Introduction to Online Information Systems

A collection of the significant papers in the field of the online retrieval of information

Edited by
David I. Raitt

Introduction to Online Information Systems

A collection of the significant papers in the field of the online retrieval of information

Edited by
David I. Raitt

Learned Information Oxford and New Jersey

ISBN 0 904933 45 8

Published by

Learned Information (Europe) Ltd
Besselsleigh Road, Abingdon
Oxford OX13 6LG, England

Learned Information, Inc.
143 Old Marlton Pike,
Medford, NJ 08055, USA

Printed in England by Cotswold Press Ltd., Oxford, England.

"Knowledge? What is knowledge? Everyone knows something!"

Contents

Introduction

"Let twenty pass, and stone the twenty-first,
Loving not, hating not, just choosing so"

Thus wrote Robert Browning in Caliban upon Setebos. Well, although to select the best one or two percent out of over 3000 English language papers on online information systems is an extremely difficult and subjective task, I trust I have applied better selection criteria than did Browning!

This anthology is aimed at the relative newcomer to the world of online. I therefore reasoned that they would be more interested in knowing what online was and how it could be used rather than wanting to know about theories. The articles chosen, hopefully, enable the newcomer to learn something, to pick the brains of others and to benefit from their experience. For students the texts can act as a convenient review or synthesis of lectures.

My criteria for selection included the clarity, readability and conciseness of the papers, their content, originality and coverage, and their pertinence and applicability. In short, whether or not a newcomer to the field could get a feel for the possibilities of online.

To this end I rejected short papers and purely descriptive papers. I have tried to stick to 'timeless' papers, ie papers which, while they were written several years ago are not really out of date in their thinking and detail. I also wanted to get a fairly representative coverage of both European and American authors — though this is made harder by the fact that the great majority of papers written are by Americans. One or two early papers are included so that the reader can see how much (and indeed whether) online systems have changed during the years.

The book is in seven chapters, with Chapter 1 giving some idea of what online is all about. Online information retrieval got off the ground in the mid 1960s although 1969 seems a big date from a number of viewpoints. This was when the European Space Agency first introduced its online service for Europe and when several large database producers started creating machine-readable tapes. Widespread online access for many librarians or information officers, certainly in so far as Europe is concerned, has only become available in the last few years. The papers chosen are all enthusiastic about online and all see it from different aspects — often conflicting. For example, Collier considers online systems are expensive, while Cuadra on the other hand says their use can be very inexpensive.

Chapter 2 looks at some search systems and databases. There are of course many excellent descriptions of databases and services and clearly it is not possible to cover them all. The Monitor survey of the world's major databases and databanks should give the tyro some idea of the complexity and flavour of databases. Going back in time Thompson's paper, although describing an interactive IR system, gives a good insight into the kind of user-friendly features that should be included in the design of an online system. Most of these features are still being included in systems today as can be seen by reference to the article by Edström and Wallin discussing a new software, 3RIP.

And of course, no introduction to online would be complete without some reference to EURONET-DIANE.

Knowing what online is all about, and being aware of some search systems and databases — how does one get started? In Chapter 3 some practical advice on choosing terminals is given in the first two papers, then Magdeleine Moureau draws on her considerable experience in industry to discuss the problems and pitfalls in setting up and operating an online information service, while Simone Klugman describes the problems of introducing an online system into the reference services of a large university.

In Chapter 4 we look at how you can use your terminal to full advantage. To this end, two papers are given on improving search strategies for maximum recall, relevance and time-saving, while another two show how the terminal can be used for other than subject searching. Black-boxes added on to terminals permit the user to prepare profiles in advance, store searches, have automated logon facilities and a paper is devoted to these issues.

Although several of the articles chosen allude to the costs of online searching it was felt necessary to include one or two papers dealing with this interesting and important topic in their own right. The papers selected

for Chapter 5 tend to show that online searching is cheaper than manual searching.

Chapter 6 contains some topics for the online user who wants to learn a bit more about the data transmission aspects of information retrieval, about query languages and methods of searching, the necessary software for micros, and about how to promote an online service effectively. In addition, a paper by Cuadra is included to stimulate the user to consider some of the problems and characteristics associated with online systems.

No book on an introduction to online could be considered complete without a look to the future. In Chapter 7, Butler examines trends affecting the growth of non-library online information systems and their use in the USA. Popper constructs an online information scenario for the 1980s, and looks at its social implications. Isotta looks even further ahead to 2001 to discuss the impact of telecommunications on information services, while Raitt examines some of the possibilities of using satellites in online information systems. Finally, it is appropriate that since we started with Martha Williams telling us, in 1977, what online is, we close with her telling us, also in 1977, what online will be!

This book could have been longer, it could have been shorter, it may even have been better if permission to reprint several excellent papers had not been withheld. However, this present selection of English language articles does provide a cross-section of the available literature and as such will benefit the new online user.

It should be noted that the content of the papers has not been modified or updated in any way and that the authors may no longer have the same title or organizational affiliation as they did at the time of writing their papers.

Chapter One

Chapter One

What it's all about

"*About it and about*"
(E. Fitzgerald. The Rubáiyát of Omar Khayyám Ed 1 XXVII)

Editor's Introduction

When digital computers first came into use some thirty years ago they were not terribly sophisticated, being bulky, expensive, and relatively unreliable. These early machines were basically number processors and were used mainly for scientific and engineering calculations. Advances in technology (especially microcircuitry) over the next decade or so, led not only to reduced bulk but also to increased reliability and speed, enabling the newer machines to be used for other arithmetic applications such as payrolls, invoicing and inventory control.

Further developments, particularly the ability of the computer to handle letters of the alphabet as well as numbers, made it of importance to those involved in the storing and retrieving of bibliographical information in addition to other management and business tasks.

Despite the fact that these computers offered fast, complete and accurate data they nevertheless suffered from one major drawback — the processing of data was done in a batch mode.

Batch-mode processing involves the sequential entry and processing of jobs which on completion would be returned to the user. Any data the user wanted to be put into the system and processed would be given to a computer operator and treated as a job. Because of the expense and time required to execute jobs, several of these would be grouped together in a batch — all subsequent operations being done on the batch as a whole.

Thus the batch would be scanned and only when that operation was complete might the batch be merged and sorted. The whole process could take hours or even days.

The major drawbacks of batch processing are that the user is denied direct and immediate access to the computer. The lack of direct access means that the user cannot modify the program during its execution or processing; while the failure to have immediate access means the user cannot interact or have two-way communication with the computer.

These drawbacks of batch processing systems have been successfully overcome with improvements in computer communications, allowing a number of users to be connected directly to a computer by a communications or telephone line, and to be therefore literally 'online' at the same time. Also computer programs can now allocate slices of processing time to each user (time sharing), giving them the required immediate access (real time).

In a batch system the computer merely executes procedures and computes problems already laid down and solved in advance by the programmer. The online system, by virtue of its immediacy and feedback, permits interaction and allows the procedure to be laid down or modified and the problems to be solved at the time of data entry.

The interaction takes place through an interface which can be defined as a physical/conceptual structure which both channels and facilitates communication.

This user interface in an online interactive system can be considered to be composed of five component parts, all of which interconnect to make up the whole interface. These parts are:

- –the terminal (console), be it a visual display unit or typewriter type, where all the response to and from the computer will take place;
- –the communication link, which enables the dialogue to be conducted with or without certain errors;
- –the database, which should enable the user to satisfy a request;
- –the operational characteristics (user language) of the system, which allow the user to proceed in his own time at his own pace to retrieve all the relevant information he needs; and,
- –the user, without whom the system could not function.

The search and retrieval technique for an online information system, is usually based on the principle of recursive searching. This means that the formulation (i.e., the selection and logical interconnection of con-

cepts) of the search question is constantly modified by the searcher after examination of the results produced by any particular formulation.

The searcher, having first selected the database in which he wishes his search to be carried out, can ask to see a list of terms alphabetically close to one he specifies. These terms can be subject terms, authors' names, classification codes, list of languages, etc. From this list the user can choose the appropriate terms that he feels match his particular concept. Each term so chosen is given a set number together with the number of references relating to it.

This process of looking up terms is continued until the searcher is satisfied he has selected enough terms to cover his subject. The searcher will then interconnect (combine) the set numbers of these terms using Boolean logic; each interconnected set of terms results in a further set with an indication of the number of references.

These resulting references can be viewed either on the screen, if the searcher is using a video terminal, or on the paper of the teletype type of terminal. By looking at the references the searcher can decide whether they are relevant or not, as well as being able to discover additional terms which could be used in a different search formulation. By selecting and combining these new terms further sets result that can be examined in the same way. The process continues until the user is satisfied that he has as much information as there is, or he can cope with, displayed; he will then order the references to be printed off-line for later mailing.

So an online interactive system could be considered as one where the input data or request is transmitted, via an interface which facilitates communication, directly to the computer from a user who may be some distance away; where the output data is transmitted directly back, via the same interface, to the same point; and where the remote user and the computer, by means of the interface, continually and alternately exchange data which constitutes a feedback response to each without the intermediary operation of punching data onto cards or tape.

Some of the reasons for using online systems and the advantages and benefits of interactive retrieval can be summarised as follows:

(a) validity checks can be made at the time of data entry,
(b) there is the possibility of using a wide variety of terminals,
(c) no middle man (interpreter) is required; therefore, distortions could be eliminated,
(d) there is no wait for batches to be processed, instead the transaction is immediate,
(e) since decisions can be made by the user rather than the computer, more accuracy should result,
(f) there is the possibility of obtaining intermediate results and modifying them there and then,
(g) the quality of output is enhanced due to being able to interact with the computer in simple steps (Meadow aptly calls this the "bargaining power"), and
(h) the possibility to search on larger bodies of information exists since no serial scanning is involved.

Online and some of its applications and advantages are discussed in the following papers.

We start with Martha William's paper on Networks for Online Data Base Access which discusses network sharing of resources and the reasons for using networks for online database searching. These include: retrospective searches, quick question, current awareness, search strategy development, training and bibliometric studies. She then goes on to discuss the requirements for online searching in terms of equipment, database access and search strategy development and the costs, advantages and acceptance of online services.

Also in 1977, Barraclough discussed the developments of online systems: both the advances in computer technology that made it possible and the use made of the changes by information retrieval systems, in such a way that the reader may, by understanding the reasons for the development, rejection, and acceptance of various techniques, judge the good and bad parts of current systems. Among the aspects she discusses are the concepts of online processing and interaction, and the information retrieval problem of producing a good definition of the document to be entered into the database so that it can subsequently be retrieved.

She then goes on to state the expectations from online systems and to give the effect of using terminals to retrieve information. This involves user instruction, knowing the command language and how to formulate queries, and leads into the performance and costs of online systems in the future.

Collier takes a look at the advantages and disadvantages of online, compared to conventional printed indexes. His view is that printed indexes offer clearer advantages in search formulation and browsing, in the length of back-files available for searching, in the cost of extensive usage, and in the ability to do certain searches. Nevertheless he believes that the online terminal is a complementary tool. Cuadra gives some figures on the number of databases available and a description of online searching to illustrate their advantages — they are fast, expensive, precise, and democratic. He follows this with some more details of the growth of commercial services, looking at some of the available systems and their features, and the factors which have led to such growth (such as the declining cost of online searching). Of the 15 or so online retrieval systems of international scope or potential operating at the time of the paper (1977) only 3 were non-US based. Compare this with today when there are now around 40 on Euronet alone plus another dozen or so on SCANNET. Concluding, he looks at the next 5-10 years and some improvements such as multilingualism.

Finally, Bourne gives an excellent summary of the history and development of online systems starting with the application of computers to batch bibliographic searching in 1954. The technology of online systems — communication lines, storage, programs — is also briefly reviewed. The economics of online services are discussed by an analogy with publishing, and a consideration of the expenses associated with online search services, and current files and pricing policies.

1 Networks for online database access

Martha E. Williams
Coordinated Science Laboratory
University of Illinois
Urbana, IL 61801, USA

New potentials for data base use in information centers emerge daily. The potential data base user may never have used a data base on-line and may not even know what is meant by an on-line system or a network. Since at any time the responsibility for choosing and implementing on-line search services for the potential user organization may arise, it is imperative to appreciate the role networks play in current data base activity.

Networks

The term "networks" can be used in different ways. Loosely defined, it can refer to any kind of formal cooperative arrangement where people agree to exchange information or resources. Many of the library networks in existence today are largely based on exchange arrangements. The kind of networking which is of concern here, however, is electronic networking that permits sharing and remote use of hardware, software, and data bases.

Reasons for Using Networks

Computer readable data bases now exist in large numbers. Ten years ago there were only a handful of machine readable, publicly available data bases. Now there are hundreds. Some of these data bases are large and they are growing; many data bases increase by hundreds of thousands of records per year. ASIS recently published the compilation entitled *Computer-readable Bibliographic Data Bases — A Directory and Data Sourcebook* [1] including more than 300 data bases and more than 100 additional data bases have been identified for possible future inclusion. Of all these data bases more than 75 are currently available on-line through networks.

Cost-sharing constitutes one of the biggest advantages of network use of data bases. Cost-sharing makes the real difference and is the primary reason why people are using networks. If the user were to do his own processing of data bases he would have to pay $1,000 to $50,000 for a one-year subscription to one data base. In a large organization which needs ten to fifteen different data bases, several hundred thousand dollars could easily be spent annually for data base leasing. Add to this the cost of software development, which can be anywhere from a hundred thousand to a million dollars for an on-line software package. The cost of computer time is also shared. (Computer time is calculated differently in different organizations; rarely do two organizations calculate it the same way, but the costs are high). The use of tape drive, disk drives, printers, and any other kind of peripheral equipment constitutes an additional cost that is shared. Finally, the user needs a communications network such as TYMNET or TELENET. By using a communications network the user pays less for telecommunications than if he were to dial direct to a particular computer, so networking again saves money in the communications area by permitting cost sharing. By using data bases on-line via networks all users of a specific system share in the personnel costs incurred by the system operators. The organization that processes the data bases has to pay personnel to design the system, develop it, manage it, maintain it, sell the services, and dozens of other expensive activities. Through the use of networks and the shared use of data bases, all these costs are shared by thousands of individuals and/or organizations.

Adequate, affordable communications technology now exists to accommodate the shared use of data bases. Ten years ago communications technology for remote use of data bases, although it existed to a limited extent, was extremely expensive and not publicly available. Communications networks that now exist include government sponsored networks, state sponsored networks, and commercial communications networks. The communications networks most commonly used for communicating with on-line data base services are TYMNET, and more recently TELENET.

The availability of on-line retrieval systems also facilities the shared use of data bases. Several years ago

there were very few on-line retrieval systems, and they were not nearly as efficient and sophisticated as they are today. In 1973, 700,000 on-line retrospective searches were conducted via the on-line services provided in the United States and Canada. In 1975 there were a million and in 1976 1,200,000. [This volume of activity includes only publicly available bibliographic data bases, and does not include the use of library-type data bases and systems such as the Ohio College Library Center (OCLC) and other types of on-line access to library files; nor does it include searches of numeric data banks.] This impressive growth rate will probably continue at a rate of 20% or more per year for at least the next five years.

Networking can take place in the profit-making sector, the not-for-profit-sector, and in the federal government. In the past, most of the activity has been within the federal government. Now there seems to be a trend working in the other direction with the private sector becoming more and more involved in networking. The government often provides the initial funds for R & D and then withdraws funds after the ground work has been laid, leaving it up to individuals in the for-profit and not-for-profit sectors to continue.

Among large bibliographic data base network activities, probably the largest is the National Library of Medicine's MEDLINE. It was one of the earliest systems to provide on-line retrieval services and has become the largest such system. Initially the service was provided free. Now, a comparatively small fee has been charged for the use of MEDLINE as well as other NLM data bases. The DIALOG system of Lockheed [2], the ORBIT system of System Development Corpor ation (SDC), and the New York Times Data Bank also rank among the country's largest on-line data base services. Mead Data, providing a variety of legal data bases, is another, and a newcomer to the field is the Bibliographic Retrieval System (BRS).

Other on-line retrieval systems include the National Aeronautics and Space Administration's (NASA) RECON, the ELHILL system, and RECON II. Several of the software systems are related. For example, the NASA RECON software was originally developed by Lockheed. A variation of the NASA RECON software called RECON II is used by Informatics. Similarly, the ELHILL software system used for NLM's MEDLINE was originally developed by SDC. SDC's own system is called ORBIT. The software system used by SUNY (State University of New York) is the IBM developed package called STAIRS. The system used by BRS is an adaptation of STAIRS. The cost of developing an on-line retrieval system is very high, and it is doubtful that many new companies will enter the on-line software business by developing totally new software. It is more likely that they will build on the foundations of others.

Network Emphasis

Various networks have different emphases in fulfilling the varying needs of users. One type of network involves use of *hardware* on which the user can run his own programs. Another type permits the use of special-purpose *software* which may be used to operate on the user's own data. In a third type of network, the user remotely accesses specific *files* (data bases) in order to answer search questions. A fourth type of network provides *communication* links between users and the computers where the data bases and software are used. The distinction between different networks is a difference in emphasis. The ARPANET network, for example, one of the earliest networks and federally subsidized, primarily provides use of hardware. ARPANET also provides software, but most users are primarily interested in buying computer time through ARPANET. Through ARPANET the user can decide, on the basis of his own needs and his specific problem, which computer in the network has the appropriate capabilities for his application. Another network which provides hardware is the NER/COM, or North Carolina Network. TYMNET represents a communications link network. The direct service provided by TYMNET is neither hardware nor software to aid the user in processing data. Rather, it provides communications and the ability for the user to reach another computer in another location. NLM, SDC, and Lockheed are examples of networks that provide data bases or files. The above are examples of different types of networks with different types of emphasis.

Data Bases

The concern of this paper is with the kind of network that provides access to files of data bases. The data bases used in these networks can be numerical, representational, alphanumeric, or mixed. A numeric data base is exemplified by the census tapes. A representational data base contains pictorial or graphic information, as in the chemical structures files that the Chemical Abstracts Service (CAS) provides. Alphanumeric data bases can be further broken down into full-text data bases, partial-text data bases, bibliographic (reference) data bases, and bibliographic-related data bases. A full-text data base contains the whole text of a document. The expense involved in keying in large quantities of information poses one of the biggest limitations on the creation of full-text bases; consequently, there are only a few full-text data bases because they are so expensive to build. An example of a full-text data base is LEXIS, which contains legal statute information. In law, it is often important enough to have information in context that the cost of incorpor ting full-text information in data bases in borne. A partial-text data base contains portions of texts: extracts, abstracts, introductions, summaries etc. The Information Bank of the *New York Times* is a partial text data base which contains shortened versions of all the articles which appear in the *New York Times* plus selected material from some seventy other publications in the United States; full text information not available on-line is available in microform, off-line. A bibliographic data base is one that contains references or citations associated with literature sources. Chemical Abstracts' CA Condensates, Engineering Index's COMPENDEX, and BioSciences' BIOSIS PREVIEWS are a few examples selected from among several hundred bibliographic data bases. A bibliographic-related data base is one that does not contain bibliographic citations but does contain information which sends the user to those

citations. CASIA (CA Subject Index Alert) is a bibliographic-related data base produced by Chemical Abstracts Service. CASIA is basically an index data base containing neither abstracts nor citations but index terms and abstract numbers (related to the bibliographic citations) which point to the abstracts.

Distinction between Terms

When dealing with data bases the user needs to distinguish between the names of communications networks, such as TYMNET, TELENET, GE, etc. and the names of parent organizations that provide data base services through networks, such as NLM, Lockheed, and SDC. Users must also distinguish between names of information service organizations within the parent organizations, and the name of the service itself. For instance, the service provided by NLM is called MEDLINE and the name of Lockheed's service is Lockheed Information Service. The user also needs to know the names of software packages supporting the services and the names of the associated command languages. The name of the software package behind NLM's MEDLINE, for example, is ELHILL 3. The name of Lockheed's software package is called DIALOG. In this case, as in some others, the name of the software package is the same as the name of the command language. Finally, the user needs to distinguish between the name of any particular data base, and the name of a data base as it is used within the search system. Most vendors of on-line search services and centers providing data base search services assign names given to the data bases which are different from the names given to the data bases by the data base producer. For example, Chemical Abstracts Condensates is called CACON by its producer, CAS. But that same data base is called variously by about thirty different names — including, CHEM 2, CA-odd, CA-even, CASCON, Chemical Abstracts and CHEMABS — in other organizations. This plethora of synonymns for data bases and confusion regarding appropriate nomenclature for communication networks, parent organizations, on-line service organizations, service names, software packages, and command languages leads to misinterpretation of information about services and systems and data bases. Unless authors and speakers are more careful in their use of terminology this confusion will be compounded.

Uses of On-Line Searching

There are many uses or reasons for on-line searching [3]. Some of them are retrospective searches; providing quick answers to questions; providing current awareness or SDI (selective dissemination of information) services; developing search question strategies and selecting terms; training personnel in the use of on-line systems and training them to observe data base differences; and for bibliometric studies [4].

Retrospective searches

People do on-line searching for various reasons. Some do it in order to get a comprehensive retrospective search. They want to get as much information as possible on a particular topic. Obviously, the user is limited by the availability of data bases and number of years' worth of data available on-line, but often a major portion of the search may be done on-line.

Quick answers

Some people use on-line searching to get a few refer nces on a topic. Other users who want to be introduced to a new field can make a very quick search to acquire some preliminary information (often called quick and dirty searches). Some want very fast answers to specific questions; the volume of this type of search will probably increase in the future as more data bases containing statistical information and handbook type information become available.

Current awareness

Current awareness represents another type of use for on-line data bases. In fact, the data base business got its start through current awareness or SDI. Several years ago, when most data bases were just beginning, the only service that could be provided was current awareness. Obviously, the first month of a data base issue certainly does not allow for retrospective searching, nor do the first six months, eight months, or even the first year; there is simply not enough material after such a short time. But enough years have gone by that now we have the retrospective search capability. The current awareness capability still exists, of course, and can be done on-line as well as in batch mode. The searcher may store his question on the on-line vendors' computer for subsequent recall and use by him. Alternatively, the question can be stored by the vendor and processed by him against incoming issues of a specific data base.

Profile/search question (strategy) development

On-line searching may also be used in the development of SDI profiles, even if the SDI service will be provided in the batch mode. A more common but less efficient means of developing search profiles for SDI involves submitting a question, waiting until the batch is run and the results come back, then rewriting the question. This method may require about three batch runs before the search question has been refined sufficiently to yield what the searcher is looking for. Search profiles (terminology selection and strategy) can be created much more quickly and easily on-line; an on-line service may be used to experiment with terms, logic and truncation forms. The refined search question development can be handled similarly.

Training

On-line searching also facilitates the training of personnel in search techniques. New searchers can learn to formulate search questions much faster on-line than they can with a batch system. Any attempt to train personnel by using a batch system will need to deal with long lag periods between the preparation of questions and the analysis of results. On-line, a new searcher can observe immediately how various applications of logic, truncation, etc. will effect search results. On-line training can also be used to aid people in observing differences in data bases. The searcher can quickly try one term

in data base A and the same term in data bases B and C to see that different results are produced from different files. Any particular term may occur with a very different frequency in each file and the term may bear a different meaning in each file. Consequently, the searcher will get very different search results. A well trained searcher may be aware of the effects that term frequency and homography have on the results of searches. These differences in term performance can readily be seen on-line.

Bibliometric studies

On-line searching can also be used to conduct various types of bibliometric studies of the data bases. Journal coverage statistics, subject coverage of specific journals, index entry statistics and many other types of statistics can be generated on-line.

Requirements for On-Line Searching

Any organization considering the adoption of on-line search services must first of all be certain that there will be users with search questions. Presumably in most organizations people do have questions. But the primary advantage of on-line searching is that the organization with the questions does not have to maintain the computer, the software, the data bases, and all the necessary peripheral equipment to answer them. All that is required is a telephone, an acoustic coupler to connect the telephone to the communications systems, and a terminal which may be either a cathode ray tube (CRT) or a hard-copy producing terminal. Also, the organization needs to use a local telephone line to provide access to a communications network. (Alternatively, one can use a WATS Line or TIE line to access the communications network or he can dial direct to the computer on which the on-line search is to be conducted.)

The relative cost advantages and ease of using a network have already been mentioned. Via the telephone lines, satellite, or other communication systems involved in the communications network the user organization accesses a data base service organization such as NLM, Lockheed, SDC, Mead Data, BRS or the *New York Times*. Once connected to the data base service organization, the user will need to be connected to the computer in organizations where the data bases reside, and to the right software package, (e.g., DIALOG or ELHILL). The user must then select the appropriate data base. Note that the user dials the communications net, is connected to the on-line vendor organization through the communications network and, once in the organization, to the right computer, the right software package, and the specified data base — all within less than a minute (usually). As soon as he is connected to the right organization the user then indicates his own password or code. This starts the meter running so that the user will be charged correctly for the use of the system. The cost varies with the data base(s) selected for his search.

In a straightforward search the user begins the search by entering search terms. A user may enter a term then wait to see the terms related to or alphabetically close to that term; the user may employ what is called a "neighbor command" — this feature of command is referred to in different ways in different systems — which gives the user a display of the desired term and the alphabetical neighbors on both sides of it. For example, an alphabetical display of the term "synthesis" will yield other terms like "synthsize," "synthesizing," synthesized," etc., all of which might be helpful in formulating the search question. The searcher then knows which variant forms of his term exist in the specific data base, and whether or not the one term he chose will be adequate or whether he should search on a truncated form of the word.

Once the searcher has selected the terms he wants, he uses a command that lets him create a file. This command takes all the citations, for example, related to the term "synthesis," and all the alphabetically close terms, and puts them in a special file (his work space). The user thus creates his own subset of the data base. He does this with all of the terms in the question. The user then employs logic operators — AND, OR or NOT — in combining the various terms and narrowing the search question. Alternatively, he may use proximity logic to specify the contextual relationship between terms.

When the searcher feels the question has been sufficiently narrowed, he may want to display some of the citations the question produces to see if they are relevant. If they are, he uses other commands to tell the system to print the citations on-line (assuming he has a printing terminal) or to print the citations off-line and mail them. After the user has looked at his citations and decided what to print and send, he logs off and the search is finished. Some searches go quickly and some require a longer time; but an average search takes about ten to fifteen minutes.

Cost of on-Line Searching

The costs involved in providing on-line search services via networks are considerably lower than the costs of processing data bases in-house to provide data base search services. If the information center of an organization already has a local telephone, suitable terminals, and accoustic couplers no additional equipment is needed. It may not be necessary to hire additional personnel if the same people who do other types of searching can be trained to do on-line searching. If equipment is not already on hand some equipment expenditures will have to be made. A terminal with an accoustic coupler rents for about $140 to $200 per month. Beyond local equipment the other out-of-pocket costs involve communication charges, data base use charges and printing charges. Communications network charges vary from about $3 to $13 per hour. Data base use charges range from approximately $15 to $125 per hour of connect time (not CPU time). The cost of printing the output constitutues an additional cost that may be related to either the number of pages or citations printed. Searchers with printing terminals can print citations on-line, at their own locations, but this method is expensive because the user must still pay the hourly data base use or connect charge just to print. For users in a hurry, of course, the extra expense is often justified

by having the citations printed immediately instead of having to wait a week for the mail delivery of low-cost on-line printed citations. By mail, prints cost between five and thirty-five cents per citation, depending on the processing organization's charging mechanism, overheads, and fees. Document delivery may also represent a cost related to the search, though this is usually not considered a part of the search cost [for additional discussion or cost see [5]].

Advantages of On-Line Searching

Many benefits in addition to *economic benefits* result from the use of on-line services [6]. The user can search *large files* quickly and can search *multiple sources* without moving from his terminal. If the user were in a library and wanted to search three different abstracting and indexing journals, he would probably have to walk around several different areas of the library to find them. But using an on-line search system, he simply pushes a couple of buttons to get out of one file and into another. And, the transition takes place in a matter of seconds. Users experience *fast turnaround time* — from the start to the finish of a question. Also, on-line vendors provide messages and *news on-line*. Notices of file changes, file additions or changes in system features are made available to the user on-line.

On-line services facilitate *question negotiation*. On-line, it is easy to develop a search question and get an immediate indication of the type of output which will result. One can also *store search questions* on-line for later use. If a searcher has developed a search question and likes the results it yields, and if he thinks he might want to run that same search on new material two months from now, he can store the search question and process it against new material at any time. Similarly, portions of search questions (lists of synonyms, for example) can be stored for later use.

On-line, the *searcher can be more productive*. A person who does manual searching might be able to accomplish two or three searches in a day (assuming they are relatively short). The same person can do approximately eight on-line searches per day including pre- and post-search time. Thus, more service is provided with fewer personnel hours.

On-line services provide *formatted copy output*. Bibliographic citations will be formatted in a similar way for each of the data bases a particular organization processes. If the user is preparing a bibliography he will not have to rearrange bibliographical information, as would a person who does a traditional library search. The bibliographic references are formatted differently in different journals and A & I sources. On-line search services permit standardization of output.

Another important advantage of on-line search services springs from the *ease of demonstrating* this type of service for sale internally within an organization. On-line services can be effectively demonstrated in a few minutes so that potential users know immediately what the search service will do for them. Also, since most people are familiar with other on-line systems, such as the airline reservation systems, on-line search services will not represent an entirely new concept. It is much easier to demonstrate and/or sell a service when the audience already has a basic understanding of the techniques employed.

Finally, and most importantly, the economic benefits of on-line searching must again be highlighted. As discussed, on-line searching can usually be done much more *quickly* and *inexpensively* than manual searching or batch processing.

Searcher/Intermediaries

Though most on-line systems were designed for the end user, most on-line systems are not used by the end user. Generally, an intermediary uses an on-line system. This does not mean the requester or any other user cannot perform the search; he certainly can. Most users, however, will not maintain familiarity with multiple command languages, and multiple data bases, and will not want to do their own searching on a regular basis. Usually an information specialist, who acts as an intermediary for the end user or requester, will do the actual searching. The specialist will get the search question from the requester, then do the search on-line. Or the specialist can do the search with the requester, then do the search on-line. Or the specialist can do the search with the requester present at the terminal; this is most desirable since the intermediary has the expertise that is required for understanding the system, the command languages, the data bases, and the vocabularies, while the end user is present to aid in question negotiation and to indicate what he does or does not want in his output. The question can be developed faster and fewer irrelevant items will be produced. The information specialist might also do the search while in remote contact with the requester, who can direct the search from another telephone. Alternatively, slave terminals may be placed in different locations with the intermediary or searcher at one location and the slave terminal in a remote library or laboratory allowing the requester to participate directly in the search without moving from his own convenient location.

Commercial organizations, academic organizations [7], governmental organizations, and public libraries all use on-line search services, but the use varies tremendously. Both SDC and Lockheed have observed that the major users of their systems are based in commercial organizations, while the lowest volume of use occurs in public libraries. The *New York Times* reports just the reverse experience; most of their users are in public libraries and the fewest are based in commercial organizations. These trends hold true for several reasons. The data bases available through SDC and Lockheed are largely of interest to researchers; it would therefore be expected (as evidence bears out) that research-oriented commercial organizations would be interested in those data bases. On the other hand, patrons of public libraries generally have less interest in scientific material and more interest in news. Marketing techniques definitely have had an effect on the kinds of users SDC and Lockheed have attracted. SDC and Lockheed directed their marketing efforts first at commercial organizations because commercial organizations have for many years spent money on research services. Public libraries have been less inclined to spend money on search services, and academic organi-

zations are inclined to purchase services only if the researcher has a grant or contract money to pay for them. Even among public libraries and academic institutions, though, the use of data bases through networks is growing.

Acceptance of On-Line Service

The acceptance of on-line systems has been rapid and continues to grow. The acceptance and growth rate can be attributed to several things. An increased familiarity with computers has developed throughout the population of potential users. Most people take some type of computer course in high school, or at the college or university level; the number of people enrolled in computer courses now is very large. This was not true ten years ago. People are also more familiar with other types of on-line systems, such as airline reservations systems, theater reservation systems, or banking systems.

The wide availability of computer terminals is another contributor to the acceptance of on-line search services and networks. Terminals are used for so many other purposes that most large organizations already have terminals. Staff members are familiar with the use of the terminals and the same terminals can often be used for on-line searching.

Prior use of SDI has greatly facilitated the spread of on-line services. People who are familiar with SDI systems know what data bases are, and what computers are, and are among the first to be interested in buying on-line services.

Back in the late sixties there were very few years' worth of any file; now many files go back five, six, or ten years. And in addition to the good retrospective search capability which now exists, there are enough data bases to serve almost every need and interest.

Finally — and this represents possibly the clearest advantage for the use of on-line search services and networks — the monetary investment required to provide on-line search services is small considering the returns. If, within a particular organization, there is not enough demand to warrant the purchase of a terminal or to train personnel to do on-line searching, that organization can buy search services from a search service, such as, EDITEC in Chicago. Organizations such as EDITEC represent a new kind of intermediary. They are third-party users of data bases, or brokers. They do not process data bases themselves, nor do they use the search results. They do on-line searching for other people or organizations, by processing searches on a variety of systems against those data bases they deem useful. This new phenomenon in on-line search service brokering reflects the increasing demand of the general user population for on-line services; the wave of demand has only begun to make itself felt.

References

[1] Williams, M.E.; Rouse, S.H. 1976. *Computer-Readable Bibliographic Data Bases. A Directory and Data Source-book*. Washington, DC: American Society for Information Science; 1976. 814 pp.

[2] Summit, R. 1977. "The New Age of Computer-Aided Information Access." (Paper presented at AAAS Meeting, Denver, CO, 1977 February 23.)

[3] Wanger, J.; Cuadra, M.; Cuadra, C. 1975. On-line Impact Study. Survey Report of *On*-Line Users, 1974–1975. (A Brief Summary Report). Santa Monica, CA: System Development Corporation; 1975 December. 17 p.

[4] Hawkins, D.; Miller, B. 1977. "On-line Data Base Coverage of the On-line Information Retrieval Literature." *On-line Review*. 1977 March; 1(1): pp. 59-69.

[5] Cooper, M.D.; Dewath, N.A. 1976. "The Cost of On-line Bibliographic Searching." *Journal of Library Automation*. 1976 December; 9(3): pp. 195-209.

[6] Williams, M.E. 1977. "The Impact of Machine-Readable Data Bases on Library and Information Services." *Information Processing and Management*. 1977; 13: pp. 95-107.

[7] Wax. D. 1975 "NASIC and the Information Services Librarian: Room in the Middle." In: *Proceedings of the Annual Clinic on Library Applications of Data Processing*. F. Wilfred Lancaster, ed. Urbana, IL: University of Illinois; 1975.

Reprinted, with permission, from *Journal of the American Society for Information Science*, **28**, 5, Sept. 1977, pp. 247–253.