description of how the personal computer can be used with Database Management Software programs to organize and maintain one’s own personal file will be given. These programs permit unlimited searching capabilities and the ability to organize the materials to suit one’s individual needs.