DOCUMENT RESUME

ED 126 942 IR 003 859

TITLE System Requirements for On-Line and Batch Retrieval.

INSTITUTION American Society for Information Science, Washington, D.C. Special Interest Group on Computerized Retrieval Services.

REPORT NO ASIS-SIG-CRS-1

PUB DATE Jun 76

NOTE 11p.; Panel Discussion of Special Interest Group on Computerized Retrieval Services at the American Society for Information Science Annual Meeting (Boston, Massachusetts, October 28, 1975)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.

DESCRIPTORS Computer Oriented Programs; Computer Programs; Conference Reports; Costs; Data Bases; Data Processing; Dial Access Information Systems; *Information Needs; *Information Processing; *Information Retrieval; Information Science; *Information Systems; Library Automation; *On Line Systems; Relevance (Information Retrieval); *Search Strategies; Systems Approach

IDENTIFIERS Batch Processing; Computer Users; New England Research Application Center

ABSTRACT

Three papers on system requirements for on-line and batch retrieval presented at the American Society for Information Science (ASIS) annual meeting are included here. At G.D. Searle, data for records related to pharmacology screening are used in a batch system, and an on-line system is used to search information on mutagenic, carcinogenic, and teratogenic chemical compounds and agents. The second paper explains the balance of batch and on-line services used by librarians to cope with information needs. Multiple data bases, data base output form, and efficient use of the computer are now important issues in library information retrieval. An information retrieval computer at the New England Research Application Center, described in the third paper, operates a fast-batch computer search system which costs less than a system using on-line searches. A summary of the comparison between batch and on-line processing is given in the final document, the ASIS panel discussion. (CH)

ED126942
IR003859

SIG Technical Publication Series

asis

CRS-1

June 1976

SPECIAL INTEREST GROUP on COMPUTERIZED RETRIEVAL SERVICES (SIG/CRS)

Panel Discussion of

# SYSTEM REQUIREMENTS for ON-LINE and BATCH RETRIEVAL

Papers from Joint Session of SIG/CRS and SIG/User On-line Interaction, 1975 ASIS Annual Meeting, 28 October 1975, Boston, Mass.







AMERICAN SOCIETY FOR INFORMATION SCIENCE • 1155 Sixteenth Street, N.W • Suite 210 • Washington, D.C. 20036 • Telephone: 202-659-3644

COMPARISON OF SYSTEM REQUIREMENTS
FOR ON-LINE AND BATCH RETRIEVAL

Joan S. Fortune

ABSTRACT

Most data retrieval systems combine both batch and on-line features, as a function of the nature of data bases and nature of queries. Two functionally different retrieval systems in current use at G. D. Searle, one primarily on-line oriented and one batch-oriented, can serve as models for discussion. A data base of 500,000 records related to pharmacology screening is used in a batch system. Since the data are highly structured and comprised of a small number of data types, the searcher has no need to interact with the data base on a sampling basis. A batch system is therefore employed. Requests are relatively few, so the file is mounted on demand. Updating costs are kept low by adding new data to the end of the file.

Another system, containing published information on mutagenic, carcinogenic and teratogenic chemical compounds and agents, requires multi-level searching on textual and chemical structural elements. Interaction with the searcher is important because of the active nature of the data base, the complexity of the frequent questions and the relationships among the data. Hence an on-line system is preferable and is used.

In practice, there are probably very few purely batch or purely on-line data retrieval systems. What is frequently encountered is some combination of batch and on-line where the combination ratio has been determined by a number of factors. Among these factors are the following: First, what is the nature of the data in the database? How many unit records are there, and how many and what kinds of entries are there for each record? A second factor relates to the type of query generally encountered. Are the queries rather straightforward requests for selected lists of stored data, such as, "Find all occurrences of "A" or perhaps "Find all occurrences of A with B"? or do the queries require more complex logical combinations, such as "Find all A and B or C and D but not any F, G, or H"? Most important, however, is whether there is any need to see a sample of hits before proceeding to the next step of the query or to another query. Also, what volume of information is generated by the query? Another factor is who searches the system? What and how much does the searcher have to know about the indexed content of the system to search it successfully? A fourth factor is what I've called "computer facilities", which includes not only the size of the computer and its associated hardware and software, but also the available modes of data storage, and the expertise and cooperation of the analysts and programmers to build and maintain a system tailored to meet retrieval needs. Finally, a point that needs no discussion -- how much does it all cost?

At G. D. Searle, there are a number of functionally different retrieval systems in use for searching scientific information. One of these is searched in a mode that is primarily batch, while another system is primarily on-line. A discussion of these two systems will, hopefully, illustrate how batch and on-line modes have been combined to meet scientific retrieval needs.

The first system that I'll discuss is the one that is primarily batch. It contains pharmacology screening data on Searle compounds. There are approximately 500,000 records stored sequentially on tape. The unit record is 72 characters long, divided into 12 fixed fields, each of which is designated for entries of a particular type, such as test category, compound tested, dose, results, date reported, etc. For any 72 character string, the data contained in the 12 fields are linked to each other only, and not to, any other 72 character string. This means that associations can be made within one string, but not from one string to another. Searching this system is one step away from a purely batch mode. The searcher accesses the computer from a remote terminal and specified, in a question and answer mode, the desired combination of data elements to be searched. In this case, I've requested data on

2 hypothetical compounds when they are active in a hypothetical test category. I have specified the fields on which the hits should be sorted, a job number is assigned, and I have signed off to wait for the computer operator to run the search against the data-containing tape. When the entire tape has been read, I will receive a sorted printout of the data. This essentially batch system is appropriate in this case for several reasons. Because the system contains a small number of data types (i.e., compound number, date, dose, results, etc.) one can specify quite precisely what is needed and anticipate the nature of what will be retrieved. There is really no need to see and evaluate a sample of the data before receiving it all, since the searcher knows beforehand what to expect. Moreover, since there are relatively few search requests, the tape may be mounted on request, and turnaround time is short. The off-line printing of relevant data in this system is advantageous since the amount of data generated is often quite large. Since there are only twelve data types, and since the computer questions are in a standard format, search requests can be defined exactly by a requestor and the search run by anyone who knows how to access the computer. Reading the entire data tape constitutes a significant element of the search cost, but updating costs are reduced since additions to the data tape are made by adding new data in sequence to the end of the file.

The second system I'll discuss is a mostly on-line, highly interactive, multilevel data retrieval system called MUTAGENS. It contains published information on mutagenic, carcinogenic, and teratogenic chemical compounds and agents. The information in the system has come from two sources. The first, and larger source, is the Environmental Mutagens Information Center at Oak Ridge National Laboratory, which has provided us with its tapes of indexed information. Each record consists of a bibliographic reference, chemical and/or physical agents studied, organism studied, test object and CAS Registry Number. To each record containing a chemical compound we have added the Wiswesser Line Notation for its chemical structure. The second source of information in this system is our own indexing of published papers on mutagenesis. We index papers in the same general format as at EMIC, but have expanded the indexing to include data on dosage, length and time of treatment with a mutagen, the type and strain of organism studied, mechanism of action, mutation frequencies, and cell type as well as the Wiswesser Line Notation for the chemical structures. There are presently 220,000 records, 10,000 of which are coded chemical structures. The information in the unit record is of four types: bibliographic references, keywords such as animals, organisms, assay type, mechanism of action, etc., fixed field entries such as fields for dosage and treatment schedules and strings of characters of variable length such as WLN or comments. The information from both sources is stored on disc in a random access, inverted file format. That is, one can get any record in the data file without having to read any other record.

This system is searched in a purely on-line mode. After accessing the computer from a remote terminal, the searcher first specifies the desired search level: Dictionary, Index, or Data. The dictionary level here is similar to the neighbor function that is in several data systems. Namely, the searcher can get an idea of the contents of the data base by questioning what the file contains. Here, I've asked to see a listing of all occurrences of the strains of bacteria derived from Salmonella typhimurium. The 35 types are listed, as is the frequency of their occurrence, and an indication of how they are used in the system. The search question is formulated at the index level and the number of hits can be listed by the system. Relevant hits are flagged by the computer and several queries can be made and flagged in succession. The searcher can then proceed to the data level where any or all of the flagged hits can be displayed. The searcher has the option of listing the reference number only (i.e., number referring to the hard copy of the hit) or listing the entire entry. For example, here I've asked for references to MNNG when it is used to induce frameshift or base substitution changes in Salmonella typhimurium. There are 4 hits in the system, and I have had one of them printed out fully. This system is well suited to our retrieval needs for several reasons. Complex search strategies can readily be accommodated by the use of Boolean logic and the logical operators, and, or, and not. The searchers can proceed in a series of logical steps, sampling the results of one step before proceeding to the next. In this way, it is possible to minimize the number of non-relevant hits by either pursuing a successful strategy or abandoning a poor one. The inclusion of a dictionary level permits a preliminary examination of the contents as well as the mechanisms of mutagenesis. Since the system actually reads a small number of entries in searching for relevant hits, the cost per search is low. On the other hand, updating the random access system is more costly than the sequentially stored system, since each new record must be split up and the segments inserted appropriately. In our case, the efficiency, accuracy and low search cost offsets the cost of update.

MUTAGENS is part of larger retrieval system which is still under development. Ultimately the search nodes in MUTAGENS will be implemented so that dosage or treatment can be specifically associated with a compound. In addition, the variable length strings (WLN, comments) and the fixed field entries (dose schedule, treatment schedule) will be string-searchable. Ultimately, also, the same type of system will be used to handle the primarily batch system I discussed earlier, since it is becoming more desirable to associate that data in more complex combinations.

In summary, I think one can conclude that the choice and/or combination of batch or on-line modes of data retrieval depends on the needs and resources of the user. I hope to have demonstrated this in the last few minutes and will be happy to discuss these systems further in the discussion period.

4

# COMPARISON OF SYSTEM REQUIREMENTS FOR ON-LINE AND BATCH RETRIEVAL

Joseph G. Coyne

ABSTRACT

Coping with information retrieval demands has led librarians, information and discipline specialists to greater use of both batch and on-line services. A balance of the two services, plus some old-fashioned tools, is needed to match demands to available resources. Over 5,000 citations are catalogued and indexed daily, and most of these eventually become accessible by computer. With this abundance of material and increasing number of users, many searches are repeated. Consideration should be given to publishing searches for which there is a high market demand, perhaps through regional services that publicize the intent to make a given search.

Coping with information retrieval demands has led librarians, information and discipline specialists to greater use of batch and on-line services. A balance of the two services, coupled with some old-fashioned tools, may be needed if the demands of today are to match the resources available. Can users efficiently merge multiple bibliographic data bases without assistance from suppliers: What should the data base output form be? Can users' needs be aggregated to make the output a more efficient function of the computer?

I would like to briefly explore these thoughts as we prepare to open for full panel discussion.

Just considering the membership of the National Federation of Abstracting and Indexing Services, over 5,000 citations are cataloged and indexed daily according to the Services 1975 membership report. I don't know what the daily figure for the universe is, but 5,000 per day is a staggering figure, and most of those citations end up in someone's machine accessible bibliographic data base sooner or later. Over 40 bibliographic data bases were cited in the Association of Scientific Information Dissemination Center's 1972 "Survey of Information Center Services" as publicly available and used for retrospective searching. Thirty-seven were cited as used for selective dissemination of information purposes. I'm not sure all the system innovations in on-line and batch retrieval techniques we can think of will help us find the resources needed to access the millions and billions of bits of information stored in computers today.

I would like to quote from Bill Knox' "Pathology of Information".[1]

> "Our laws and practices assume that there is a scarcity of information and that a good society results from a maximum flow of information of all kinds. But we have moved, in a generation, into an era when the average citizen suffers, not from a scarcity, but from an over-abundance of information."

Is it possible to harness this over-abundance? For example, how many searches are being asked of bibliographic data bases every day? Whether batch or on-line techniques are used, it seems to make no difference for in a recent progress report on a study being conducted for NSF, results were reported showing on-line searching equally cost effective as batch searching. Another estimate is that 1.5 million terminals will be operating in the United States in 1980. So, shouldn't we start asking ourselves some questions? How many times are the same searches repeated in the thousands of terminals now linked to bibliographic data bases? You may say, so what? Redundancy in this business is a way of life, and we must accept the fact that our data bases overlap, and our questions may duplicate one another. But at what cost? When labor costs, computer costs, terminals, paper, postage, and everything else is considered, what price can users really afford to pay for a bibliographic search? And think of the added cost when the same question asked of one data base must be placed against other data bases to insure comprehensive coverage for the searcher. In your

[1] "The Pathology of Information", William T. Knox Publ. in Book Prod. Industry, June 1971.

5

own experience, how many data bases must be searched to give you reasonable satisfaction--Three? Five? Ten? To formulate and retrieve the answer to a search, charges of $50 - $200 were common back in the 1972 ASIDIC Survey. Indeed, several ranged above the $1,000 mark.

Is it possible to turn this question around and ask -- Is there a better way to take advantage of the Batch and On-line technology available today?

Can we as information scientists think of ways to anticipate our users' needs?

If we can better anticipate and provide market aggregation of select searches, what would happen to unit costs per individual satisfied?

Does it make sense to think of a balance between published bibliographies, where market aggregation is possible, and on-line or batch services, where market aggregation is not possible, that will help us to conserve scarce resources?

NUMERIC DATA-BASES

Let's extend the problem from bibliographic to statistical or numeric data bases stored in computer. Let's use as an example, a request to pull from a statistical data base, an analysis of geographic areas in the United States where the characteristics of the inhabitants make the areas candidates for a proposed consumer product. To develop the strategy, pull the data off the file and amortize the cost of loading and maintaining the data may cost as much as $1,200.00. Since retrieval of information from statistical data bases often takes place off-line, how about taking a different look at market aggregation possibilities. If ten people needed the same information contained in our example, look what happens to the cost. It is reduced from $1,200.00 for one user to $120.00 per user.

Enough on the problems. How could such market aggregation take place?

Perhaps through a regional service, notice of intent to retrieve data in response to a specific need could be stated in a newsletter or other mailing. The notice of intent would represent for those having comparable or similar data needs, an invitation to join in the data retrieval costs for the search (which would be estimated). Or, perhaps, a more timely method would link all patrons of a certain statistical data base through a conferencing network. This would quickly enable the patron to vote "Yes" or "No" on sharing the front-end cost of a proposed interrogation.

A suspense date or cutoff for RSVPing would be cited, and the data would be developed with an ad hoc consortium. If no one responded, the initiator would pay the full cost (or not proceed with the interrogation). If there were five respondents, the cost would be shared in fifths, 10 respondents in 10ths, and so on.

Now, in conclusion, I'm sure many of you have thought about this problem and probably have better ideas along this line. Most important is that we continue to think and work together solving retrieval needs, so that we do not become overwhelmed by the mass of information that is now accessible by machine.

A COMPARISON OF COSTS BETWEEN ON-LINE AND FAST-BATCH SEARCHING

Daniel U. Wilde

ABSTRACT

The presence of a dedicated computer solely for information retrieval at NERAC creates a unique situation with regard to turnaround time. Therefore, NERAC operates a fast-batch system. Delivery of results to requestors is generally faster than that of material searched on-line and printed off-line as is the normal case for most current "on-line" searches. Costs for the number of searches done at NERAC in a year, if done at the average cost of current on-line searches, would be close to $600,000.00. This compares to a total cost for fast-batch of $190,000, which includes all equipment, file, personnel, and overhead expenses.

BACKGROUND

The New England Research Application Center (NERAC) was established at the University of Connecticut in 1966 as part of NASA's Technology Utilization Program. The purpose of the Center is to help business and industry gain access to appropriate technical and business information. NERAC attempts to take technology that was invented at one location and tries to put it to work at another. In other words, NERAC tries to help business and industry benefit by using someone else's technology.

It is well accepted that technology transfer is best accomplished by human being to human being communication. Consequently, NERAC uses technical specialists as interfaces between its users and its data sources. All of NERAC's technical information specialists have graduate degrees and years of industrial experience in their various subject fields. Here, users can themselves be information specialists at participating industrial firms or at other technical information centers. They can be company engineers or corporate officers who are the actual end users of NERAC's information.

Perhaps the singular difference between NERAC and other Information Analysis Centers is that NERAC has its own computer dedicated to serving just the participants of its Technology Utilization program. Presently, NERAC is searching 18 different abstract files on this inhouse machine. These include all major files, such as Chemical Abstracts CONDENSATES, Engineering Index, Inspec, SPIN, etc. and the complete NASA STAR and IAA files.

By having its own machine dedicated to just its participants, NERAC's technical information specialists need not worry about turnaround time.[1] NERAC's computer staff operates its machine three shifts a day, seven days a week and can search all files in a 24-hour period, if and when the need arises. NERAC's technical information specialists do not complain about turnaround time. Consequently, NERAC's mode of operation might be called a "fast batch" system.

COST CONSIDERATIONS

It is very difficult to compare online versus batch because no center has yet run both types of operations in parallel long enough to generate complete and valid data. However, the staff of NC/STRC has done perhaps the "best" job and has actually run comparisons using the same people on the same problems using parallel systems.[2] Nevertheless, it is possible to make some comparisons of NERAC's fast batch system with various online systems.

In a typical month during mid-1975, NERAC's computer processed an average of 1,000 retrospective file searches and 4,000 current awareness file searches. Here, a retrospective file search is defined as a search of one complete file run against one update issue. Thus, an Engineering Index profile would be run once a month while a Chemical Abstracts CONDENSATES profile would be run four or five times per month.

For online expenses, it is necessary to make some assumptions because the literature only indicates costs for very quick lookup searches without exhaustive file examinations. Consequently, online costs are understated. Nevertheless, an estimated price for an online retrospective search including citation and abstract printing costs can be expected to exceed $30.00 for a complete and thorough examination of a typical file.[3] Likewise, a cost of $5.00 per current awareness is reasonable. Using the above search loads and prices, it would cost NERAC close to $600,000 per year to perform its searching via an outside online service.

In contract, for 1975, NERAC's total "fast batch" costs will be approximately $190,000. This figure includes the following:

- Computer rent
- Computer utilities
- Paper, ribbons and supplies
- Data bases
- Data base royalties
- Full-time programmer
- Full-time 3rd shift operator
- Complete overhead, fringe and burden.

Thus, NERAC's fast-batch operation costs approximately one third an external online system. Furthermore, NERAC's system has recently been upgraded and still has unused search capacity.

## SUMMARY

NERAC is organized to help its participants. When users telephone, our technical staff is available to answer their questions immediately. Callers do not have to wait because NERAC's staff are at terminals with interactive searches in progress. Participants are never put on hold or told to call back. NERAC tries to never keep users waiting.

As a result, NERAC time-shares its technical information specialists. They wait for user calls while they are reviewing and evaluating searches that were run the previous night and while they are designing strategies for the next night. When a user calls, the appropriate technical specialist is ready at his or her desk available to dialogue the request. Once the user finished the call, the user is free to go on about his more important work while NERAC's experienced technical specialists do the search work.

Finally, NERAC maximizes the use of its computer. The computer is a tool that should be used whenever possible to help participants. In contrast, if NERAC was buying service online, its costs would go up as more searches were run. Increasing costs would inhibit greater usage.

NERAC does everything possible to encourage its users to submit many questions. Because NERAC's search costs are fixed, it can pass the savings onto its participants. This makes it easier for them to ask more questions. And, when a user asks a lot of questions, they are sure to benefit from NERAC's service.[4]

## REFERENCES

1. Wilde, Daniel U., "Using a Small/Low-Cost Computer in an Information Center," Frank Slater, Ed., American Society for Information Science, June, 1973, pp. 141-149.

2. Smetana, Fredrick O., et al., "A Study of the Relative Effectiveness and Costs of Computerized Information Retrieval in the Iterative Mode," North Carolina Science and Technology Research Center, Report STR-505, November, 1974.

3. Barbara, Lawrence, et al., "Making On-Line Search Available in an Industrial Research Environment," Journal of the American Society for Information Science, November-December, 1975, 00. 364-369.

4. Wilde, Daniel U., "Maximizing User Benefit from a Technical Information Center," Conference on The Problem of User Benefit in Scientific and Technological Information Transfer, AGARD-NATO, Copenhagen, Oct. 8-9, 1975.

ASIS PANEL - "SYSTEM REQUIREMENTS FOR ON-LINE VS. BATCH"

Peter J. Chenery

ABSTRACT

While "on-line" frequently is thought of as "instantaneous" and "batch" is considered synonymous with "delayed", the system at NC/STRC does not fit these concepts. On-line systems used with off-line printing (the typical situation) often engender delays measured in days. And batch searching of a computer in the next room can provide very fast, if not instantaneous, results.

A comparison of the same questions against the same data bases was made by NC/STRC staff. Quality of results was found to be a function of staff member capability and time spent formulating the query. For a given searcher, quality was independent of system mode. Search costs are higher for an on-line system, but the disparity decreases as searches get larger. For a small search, on-line cost is 1.6 times as much as batch, while for a large search, it was only 1.2 times as much.

Like NERAC, NC/STRC is a NASA-supported, industry-oriented information center. Like Searle, we use both batch and on-line search systems.

In planning my comments for this panel, I decided that, regardless of whether I was to speak first or last, it would be useful to give you my definitions of the terms "on-line" and "batch". These are polarized words which evoke unstated assumptions when we hear them.

(Figure 1) Thus, "On-Line" means a system in which the user's terminal is connected by a communications link to a (usually distant) computer; a system which accepts data and commands from the user and responds to the user "immediately", as if the user were engaged in a conversation with the system.

The nature of the response can vary widely, form simple acknowledgement that a job has been received for later processing, to the examination of indexes and files in order to develop a search strategy, to the actual search itself, with the user able to display or print search output at his terminal.

(Figure 2) "Batch" is perceived by many as the return to yesterday's computer systems, with data and commands being made machine-readable by keypunching, the cards transported to a computer center where they are grouped with other similar jobs to be processed against a set of files at one time. Thus, "batch" means, to most people, long delays while questions are accumulated for economy of processing.

(Figure 3) Let me point out that, while an on-line search usually provides some assurance that the right question has been asked, and that answers will be relevant, the complete set of answers is ordinarily delayed for days by printing them at the computer center, and mailing them to the user's location.

In a few cases, where a "quick fix" is sought, and any relevant information will solve the problem, the delay for mailing the complete output is of no consequent.

(Figure 4) What about systems that fall between the extremes of batch and on-line as I have defined them? One such system is in operation at our center in North Carolina. My system diagram looks very much like the on-line system I showed first; in fact, the diagram is identical. What are the differences between systems? The first important difference is in the amount of interactive dialogue between the system and the user. Instead of referring to computer-stored indexes and thesauri displayed at the terminal to construct a search strategy, our analyst uses microform indexes and printed thesauri at his desk. Instead of browsing intermediate results displayed on a terminal, our people browse their results in the printed journals. The final search strategy is entered interactively via terminal, with the computer response limited to checking for validity and completeness in the strategy.

The second important difference between our system and the typical batch system is in our use of search files mounted on direct access storage devices. This makes it economic to run single searches on demand, since the specific parts of the file needed can be examined without searching the entire file.

(Figure 5) While our system can provide answers in minutes, the computer center we use has an incentive pricing system which places a premium on quick responses. Our normal time for search output is overnight, but we can get results in an hour or two if we need them. And our results are complete when received.

Dan Wilde has referred to a study we made about a year and a half ago; some of you may have seen reports on it, by Fred Smetana of NCSU. What we did was this: we used our staff of information specialists and a few less-skilled people (like me) to search the NASA data base in two ways. The first method was to use our batch search system in the way I have just described. The second way was to make an on-line search of the identical NASA data base using a terminal connected to the NASA computer then in College Park, MD. We arranged for each person to search the same question using both systems. To minimize the effects of learning from the first search on the quality of the second search, we alternated the system which was used first, and we withheld the full search output until both searches had been made.

What were the results? We expected that the analyst at his desk would spend less time in developing his strategy than the person at the terminal, because of system delays in responding, poor typing skills, and so on. This did not occur. The total time spent varied widely, from person to person, but most people spent about the same time per question at the desk or at the terminal.

Our evaluation of search quality showed that, regardless of system used, the more time spent on the strategy, the better the search results. In general, both searches by an individual were equal in quality.

As for search costs, we applied commercial rates to our use of the NASA Recon on-line system and found that our on-line searches cost from two to three times as much as the batch searches. I have said that labor was about the same for both systems; the big difference was in computer costs. Input data was keypunched for these batch searches; the on-line system was then on a 360/50 machine with relatively slow response due to high user loading.

Within the last month we made test searches of the ERIC data base using Lockheed's DIALOG system and our system with interactive terminal entry of strategy. (Figures 6 & 7) Here the cost difference is less. Our terminal entry system was designed for student use, and is relatively slow because of the tutorial features included. The Lockheed system has be optimized for quick response, and our analyst is skilled in its use (Figure 8). The cost basis for this comparison is shown here.

In conclusion, what are the trends, and what improvements are needed to make mechanized searching more effective at lower cost? It is clear that the return of search output by mail from a distant computer center is the principal source of delay for on-line searches now. The next upgrading of the computer networks, such as Tymshare's Tymnet, through which many users communicate with on-line services, might well include the location of high-speed printers at most of the network nodes; this would be similar to the Mailgram service now offered by Western Union and the Postal Service.

A belief which I hold is that much of the demand for searches comes from small and medium-sized organizations which cannot support or keep busy a skilled searcher. After the initial glamour of "I can do it myself at my own terminal" passes, many of these searches will be given to intermediaries at local and regional search centers. The computer systems used by these centers will be designed to meet economic constraints which are now, for the largest part, ignored by users who are fascinated with the ability to talk to the computer.

Note: The cost of 48¢ per minute for communications shown in Figure 8 is for a direct-dialed call from Research Triangle Park to Palo Alto. Now that a Tymsat has been located in the Park, that cost drops to about 17¢ per minute.

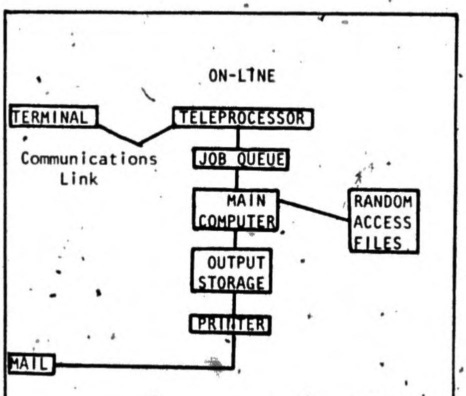


FIGURE 1

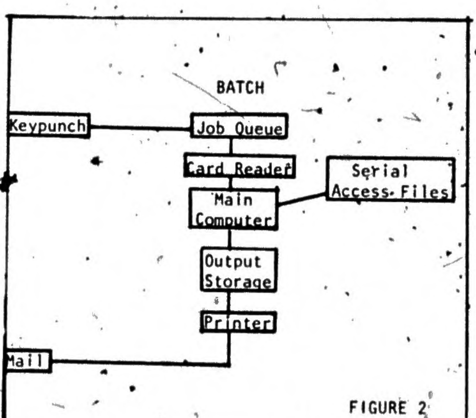


FIGURE 2

| OPERATION | ON-LINE | BATCH |
|---|---|---|
| Strategy Developed | At Terminal | At Desk |
| Results Browsed | At Terminal | In Journal |
| Output Size | Given by Computer | Estimated |
| Strategy Changes | During Search | After Search Returned |
| Final Output | Mailed from Comp. Ctr. | Mailed from Comp Ctr. |

FIGURE 3

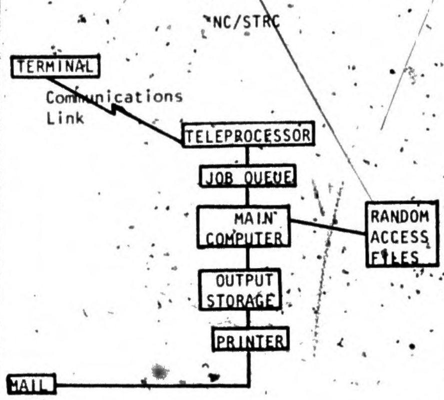


FIGURE 4

| | ON-LINE SERVICE | NC/STRC |
|---|---|---|
| Interactive: | Strategy Development | Strategy Entry |
| Response Time: | | |
| Output Size | Minutes | Hours |
| Sample Hits | Minutes | Hours |
| Full Output | Days | Hours |

FIGURE 5

SEARCH COSTS

Small Search: 6 terms, 1 Intersection, 200 hits

| | Labor | Computer | Total |
|---|---|---|---|
| ON-LINE | $5.83 | $35.51 | $41.39 |
| NC/STRC | 6.76 | 18.88 | 25.64 |

FIGURE 6

SEARCH COSTS

Large Search: 23 terms, 2 Intersections, 500 hits

| | Labor | Computer | Total |
|---|---|---|---|
| ON-LINE | $10.26 | $70.63 | $80.89 |
| NC/STRC | 7.13 | 58.99 | 66.30 |

FIGURE 7

SEARCH COSTS

| | |
|---|---|
| Engineering Labor | $12.50/hour |
| Clerical Labor | 3.50/hour |
| On-Line Computer Charges | 25.00/hour + 10¢/item printed |
| Communication Charges | 48¢/minute |

FIGURE 8