

THE NEW INTERACTIVE SEARCH SYSTEMS

There has been good user acceptance of information retrieval systems in the past five years, and the main reason seems to be the new interactive search systems. In theory (and, in some cases, in practice), the end user interacts directly with large bibliographic files and can get selected citations and abstracts in a few minutes time, with these systems. It is more generally the case that a trained search specialist operates the terminal, but even then the end user gets a very rapid response. The technology now supports rapid searching of the bibliographic files just when social and business pressures are building up to require it. Here are the experiences of some pioneer users.

The Library of Congress, located in Washington, D.C. adjacent to the U.S. Capitol Building, serves many of the information needs of the U.S. Congress. In addition, the Library has other responsibilities, such as operating the U.S. Copyright Office.

The Congressional Research Service (CRS) is one of the six departments of the Library of Congress; it serves special information needs of the Congress. CRS has seven subject matter divisions, plus a reference division and a bibliographic division. In recent years, CRS has made increasing use of computer-based interactive information systems for quickly locating citations to desired literature. Mechanized searching is important to CRS because it is receiving some 200,000 requests per year for searches, ranging from simple to very detailed, complex requests.

CRS uses a search system developed in-house by the Library (SCORPIO), for accessing three important information files, and it also subscribes to three outside interactive services.

In-house system. The Library's SCORPIO system was put into operational on-line use in 1974. It is

operated on the Library's computer system. It accesses three information files on an interactive basis. One of these is the Legislative Information file, which contains the history and current status of all public bills and resolutions of the 93rd and (current) 94th Congresses. The data in this file is current within 48 hours, or less, of the events pertaining to these bills and resolutions. The second information file is the Issues Briefing System, dealing with current public issues such as air pollution, energy, and so on. The file contains both brief and detailed definitions of each issue, references to pending legislation, chronology, and sources of additional information. The third interactive file is the Bibliographic file, with citations to all of CRS' own reports and memoranda, and citations to significant material in some 3,000 periodicals, United Nations documents, and other such material.

Outside services. CRS subscribes to the New York Times Information Bank service, the National Library of Medicine's MEDLINE service, and the Department of Justice's JURIS service. We will discuss the NY Times Info Bank service later

Reproduction prohibited; copying or photocopying this report is a violation of the copyright law; orders for copies filled promptly; prices listed on last page.

in this report. MEDLINE includes citations to biomedical literature and JURIS contains the U.S. Code, some U.S. Attorney General briefs, Supreme Court cases, and so on.

How does CRS use these services? For example, a senator or a congressman, or one of their staff members, might ask for a list of all legislation that is under discussion on a given subject—abortion, foreign aid, control of energy usage, or whatever. From this list of pending legislation, the requestor might want to know the current status and location of some of the bills, as well as abstracts of supporting literature such as news stories, articles, and editorials. In performing such a search, the CRS specialist would probably use both the in-house and the outside services.

Another not-unusual request is for a senator or congressman to ask for a list and the current status of all bills that he or she has either sponsored or co-sponsored.

A more complex type of request arises in research studies to support the drafting of new legislative bills. The request might ask for a listing of all proposed legislation on the subject that has been submitted in the past X years, how far each bill went through the legislative process, how the final House and Senate bills compared for legislation that was passed, and so on.

Of the requests that can be handled by queries on CRT terminals, CRS finds that they can handle about 70% on a “same day” response basis. The other 30% of these requests involve a longer period of time. Other requests may require the tracking of a particular subject in the public literature over a period of weeks or even months.

The network of SCORPIO terminals extends throughout the seven subject matter divisions and in two Congressional reference centers—one in the Senate and one in the House. In addition, terminals are located in the offices of some 30 Senators and a lesser number of Representatives. About 500 people have been trained in the use of the SCORPIO system.

It is not unusual, we were told, for staff lawyers to operate the terminals and to search the legislative information file and the bibliographic file. Also, some staff members use the services to help them in answering letters from constituents. In this use, they take the letters right to the terminals, enter the queries, have the answers printed out on the terminals, tear off the responses and

clip them to the letters, then take the letters back to their offices to draft the replies.

One senator became a believer in the new services when an influential and irate constituent called his office and asked why the senator was sponsoring a bill favoring abortion, which she very much opposed. A staff member had ready access to a terminal and made an immediate search of the legislative information file—and within a matter of seconds could tell the constituent that her information was incorrect; the senator was neither a sponsor nor a co-sponsor of any bill dealing with abortion.

So far, the several interactive systems use different types of terminals, different languages, and different procedures. The people at CRS would very much like to see more standardization in these areas. They would prefer one terminal type, which would have a CRT for speed plus a selective hardcopy printout capability. Common languages and procedures would be desirable, since it is confusing to have to change either or both in the middle of a search. A high degree of system reliability is a “must,” they say, including not only being operational when the system is needed but also providing supervisory help when it is needed. Finally, the CRS people say that some means of making the cited source documents available rapidly is a real need; this might involve having the documents in microfilm or microfiche form.

The U.S. Library of Congress has been one of the pioneers in the use of interactive search services, for supporting the information needs of Congress.

Atlantic Richfield Company

The Atlantic Richfield Company (ARCO), with headquarters in Los Angeles, California, is a major petroleum and chemical company. It has annual sales in the order of \$7.4 billion and employs about 26,000 people.

In early 1974, the library at the ARCO headquarters signed up for the SDC Search Service, offered by System Development Corporation in Santa Monica, California. (They also signed up for the Lockheed DIALOG service and have recently installed the NY Times Info Bank service.) Previous to 1974, all bibliographic searches were done manually, using the reference sources

available in the library plus those at nearby public libraries.

The library serves people throughout ARCO, including the 3000 people at headquarters, as well as executives, managers, engineers, and so on at the company's other West Coast operations. These operations include two refineries. The search services are often used to aid these people in making planning-type decisions. And planning decisions in energy companies today can involve a wide variety of disciplines—geology, marketing, finance, environmental impact, government regulations, etc. So the subject matter of the bibliographic searches can involve many types of information files.

To illustrate the use of interactive searching, consider the following example. It is not an unusual case, we were told, but the necessary speed of response probably was more urgent than is true for most requests. An engineer at one of the ARCO refineries called in. He had witnessed an undesired and unfamiliar reaction in a catalytic converter at the refinery and he wanted information on what to do about it—and as quickly as possible. Over the phone, he told the library search specialist the key words applying to the information he wanted.

The search specialist formulated the search query while the engineer was on the phone. Using a key word list, the search specialist made sure that he had all of the terms that might be appropriate. After finishing the phone call, he decided that he would use the SDC Search Service. He placed a telephone call to this service (a local call, in this instance, but the service is available on a time sharing network, as will be discussed) and connected the terminal via an acoustic coupler. He chose the American Petroleum Institute bibliographic file as the first one to use; it was one of several that SDC offered and that were appropriate to this search.

The search was commenced with the fewest but most appropriate terms for the problem at hand. The search specialist entered the terms, connecting them by AND, OR, and NOT as required. The system responded very quickly, giving the number of "hits"—the number of citations that met the search criteria. (If the number seems unduly large, the search specialist can enter additional terms suggested by the requestor, so as to narrow the search.) The next step was to look at

the abstracts by way of the terminal. Those that had any relevance at all were printed out on the hardcopy printer.

The same search process was continued with other appropriate bibliographic files available through the SDC system. These included chemistry abstracts and engineering abstracts. (Some searches can involve 16 or so bibliographic files, we were told; even so, not all literature is yet covered by the interactive search services, so some manual searching may also be required.)

When the appropriate abstracts had been printed out, they were transmitted to the engineer by facsimile transmission, due to the urgency of the case. The engineer reviewed the abstracts and selected some where he wanted to see the source documents. Those that he did not have local access to he ordered from the library. Total elapsed time, from the time the engineer first called in until he had identified the source documents he wanted, was less than one hour.

The people at ARCO say that the procurements of the source documents is becoming a full-time job for one person. The search services provide citations to so much of the world's literature that it is impractical for ARCO to have it all on hand. If a desired source document is not on hand, they next try other nearby libraries, including several university libraries. If not available there, they keep on searching further afield.

In less urgent cases, the ARCO search specialist does not have the abstracts printed out on the hardcopy printer. Rather, he requests that they be printed out on the SDC high speed printer and then mailed to ARCO. (For more distant customers, air mail is used.)

Also, where it is feasible, the search specialist asks the requestor to participate in the search by sitting beside the terminal and reviewing the retrieved abstracts. While some of the employees can and do operate the terminal, ARCO finds that it is usually preferable for the search specialist to conduct the search, with the requestor viewing the results.

In the future, they hope to install "slave" terminals at a number of ARCO locations. A requestor would sit at a slave terminal, and would see the results, while the search itself was being conducted by the search specialist.

What have been the results of this new searching capability, we asked? "Our credibility as an

information resource within ARCO is now much improved," we were told. "We now have access to so many information files that we can give a much improved service over what we used to do manually. We are encouraging ARCO employees to use this new service, as a way of helping them do their jobs better."

Los Angeles County Superintendent of Schools Office

The Los Angeles County Superintendent of Schools office, located in Downey, California, serves 93 school and community college districts in Los Angeles County.

In September 1973, as a service for school district administrators, counselors, and teachers, an education resource service (LANCERS) was started, using an interactive search service offered by a local university. The information file in which they were primarily interested was the ERIC file—the Educational Resources Information Center file.

In September 1974, they also signed up with the Lockheed DIALOG service, offered by the Lockheed Missiles & Space Company, Inc., Palo Alto, California. DIALOG offers access to 20 bibliographic files, including the ERIC file. So the LANCERS service currently has access to the ERIC file through either system. Connection to DIALOG is provided either via Tymshare Inc.'s Tymnet (which we discussed in our January 1973 issue) or by direct dial telephone call; connection via Tymnet is less expensive and is generally used. For their terminal, they use a CRT with attached 30 character per second hardcopy printer.

Requests come into the LANCERS office, generally in written form, from any of the 93 districts. School administrators might ask for citations to specific pieces of legislation or court decisions dealing with, say, school district financing. Or they might ask for citations to literature on specific aspects of staff development. Counselors might ask for citations to literature on ways to change attitudes of parents toward innovative educational methods. And teachers might ask for citations to literature dealing with ways of teaching sixth grade arithmetic to disadvantaged children. These, of course, are only examples of the requests; the actual searches involve a wide range of subjects.

If at all possible, the search specialists at the LANCERS office prefer that the requestor be on

hand at the library when the search is made. The requestor can see the results and can guide the search into narrower or different channels. But since the 93 districts are spread over a wide geographic area, it generally is not feasible for the requestor to be present. Hence the use of written requests.

As mentioned, when a search is to be made, DIALOG is accessed via Tymnet, which in turn is accessed by calling a local telephone number. The search specialist then enters one key word at a time, from the key words submitted by the requestor. (Some information files use controlled vocabularies, so the search specialist must look up the acceptable key words in a thesaurus that is provided.) DIALOG provides a hierarchical breakdown of search terms, so the specialist can ask for related terms, more detailed terms, etc. As each key word is entered, DIALOG responds by telling the number of hits in the file for that key word. When all key words have been entered and the search specialist has a feeling for what is in the file, she enters the boolean search criteria—the desired key words connected by AND, OR, or NOT. When the boolean search criteria have been entered, DIALOG responds with the number of citations that meet the criteria.

The search specialist can then either narrow the search, if the number of hits is too large, or can begin looking at titles and, if desired, abstracts. If the abstracts look appropriate for the question at hand, and if there are more than just a few hits (say, more than five or so), she requests that they be printed on the Lockheed high speed printer and the printout mailed to the requestor.

To expand the use of this new information service, the Superintendent's office currently is offering it free of charge to the 93 districts. However, each request must be approved by a school principal or school district administrator. The use of the service has been accelerating in recent months, we were told.

Union Research Center

The Union Research Center, located in Brea, California, is a part of the Union Oil Company of California. There are about 300 professionals (engineers, researchers, etc.) employed at the center.

In the late 1960s, the people at the Research Center wanted an interactive search capability for searching bibliographic data. The then-

existing batch systems just did not meet their needs; the slow response time was discouraging the Research Center's staff from using the service. And, at that time, no interactive search service was available commercially. So they decided to develop their own system.

In 1970, their in-house search system became operational and they have continued to use it heavily since that time. In addition, they recently subscribed to the SDC Search Service to supplement the in-house system. The bibliographic file used with the in-house system contains citations to petroleum literature, in the main. It includes the petroleum exploration and production abstracts prepared by the University of Tulsa and the American Petroleum Institute file on refinery literature. Also, all company research reports are included. In all, over one million documents are cited in this bibliographic file.

The system uses one CRT terminal, located in the center's library. The system is operational during all regular business hours.

In a typical case, the requestor comes to the library and describes his or her request to the search specialist. The search specialist formulates the search query using a controlled vocabulary, and enters the query on the terminal. Since the library is only a few hundred feet from the computer and since the terminal is hard wired to the computer, transmission is high speed. The information is displayed instantaneously, as far as the human eye is concerned. The system displays the number of hits in the file. Because of the high speed of the system, the requestor (sitting beside the search specialist) might choose to scan through all of, say, 50 or even 100 abstracts—which the people at Union Research Center feel could be tiresome and costly in slower speed systems. When a printout of abstracts is desired, the search specialist so indicates to the system. They are printed out on the high speed printer in the computer room, and are available by the time the requestor has walked from the library to the computer room.

So, whether the abstracts are scanned via the terminal or via a printout, it may well be just a matter of minutes from the time the search began until the requestor has indicated the source documents he or she desires. And, as far as the in-house system is concerned, in most cases those source documents are located in the library.

The Research Center does all it can to encourage its staff members to make literature searches "right now." If a need arises, make the search, they say. True, some searches will prove to be unnecessary—but those are more than compensated for by the ones that uncover valuable information.

So Union Research Center is one of the early users of an interactive search system, one that they created in-house. Moreover, its features, including its speed of response, are such that it compares favorably with the commercial systems now available. But the center is subscribing to an outside service in order to have access to bibliographic files that it would be uneconomic to put on their in-house system.

Where things stand today

Our January 1970 report was concerned with the status of information retrieval systems. At that time, we reported that we found a number of batch-type systems in existence—but not in heavy use. The long delays in getting back a list of citations—delays of days and even weeks—seemed to discourage users.

As indicated earlier in this report, interactive search systems began to make their appearance shortly after that report was written. These systems have caught on very well to the point where some of today's services are economically viable—which certainly did not seem to be true six years ago. Batch search systems are still in use but we believe that they are losing ground rapidly to the interactive systems.

In this report, we will concentrate on the current status of the interactive search systems, operating in business-type and management-type environments. We will not discuss batch-type systems. Also, we will not discuss research that is under way to improve the efficiency of indexing, of abstracting, of searching, etc. These are important topics but they are not crucial to the expanded use of today's systems. As we say, we will concentrate on discussing how some of the interactive search systems are actually being used today.

Digressing for a moment, the reader who is familiar with the field of information retrieval might wonder why we have used the term "bibliographic files" instead of the more common term "data bases." Within the information re-

trieval field, just about all of the bibliographic files are referred to as data bases. But we have been using the term "data bases" for some years now in another context and prefer to remain consistent in that usage. Also, "bibliographic files" or "information files," while longer terms, are more properly descriptive of what is in these files.

The experiences of the above-discussed organizations should indicate that the technology today will support the on-line searching of massive information files—files with millions of citations or similar information records. And, as it happens, there are social and business pressures building up which will encourage the use of such searching technology.

Just what are these social and business pressures? They are changes in attitude on the part of the general public and on the part of governmental regulatory bodies. These changes in attitude result in the need for organizations to better "prepare their cases" when they plan to do something that impacts the general public. In "preparing their cases," such organizations will find that they must search the literature quite thoroughly, to learn of the experiences of other organizations that have tried to do the same things that are being proposed. Here are some examples that we came across during our study, on the need for such searches:

PRESSURES FOR BIBLIOGRAPHIC SEARCHES

1. Environmental impact of proposed actions—finding examples of what other organizations have done and the problems they encountered.
2. Consumer protection activities—involves (1) searching out problems ahead of time, based on reports in the literature and in the press and then taking steps to avoid the problems, and (2) preparing rebuttals when attacks are made on products.
3. Class action law suits—again, searching out and identifying problems ahead of time, plus preparing a case should a suit be entered.
4. Corporate responsibility to the community in which it is located—finding out what companies do to help improve the quality of life in their communities.
5. Companies considering entering new lines of business—finding out what other companies are in the same business and some of the problems they have encountered, as reported in the literature and in the press.

The use of interactive search systems is growing rapidly. Engineering and scientific users probably predominate today. But business-type use

and management-type use are growing. We suspect that the basic history of the computer field will be repeated for interactive searching. The early computer users were mainly in engineering and scientific disciplines, but before too many years passed, business uses dominated from the volume standpoint.

With this background, let us give a brief discussion of three of the leading commercial interactive search services, presented in alphabetical order.

Lockheed's DIALOG service

The DIALOG on-line interactive information retrieval service is offered by the Lockheed Missiles & Space Co., Inc., Palo Alto, California, which is part of the Lockheed Aircraft Corporation. Lockheed first began providing on-line information retrieval services for the U.S. government agencies in 1965. DIALOG was developed in the mid 1960s, and in 1970 the services were offered commercially to industrial and academic users. DIALOG continues to expand its services.

DIALOG currently offers customers access to 20 bibliographic files. These files are obtained from the organizations which do the abstracting, indexing, and creating of the files. The files cover the fields of science, social science, business, and agriculture, including in particular chemical abstracts, physics abstracts, psychological abstracts, U.S. government report abstracts, market abstracts, abstracted business information, and so on. In all, DIALOG provides on-line access to over six million citations.

The DIALOG service provides on-line access to the bibliographic files at all times during a 60-hour week. In all files, records can be accessed by keywords in titles (adjacency or proximity of keywords can be specified), author, year of announcement and/or publication, corporate source, index terms or subject headings, journal name or publication type, and several other means. Commands are simple, English language instructions which provide for the display of the file index, selection and combination of terms, display of results, and off-line printing of results.

Several speeds of service are available. The Lockheed high speed data network operates at 480 characters per second, so that users who can access this network can operate at this speed. Direct dial phone access, including Wide Area Tele-

phone Service (WATS), can provide service at 120 characters per second. Access via Tymshare's Tymnet gives from 10 to 30 characters per second service. And Teletype TWX service operates at from 8 to 10 characters per second. These speeds are output speeds, of course; input speed is at the terminal operator's typing speed.

DIALOG can operate with a wide variety of terminal types; the main requirement is that the terminal accept the ASCII code. For high-speed use (120 and 480 characters per second), a CRT terminal is needed; this is usually complemented with a typewriter-type hardcopy printer for printing selected information. For the medium speeds (15 to 30 cps), either just a typewriter-type terminal or CRT-plus-typewriter combination can be used. And at the low speeds (8 to 10 cps), teleprinters are used. The most common terminal types in use with DIALOG are the typewriter-type, operating at the 15 to 30 cps speed.

Typically, users use the hardcopy printer for printing out only selected information—search criteria, number of hits, titles of articles, and such. If there is any volume of abstract printing to be done, users generally request that the abstracts be printed on the Lockheed high-speed printer in Palo Alto and sent to them via air mail.

An EXPAND command provides the user with a display of the *alphabetically-near* index terms to the term that has been entered, along with the number of citations for each term. EXPAND also can provide the user with a display of *related* terms, in the same format. So the EXPAND command allows the user to find the indexed key words that are most appropriate to his search. A SELECT command allows the user to specify which of the key words are to be used in a search. A COMBINE command allows for the boolean combination of the selected terms.

Prices for the DIALOG service include the terminal connect time (usually \$25 to \$65 per hour, depending upon the bibliographic file being used), plus data communications charges (for example, \$10 per hour for domestic Tymnet use, and \$22 per hour for trans-Atlantic use), plus off-line printing charges (10¢ per abstract, including air mail postage). In addition, the user must provide the terminal and modem or acoustic coupler. There are no installation or monthly minimum charges.

The off-line printing is done in both upper and

lower case, and is easily readable. The content of the abstracts depends upon the bibliographic file supplier, of course, but the samples that we saw were very complete. Not only did the abstracts give a reasonable idea of the contents of the source document but also full information was given on where to procure the source document, including the prices of photocopies or microfiche copies of reports.

New York Times Information Bank

The New York Times Information Bank service was put into commercial service in 1973 by the New York Times Company, New York City. The major type of information in the file is selected material from the *New York Times*, from January 1, 1969, to the present. But in addition, the file includes citations and abstracts from some 60 other U.S. and foreign newspapers and magazines; the time period for these generally begins in 1972 and 1973. The file now has almost one million citations and abstracts, about evenly divided between the *NY Times* and other publications, and they are being added at the rate of about 20,000 per month.

Info Bank uses a controlled vocabulary and each user installation is provided with a thesaurus. Search requests are formulated by looking up the appropriate terms in the thesaurus. In addition, however, Info Bank provides a cross-reference to related terms to aid the user in finding the appropriate terms for his search. All indexing and abstracting of the *NY Times* material is done by the *NY Times* staff.

The Info Bank service also provides microfiche copies of all *NY Times* source material. If and when arrangements can be made with the publishers of the other periodicals, microfiche copies of this other source material will also be provided.

The bibliographic files contain significant news items, business and financial news, interpretive articles, editorials, surveys, and even advertising matter that is related to public interest topics (*not* product related). The material, in order to be included in Info Bank, must be of broad, general interest; material of local or narrow group interest is not included.

The 60 other publications (in addition to the *NY Times*) include such periodicals as *Advertising Age*, *Black Scholar*, *Business Week*, *Forbes*, *Fortune*, *Economist of London*, *Psychology Today*,

Reader's Digest, *Time*, *Wall Street Journal*, and *Women's Wear Daily*. So there is a wide coverage of sources dealing with issues of current interest.

The basic speed of service is 1200, 2000, or 2400 bits per second (essentially 120, 200, or 240 characters per second). To date, Datapoint 1100 and 2200, Incoterm SPD 10/20, Hazeltine 2000, IBM 3270 and 3275, and Bell Canada VuCom CRT terminals can be used. The people at Info Bank recommend the use of an attached hardcopy printer which operates in the 30 to 120 characters per second speed range. At the data transmission speeds that are used, access may be either by private line or by direct dial phone line.

Because of the relatively high speed of transmission, the user's train of thought is not distracted by the system. New lines of information appear very rapidly on the CRT screen, so the user is seldom waiting on the system.

The people at Info Bank urge that users provide a microfiche reader-printer and subscribe to their source document microfiche service. Six years of *NY Times* material requires only about two square feet of shelf space. When a user learns enough from an abstract to indicate that he would like to see the original document, it generally takes a minute or so to find the proper microfiche, locate the document on it, and get a print made.

The Info Bank service also provides for the off-line printing of abstracts on their high-speed printer; the abstracts are then mailed to the user. But in the more typical use, users bypass this step. They scan the abstracts on the CRT—and, in fact, may get all the information they need from the abstracts. If not, the source documents may well be available in the microfiche file.

The system is used in-house by *NY Times* editors, reporters, and so on. A number of terminals are available to newsroom personnel and they are encouraged to use it. The typical use here, we are told, is to ask for the selected abstracts to be printed off-line and delivered to the newsroom within two hours. Background material for news reports, articles, and editorials can be obtained in this manner.

Prices for the Info Bank include terminal connection time (in the order of \$45 to \$55 per hour, depending upon speed of operation and number of hours per month used) and the microfiche service (\$270 per year for all back copies, if ordered at time of original contract, or \$540 per year if or-

dered more than 90 days later, plus \$45 per month for updating). User costs would also include the terminal cost, data communications charges, and microfiche reader-printer costs. Off-line printing prices are 1¢ per line. There is a monthly minimum of four hours of service.

The SDC Search Service

The SDC Search Service, offered by System Development Corporation, Santa Monica, California, began commercial operation in January 1973. It has grown at a steady rate since that time. At present, its main customer base is scientists, engineers, and researchers. But the list of non-technical users is also growing; these include banks, insurance companies, and managers in companies that use the technical files.

The SDC Search Service offers access to 19 bibliographic files; 12 of these are offered exclusively through SDC. Technical files include the subjects of petroleum, chemistry, engineering, the geosciences, the life sciences, and a file on research in progress. Other subject areas include education, business management, business news in petroleum and energy, and agriculture.

In the business literature file, six means of access are provided: by title, by author, by publication year, by abstract (including key words), by journal code, and by update code. The directly searchable subject index terms (key words) are derived from the abstract and title categories.

A variety of terminal types can be used with this service. The basic requirement is that the terminal accept ASCII and EBCDIC code. On direct dial phone lines, speeds up to 1200 bits per second (essentially 120 characters per second) can be provided. Access is also available over Tymshare's Tymnet, operating at up to 30 characters per second. The bulk of the terminals in use with this service, we were told, are typewriter-type terminals operating in the 10 to 30 cps speed range.

In most instances, users do only selective printing on their terminals—search criteria, number of hits, and perhaps titles and a few abstracts. If there are a good number of abstracts to be printed, users generally request that these be printed off-line on the SDC high speed printer and then sent to the user via air mail.

Special features of the SDC Search Service include the ability to enter multiple terms in one search statement, the full truncation of word

stems, and the ability to serially scan any part of a set of retrieved records for any specified string of characters.

Prices for the SDC Search Service include terminal connection time (ranging from \$40 to \$120 per hour, depending upon the bibliographic file used) plus data communication charges (\$11 per hour for domestic users of Tymnet and \$22 per hour for trans-Atlantic use) plus off-line print charges (8¢ to 25¢ per citation, depending upon the bibliographic file used). The user must provide the terminal and modem (or acoustic coupler) equipment. There are no installation charges or monthly minimum charges for this service; charges are for usage only.

Bibliographic files

We have made reference to the fact that frequently the interactive search services obtain their bibliographic files from other sources—namely, from organizations that supply such files on a fee basis. In the cases discussed above, only the Union Research Center and the New York Times Info Bank service are creating their own bibliographic files, and in Union's case, it is only their own internal research reports that they are inputting.

The availability of good quality bibliographic files is one of the essentials for a good search service, of course. We will have more to say shortly about some of the reported quality differences among these files. But more and more files are becoming available, and as they do, the value of the search services increases.

In the brief tests that we made of some of these files, we were impressed with the breadth and depth of coverage of subjects. With all of the literature searching we do, we still were able to find citations to some literature that looked most interesting to us and that we had not seen referenced before.

Individual user organizations may obtain many of these bibliographic files directly from the suppliers. We were told that some organizations not only subscribe to one or more of the commercial interactive search services but also obtain some of these bibliographic files for in-house searching. The reason would seem to be an economic one; if one bibliographic file is to be searched very frequently, it might be less expensive to do it in-house. The commercial search services would

then be used for the files that are searched less frequently.

The U.S. National Bureau of Standards is keeping abreast of the development of information systems and has issued a number of reports on this subject. In Reference 4, a description is given of a good many of the available bibliographic files, along with price information and the person to contact. Most of the files listed are scientific and engineering oriented. The reference also lists a number of the available search services and current awareness services.

Using interactive search services

Some of today's commercial search services charge only on a usage basis and have no installation charges and no monthly minimum charges. They thus provide a relatively inexpensive way for a using organization to try out this new technology. A complete search often can be conducted in 10 to 15 minutes and the total charges, including the off-line printing of a number of abstracts, might be in the order of \$15 to \$25.

We were told of one example where a portable terminal was taken to one prospective user organization. Access was made to a search service on a demonstration basis. The user decided to make a search on a subject that two people had been making a search on manually. One of these people had been working essentially full time for the better part of six weeks on it, and the other person—a consultant—had been working part time on it over this same period. The mechanized search took 15 minutes. The evaluation of the results by the user: the manual search was *almost* as complete as the mechanized one! This was perhaps an extreme case but it does point up the economic benefits of interactive search services.

There still is an important role for in-house search facilities. For one thing, an in-house system may have access to literature that a public service does not have, such as company-confidential research reports. Then, too, there can be economic advantages for in-house searching of frequently used files. If the terminal and the computer are at the same site, there can also be speed advantages for getting printouts of abstracts, as well as higher speed of transmission from computer to terminal for faster searching. The National Bureau of Standards has developed a list

(Reference 5) of software packages that can be installed and used for interactive searching of bibliographic files at the user's site, as well as those available as a service running at the vendor's site. It should be pointed out that the packages will not provide all of the conveniences and services that the commercial services provide, but they do provide some basic capabilities.

We suspect that speeds of terminal operation will be increased on the commercial services. We have heard of one study of search services that concluded that a speed of ten characters per second was just too slow; a user's mind would wander while the output was being printed. However, it was concluded that a speed of 30 characters per second was sufficiently fast that the delay involved would not break a user's train of thought. But we have to admit from what we have seen: a speed in the range of 100 to 500 characters per second is nice. When you call for the next abstract, zip, there it is; there is no distraction from seeing the text appear one character at a time.

These higher speeds of operation imply the use of CRT terminals although, of course, there are printing terminals that operate in the 100 cps range. But the speed and quietness of the CRT terminals are real advantages. There still is need, though, for selective printing. If the user has only a CRT terminal, he may need to manually copy down his search criteria, the number of hits, and perhaps some of the citations and portions of abstracts. A history of the search is important, we were told; this is one of the advantages of the typewriter-type terminals in that the printed copy provides a history of what has gone on. So, with CRT terminals, some facility for providing selective hardcopy probably is needed.

Typically, we were told, there will need to be a trained search specialist conducting the search—but it is very desirable for the requestor to be on hand as the search is conducted. For one thing, the occasional user might forget how to use a service properly, or might not search all appropriate files, or such. Also, the sad fact is that a lot of people do not like to read—so they might skim over titles and abstracts and thus miss points that they should see. A trained search specialist hopefully would be a good reader and could point out such points to the requestor. The concept of "slave" terminals is an interesting one; it would allow the search to be conducted by a search spe-

cialist but with the requestor(s) participating at one or more slave terminals. There is no real reason, of course, why these interactive search services cannot be used directly by the requestors—and we cited examples of such at the U.S. Congress and at the New York Times; also see Reference 6. But from a number of the people that we interviewed, we received the same message: don't count on eliminating the trained search specialist from the picture.

Another important feature of interactive search services is the supervisory "help" service. If a user, even a trained search specialist, cannot get the desired response from the system or is having trouble with some aspect of a search, he or she should be able to call for help. For domestic users, a long-distance phone call is the typical means used. The supervisor at the service cuts in to the user's line and sees the same message on his terminal that the user sees. The user explains over the phone what he or she is trying to do—and the supervisor helps solve the problem. This service should be available at all hours that the service is operating. For trans-Atlantic service, or elsewhere in the world, the user can type a message to the supervisor on the terminal, and communications can take place via the terminals.

Another feature to look for in interactive search services, we were told, is multiple access paths to the information. Citations should be retrievable by way of key words, titles, author's name, journal name, and so on.

But all is not milk-and-honey with today's interactive search services. Here are some of the problems we encountered in our interviews.

Problems with search services

The problem area that seemed to be mentioned most frequently to us was that of variations in quality and consistency of the bibliographic files. There are big variations in the quality of indexing and of abstracting among the various files. Poor assignment of key words and/or poor abstracting can give false hits, thus wasting time and causing annoyance on the part of the user. False hits will never be completely eliminated, but hopefully they should not be caused by poor indexing. Also, inconsistent punctuation can lead to missed citations. For instance, the term *time sharing* might be entered as a key term in several variations in a bibliographic file with an uncontrolled vocabu-

lary: "timesharing," "time sharing," "time-sharing," etc. The collation sequence of the computer might space these terms quite far apart and the user might see one and not the others.

Another common complaint is the differences in procedures for using the several bibliographic files within one search service. For instance, one file might use a controlled vocabulary where the searcher must first reference a thesaurus before entering the key words, while the next file searched has an uncontrolled vocabulary.

Then, too, there are differences in semantics among the various bibliographic files—two terms used that mean the same thing, or two or more meanings for the same term. For instance, one file might have some entries under the key word *computer* while in another file those same entries would be found under *data processing*.

Also mentioned frequently to us were the differences in procedures among the different search services. When a user uses two or more search services, he or she may tend to use one system's commands and procedures on another system. There is user interest in a set of standard procedures, based on the "best" of those now available. But progress toward this end probably will be slow. For one thing, each commercial search service will most likely want to exploit the advantages it believes it has, in the marketplace.

Still another point mentioned was the differences in terminal requirements. In some instances, services require the use of specific types of terminals. It would be desirable for users to be able to access any service to which they subscribe from a single terminal type.

The main point one can make from the above complaints is that users desire standardization and consistency—within bibliographic files, among bibliographic files, within a search service, and among search services. The workings of the marketplace may well lead to this, to keep up with the leaders, competitors may find that they have to provide services that are compatible with the leaders.

But there are still more aspects where users would like to see improvements. There are significant differences in the support that the various services provide, we were told. This support includes: (a) hours of the day and days of the week that each bibliographic file is available for searching; (b) the quality and availability of supervisory

"help" services; (c) the reliability of the equipment used which translates into the reliability of getting service when it is desired; (d) the quality of the instruction manuals provided, which describe what actions to take to get desired responses; (e) the speed of response of the system, which may be affected by the computer being shared with other workload; (f) providing multiple means of access to the citations, plus system-assistance in leading the user to the means of access he needs; and (g) providing both upper and lower case alphabets for printouts, for ease of reading. Some services do very well on these points, we were told, while others have shortcomings.

In some instances, search system designers and bibliographic file indexers have used pedantic terminology—"initiate," instead of "start" or "begin"; or "educational institution" instead of "school." The terminology should be selected for the market that is served, it was pointed out to us.

Some of the problems have little or nothing to do with the search services themselves. One problem that applies to all search services, including in-house ones, relates to human reading capabilities. Enough people do not like to read that it causes a problem. They resist using a service that forces them to read, so they do not perform a bibliographic search—and they may end up trying to resolve problems that have already been solved by others. Then, if they do use a search service, they may not read the abstracts carefully enough (perhaps missing points made obliquely) and hence miss valuable information.

Another very basic problem is that of gaining access to source documents once they have been identified by a search. It is nice to find out in a few minutes time that a document exists which bears directly on the problem you are working on; it is disconcerting to find out that it is only available from a source several thousand miles away and that it will take at least several days to get a copy. The eventual solution might well be microfiche copies of all documents referenced in a search service being available from that search service, but there may be copyright and royalty payment problems to be solved. Note that some microfiche services *are* available, such as University Microfilms, Institute for Scientific Information, NTIS, and ERIC (Reference 7).

It is also worth noting that one of the reasons

that some bibliographic files have not been combined is the copyright problem, we were told. Most copyright problems could be easily solved if the copyright holders could be assured of getting royalty payments for every copy made, including displays on CRT screens, photocopies made, etc. But the information services field has not yet reached this point.

Finally, a problem that was pointed out to us is the rather lukewarm attitude toward information services by middle management in many organizations. When budgets must be cut, the libraries and related services seem to be one of the first candidates for cutting. In the high technology organizations, however, the value of information services is usually recognized by both middle and top management. This is apparently not the case in the bulk of user organizations, we were told.

We can summarize the status of today's interactive search systems by saying that they can provide rapid and efficient searching of huge bibliographic files, involving millions of citations, on an economic basis. These services are only a few years old. Acceptance is growing. But there are problems of the lack of consistency and standardization that annoy users, plus human reading problems, as well as some difficulties in getting access to source documents. In short, these systems are performing a valuable service but there are still obstacles to overcome.

What about the near future? Will some of these problems be solved?

Future trends in search services

From our interviews, it appears that the following are the main enhancements to interactive search services to be expected in the near future.

Stored SDI services. An SDI (selective dissemination of information) is a current awareness type of service; a subscriber receives notices of all new articles, documents, etc., that meet his or her interest profile. The interactive search services are beginning to offer this service. A user performs a search for the type of citations in which he is interested. He then asks that the search criteria be saved, assigned a number, and that his name and address be attached. Then, as the bibliographic files are updated with the new additions (generally monthly), the new additions are passed against the stored search criteria. The subscriber receives a printed listing of all hits, in just

the same manner and format as he receives a list of abstracts printed off-line.

Slave terminals. The idea of slave terminals was brought up during one of our interviews. In asking about it subsequently, we found that the idea has been widely considered. There are technical details to be worked out, of course. The master terminal probably must conduct the search. But a slave terminal must have some way of breaking in and sending a message to the master. Also, the system should probably allow any of the terminals to become the master for a given search.

Higher speed of transmission. As we have pointed out earlier, a 30 character per second speed of display on the terminal is acceptable. But higher speeds do have benefits. Perhaps with the arrival of the packet switching common carriers, it will be possible to get a higher speed of transmission at the cost levels claimed for packet switching.

Microfiche of source documents. The first point is to make here is that the source documents are not always needed. For instance, abstracts of news stories may give all the information that the requestor is seeking. But when the source document is needed, it appears that microfilm or microfiche offers the most economical solution for the foreseeable future—provided that copyright and royalty payment problems can be solved. Eventually, the full source documents may be stored and transmitted electronically, but the economics do not support this yet and the copyright problems would still exist.

More management-oriented data. We gather that the use of the interactive search systems by managers (as requestors) is increasing at a good rate. So, while the bulk of the bibliographic files on today's systems are oriented to science and engineering, it is likely that more and more files will appear that are management and business oriented.

Quantitative data and the analysis thereof. The people at the Lockheed DIALOG service say they expect statistical-type data to be stored in interactive search systems and that the systems will provide analysis programs for handling the data. For instance, a variety of economic data, in time series form, can be stored and retrieved on demand. The user may want to make projections for a period of time into the future, using as a base the historical time series data he has selected.

Speculation on the handling of management-type data and statistical data leads one to wonder just what these services might be like in, say, five years or so. They might well begin offering data services, such as the storage, retrieval, and analysis of up-to-date market data, trade association data, industry data, and so on. They may well ex-

pand their services to doctors, lawyers, etc., for performing medical and legal searches connected with their work.

In any case, practical, mechanized information retrieval is now off and running. It will be interesting to see how it develops in the years ahead.

REFERENCES

1. For more information on the Lockheed DIALOG service, write to Lockheed Information Systems, Org. 52-08, Lockheed Missiles & Space Co., Inc., 3251 Hanover Street, Palo Alto, Calif. 94304.
2. For more information on Information Bank, write to New York Times Information Bank, 229 West 43rd Street, New York, N.Y. 10036.
3. For more information on the SDC Search Service, write to System Development Corporation, 2400 Colorado Avenue, Santa Monica, Calif. 90406.
4. Marron, B., E. Fong, and D. Fife, "A Mechanized Information Services Catalog," National Bureau of Standards Technical Note 814, Feb. 1974; order from Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402; price 90¢.
5. Fife, D. et al, "A Technical Index of Interactive Information Systems," National Bureau of Standards Technical Note 819, March 1974; order from Superintendent of Documents (address above), price \$1.20.
6. Moghdam, D., "User training for on-line information retrieval systems," *Journal of the American Society for Information Science* (1155 16th Street N.W., Suite 210, Washington, D.C. 20036), May-June 1975, p. 184-187.
7. "ERIC: The model micrographics information system," *Information & Records Management* (250 Fulton Avenue, Hempstead, N.Y. 11550), March 1975, p. 42-45.

The question of "information privacy" is looming for most industrialized societies. Privacy legislation has been enacted by some governments—for instance, President Ford signed the Privacy Act of 1974 on January 1, 1975, for application to agencies of the federal government—and is under consideration by many other legislative bodies. The impact of privacy regulations on the data processing function can range from moderate to fairly severe, depending upon the wording of the legislation. Next month, we give the first of two reports on "the debate on information privacy."

EDP ANALYZER published monthly and Copyright© 1975 by Canning Publications, Inc., 925 Anza Avenue, Vista, Calif. 92083. All rights reserved. While the contents of each report are based on the best information available to us, we cannot guarantee them. This report may not be reproduced in whole or in part, including photocopy reproduction, without the

written permission of the publisher. Richard G. Canning, Editor and Publisher. Subscription rates and back issue prices on last page. Please report non-receipt of an issue within one month of normal receiving date. Missing issues requested after this time will be supplied at regular rate.

SUBJECTS COVERED BY EDP ANALYZER IN PRIOR YEARS

1972 (Volume 10)

Number

1. Computer Security: Backup and Recovery Methods
2. Here Comes Remote Batch
3. The Debate on Data Base Management
4. Intelligent Terminals
5. COBOL Aid Packages
6. On-Line Development of COBOL Programs
7. Modular COBOL Programming
8. New Training in System Analysis/Design
9. Savings from Performance Monitoring
10. That Maintenance "Iceberg"
11. The "Data Administrator" Function
12. The Mini-Computer's Quiet Revolution

1973 (Volume 11)

Number

1. The Emerging Computer Networks
2. Distributed Intelligence in Data Communications
3. Developments in Data Transmission
4. Computer Progress in Japan
5. A Structure for EDP Projects
6. The Cautious Path to a Data Base
7. Long Term Data Retention
8. In Your Future: Distributed Systems?
9. Computer Fraud and Embezzlement
10. The Psychology of Mixed Installations
11. The Effects of Charge-Back Policies
12. Protecting Valuable Data—Part 1

1974 (Volume 12)

Number

1. Protecting Valuable Data—Part 2
2. The Current Status of Data Management
3. Problem Areas in Data Management
4. Issues in Programming Management
5. The Search for Software Reliability
6. The Advent of Structured Programming
7. Charging for Computer Services
8. Structures for Future Systems
9. The Upgrading of Computer Operators
10. What's Happening with CODASYL-type DBMS?
11. The Data Dictionary/Directory Function
12. Improve the System Building Process

1975 (Volume 13)

Number

1. Progress Toward International Data Networks
2. Soon: Public Packet Switched Networks
3. The Internal Auditor and the Computer
4. Improvements in Man/Machine Interfacing
5. "Are We Doing the Right Things?"
6. "Are We Doing Things Right?"
7. "Do We Have the Right Resources?"
8. The Benefits of Standard Practices
9. Progress Toward Easier Programming
10. The New Interactive Search Systems

(List of subjects prior to 1972 sent upon request)

PRICE SCHEDULE

The annual subscription price for EDP ANALYZER is \$48. The two year price is \$88 and the three year price is \$120; postpaid surface delivery to the U.S., Canada, and Mexico. (Optional air mail delivery to Canada and Mexico available at extra cost.)

Subscriptions to other countries are: One year \$60, two years, \$112, and three years \$156. These prices include AIR MAIL postage. All prices in U.S. dollars.

Attractive binders for holding 12 issues of EDP ANALYZER are available at \$4.75. Californians please add 29¢ sales tax.

Because of the continuing demand for back issues, all previous reports are available. Price: \$6 each (for U.S., Canada, and Mexico), and \$7 elsewhere; includes air mail postage.

Reduced rates are in effect for multiple subscriptions and for multiple copies of back issues. Please write for rates.

Subscription agency orders limited to single copy, one-, two-, and three-year subscriptions only.

Send your order and check to:

EDP ANALYZER
Subscription Office
925 Anza Avenue
Vista, California 92083
Phone: (714) 724-3233

Send editorial correspondence to:

EDP ANALYZER
Editorial Office
925 Anza Avenue
Vista, California 92083
Phone: (714) 724-5900

Name _____

Company _____

Address _____

City, State, ZIP Code _____