DOCUMENT RESUME

ED 043 348 LI 002 139

TITLE	Project Intrex. Semiannual Activity Report, 15 March 1970 to 15 September 1970.
INSTITUTION	Massachusetts Inst. of Tech., Cambridge.
SPONS AGENCY	Carnegie Corp. of New York, N.Y.; Council on Library Resources, Inc., Washington, D.C.; National Science Foundation, Washington, D.C.
REPORT NO	PR-10
PUB DATE	15 Sep 70
NOTE	107p.
EDRS PRICE	EDRS Price MF-$0.50 HC-$5.45
DESCRIPTORS	Automation, *Cataloging, Computer Programs, *Electronic Data Processing, Information Centers, *Information Retrieval, Information Services, Information Storage, *Information Systems, Libraries, Reports, *Use Studies
IDENTIFIERS	Library Automation, *Project Intrex

ABSTRACT

User experiments are the main theme of this tenth issue in the series of Project Intrex semiannual activity reports. This system is undergoing tests with users who are motivated by their need for information rather than an interest in the machinery by which it is stored, retrieved and presented. Test results will indicate which fields in the augmented catalog are most helpful to the user, which of the retrieval program features are most effective and which display techniques are most congenial in continued use. This kind of information will make it possible to design the future information systems in which libraries and information centers will utilize the new computer-communications technology. The seven major sections of the report are: (1) introduction, (2) research and development activities, (3) model library, (4) Project Intrex staff, (5) current publications, (6) past publications 1969-1970, and (7) past publications 1966-1969. The research and development activities section covers: (1) status of the program, (2) user experiments, (3) augmented-catalog inputting, (4) storage and retrieval, (5) display consoles, and (6) full-text storage and retrieval. The model library section includes: (1) status of the project, (2) point-of-use instruction, and (3) library pathfinders. (NH)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT INTREX

SEMIANNUAL ACTIVITY REPORT

15 March 1970 to 15 September 1970

PR-10

15 September 1970

CAMBRIDGE MASSACHUSETTS

ACKNOWLEDGMENTS

The research reported in this document was made possible through the support extended the Massachusetts Institute of Technology, Project Intrex, under grants from the Carnegie Corporation, the Council on Library Resources, Inc., and the National Science Foundation.

TABLE OF CONTENTS

PROJECT INTREX

Activity Report

I. INTRODUCTION

In a period of shrinking research budgets, it is gratifying to release a report on expanding results. The steadfast support of our sponsors during five years has enabled Project Intrex to establish the technical feasibility of a new system of providing access to information resources. This system is now undergoing tests with an increasing number of users who are motivated by their substantive need for information rather than any interest in the machinery by which it is stored, retrieved, and presented. The results of these tests are the essential output of Project Intrex. They will tell us which fields in the augmented catalog are most helpful to the user, which features of the retrieval programs are most effective for him, which display techniques are most congenial in continued use. Only when this kind of information is available from statistically significant experiments will it be possible to design the future information systems in which libraries and information centers will utilize the new computer-communications technology.

User experiments, then, are the main theme of this tenth issue in the series of Project Intrex semiannual activity reports. This theme will be encountered not only in the sections that describe the augmented-catalog and text-access activities, but also in the section on the more recently initiated Model-Library Project, where the objective is to design user aids for a library in transition between old and new technologies.

The Computer Science and Engineering Board of the National Academy of Sciences has established an Information Systems Panel under the chairmanship of Ronald L. Wigington of Chemical Abstracts Service. The members serving with Mr. Wigington are:

F. T. Baker	IBM Corporation
Joseph Eachus	Honeywell
Douglas Engelbart	Stanford Research Institute
Gerald Salton	Cornell University
James E. Skipper	University of California

The Panel's assignment is to determine to what extent the development of national information systems and networks is held back by technological limitations, both in hardware and software, and to survey what is being done to overcome such limitations. Following a visit to Project Intrex on 25 March 1970, the Panel asked for written comments on a number of questions related to the possible broad-scale use of the concepts and techniques involved in our experiments. We regard this exchange as the beginning of what we hope will be a continuing and fruitful interaction with the Panel.

Carl F. J. Overhage
Cambridge, Massachusetts

15 September 1970

II. RESEARCH AND DEVELOPMENT ACTIVITIES (Electronic Systems Laboratory)

A. STATUS OF THE PROGRAM

Professor J. F. Reintjes

Experimentation and analysis constituted the major parts of our efforts during the preceding six months. Controlled experiments with invited users and noncontrolled use of the Intrex system in the Baker Engineering Library are providing a substantial body of data for making analytical studies of system operation and performance.

Experiments and analyses centered on the computer-stored catalog of documents, the number of which now exceed 12,000. Our objectives with respect to the catalog remain unchanged: we wish to determine the proper content and scope of a catalog of documents when an interactive computing system is available as an aid in the search process. By "proper" we mean the amount and kinds of catalog information that will satisfy a user community's information needs in terms of completeness, relevance and convenience, and at a cost that can be justified.

We recognize the possibility that two types of catalogs which differ substantially in content may emerge from our studies. One type, containing a substantial amount of information per entry, would be of greater value when catalog information only is readily available at user display terminals; the other type, which could turn out to contain much less information per entry, would be employed when the full-text of documents, as well as catalog information, is provided at the user terminals. The Intrex system is designed for tests under both conditions.

Several observations can be made about our experimental program thus far. One is that each console session with a user provides us with a flood of information which, through subsequent analyses, has potential for contributing to our understanding of the effectiveness of the system. We are therefore being challenged to ask the "right questions" so as to narrow the choices among the many diverse analyses that are possible. Another observation is that it is possible to test the same basic features of the catalog in several ways. Thus, we have an opportunity to make cross checks before drawing firm conclusions. Substantial thought is being put into the design of experiments and their scope in order to minimize statistical roughness of data which may be attributable to looseness of experimental procedures.

In support of the experimental and analytical program, continued contributions are being made by the data-bank inputting group and by the software and hardware groups. The inputting group has selected nearly 15,000 documents for inclusion in the system. Thus, our information source continues to be up-to-date and hence relevant to user-community needs. Additional features have been included in the retrieval programs by the software group in an effort to simplify use of the system and to

improve the clarity of output presentations. For example, it is now possible to order the full text of a list of documents at the display terminals simply by typing the command "output text." Formerly the microfiche number and the page location or the fiche had to be entered manually for each document being requested. A new software package has also been added which presents in conventional form superscripts, subscripts and Greek characters.

The hardware group has completed the modifications of an ARDS Console so that catalog and full text can be displayed on the same cathode-ray-tube screen. This feature has proven to be so appealing that the second ARDS Console is being similarly modified. An inexpensive hard-copy capability has been added to the Intrex Display Console, and a hard-copy device for ARDS terminals, originally developed but not finalized by another Laboratory group, has been turned over to us for possible application in the Intrex System.

B. USER EXPERIMENTS

Staff Members

Mr. A. R. Benenfeld
Mrs. S. Brown
Professor L. S. Bryant
Miss M. A. Jackson
Mr. P. Kugel
Miss L. T. Lee
Mr. R. S. Marcus
Miss V. A. Miethe
Professor J. F. Reintjes
Miss L. Rossin

Consultant

Dr. P. W. Holland (Psychology and Statistics)

SUMMARY

In the area of user experiments this reporting period witnessed the beginning of major use of the Intrex facility in the Engineering Library in an essentially "open" environment, as well as the continuation of our controlled experiments in the three categories previously described.

During the summer we have had the benefit of the experience of a psychologist, Mrs. Susan Brown, in the conduct and analysis of our controlled experiments. Dr. Paul W. Holland of the Harvard University Department of Statistics consulted briefly with us toward the end of the spring academic term. He will rejoin our group in the fall as a consultant in matters of experiment design. Professor L. S. Bryant of the Humanities Department, M.I.T., continued to participate with us on matters pertaining to user aids.

The Intrex facility in the Engineering Library has been operating on a regular 3-hour-a-day basis since March. No major problems in system operation have been encountered. By August 1, 1970, over 350 individual uses of this facility have been made. The typical user's response was overwhelmingly enthusiastic. The main complaint is the lack of coverage of the data base within the user's own particular area of interest; Intrex, by intent, covers only selected areas of materials science and materials engineering. While most of the uses have been for demonstration or to satisfy curiosity, we have identified for analysis purposes a set of 23 sessions involving serious use of the system. This analysis is continuing and has so far concentrated on a statistical summary of the use of the different catalog fields and the comparative effectiveness of different ranges of subject-index terms in retrieving relevant documents—relevancy being projected on the basis of user request for full text. A major effort involved in the setting up of the library facility was the training of librarians to aid users. In both the training sessions and the actual user sessions we obtained many valuable comments and observations on the effectiveness of particular system features.

Several additional experimental subjects have participated in our controlled experiments — most notably five new subjects for Category II — but the prime emphasis in this period has been on the further analysis of experiments already run.

Further analysis of Category-I results has begun to suggest the magnitude of the relative improvement of Intrex retrieval over abstract-journal retrieval, at least for some simple one-word searches. Similarly, additional compilation of Category-II data has begun to quantify our estimation of the relative utility of several of the catalog fields for indicating the ultimate relevance of a document to the user's needs. We have continued to consider the problem of refinement of our experimental techniques and the need to establish in detail whether we are asking the right questions in our experimental designs.

INTREX FACILITY IN THE BARKER ENGINEERING LIBRARY

System Configuration and Usage. The Intrex system became available to the M.I.T. community with the opening of the Barker Engineering Library at M.I.T. on March 5, 1970. Access to the catalog and text-access systems was made available daily for regularly scheduled three-hour periods. No charge was made for use of the system — except 10 cents per page for full-size hard copy of full text — and users were accommodated on a first-come first-served basis. By August 1, 1970, more than 350 persons had used the system at the Engineering Library.

The initial configuration of the Intrex retrieval facility in the Engineering Library (see Figs. B-1 and B-2) includes an ARDS 101A display console for access to the augmented-catalog information in CTSS, the time-sharing computer system and a stand-alone text-access console for display of full text transmitted approximately 1/3 mile over coaxial cable from the remote text-storage station located in the Electronic Systems Laboratory. An IBM 2741 typewriter console is also available, on request, or as a backup device, for access to the catalog information and on which hard copy of catalog material may be obtained. Hard copy of catalog information may also be obtained by requesting Intrex personnel to order off-line computer printouts on an overnight basis. Users have tended to prefer the latter facility in spite of the delay, probably because it requires less effort on their part.

Hard copy of the full text of documents is obtainable through the text-access automatic film station. However, instead of relying on this facility we have so far concentrated on the use of the duplicate fiche collection, located in the microreproduction room in the Engineering Library; this room contains facilities to provide either microfiche output or hard copy. Of course, users may also locate the original documents in the library stacks if they prefer to do so.

Instruction and Monitoring. Librarians specially trained in the use of Intrex are available at the consoles to assist users in the event that they encounter difficulties and to give demonstrations of the system to visitors. A variety of printed material is available for users at the station including the Intrex Guide, bibliographies generated as a result of Category-II experiments, an introduction to the system, an

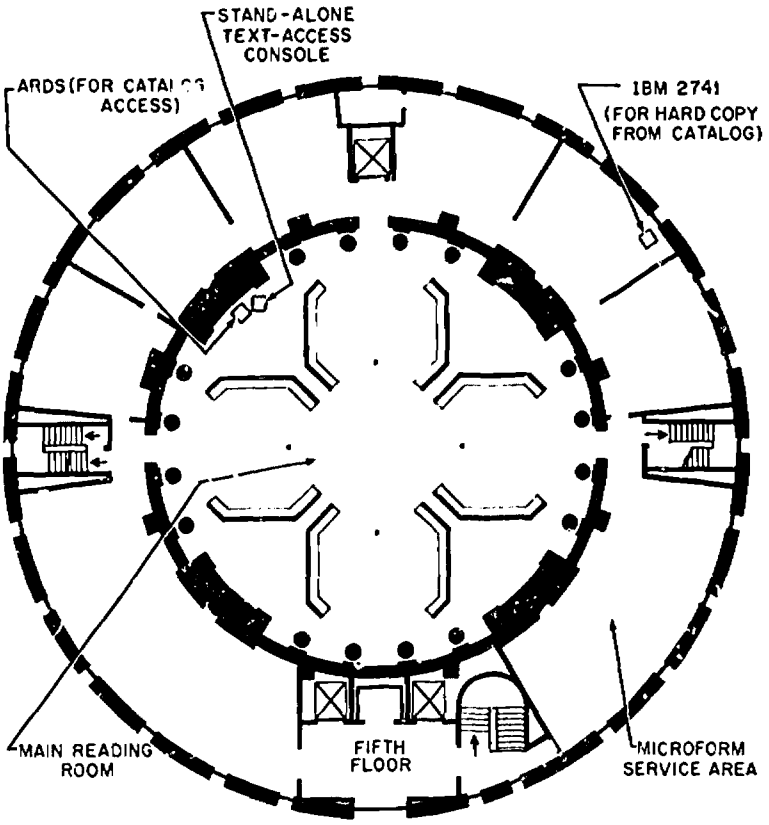


Fig. B-1 Intrex Facilities at Barker, the Engineering Library

Fig. B-2 Photo of Display Consoles in the Barker Engineering Library

overview of the system and other aids. Experiments are being performed to determine the effectiveness of the printed material in absence of a librarian.

The librarian, in addition to his (or her) role as system instructor, observes user behavior and records information not otherwise captured in the monitoring process. The librarian also encourages the users to make comments about the system via the Intrex comment command and in a users' comment book that is available at the station.

The catalog system automatically monitors all user-system interactions that take place. These are stored on disc files and printed out daily. The resulting printouts constitute the primary documentary record of system utilization on which analysis is based. Monitor files record user requests and system responses as they are made, as well as data concerning the timing and utilization of system modules.

Users' Responses. The over-all response of a typical user is very enthusiastic. The speed and relative ease of access to information compared to traditional library procedures seem to impress the users. The user's main complaint is the lack of coverage of the data base within his own particular area of interest. Of course, of the hundreds of users of the system who, for the most part, had come to use the Engineering Library and not Intrex, only a few had problems falling largely within the confines of our specially selected data base. For those users who did make serious use of the system, a start has been made in analyzing just how well various features of Intrex are suited to their problems. This analysis is described in the second section below. Some general observations on system operation follow in this section. A more comprehensive analysis of system operation and effectiveness is under way.

Analyses of the monitor files, and of the users' and librarians' comments have enabled us to detect and correct errors in the catalog data, in the text-storage system and the catalog software. Most of the errors that have been detected have been minor and easily corrected. The quality of the inputting process, as reflected in the low error rate of the data appears extremely high. Users appear to find the system easy to use — at least, with initial librarian help — and when system errors do occur, the recovery procedures appear to operate satisfactorily in the majority of cases.

A study of the timing data and module utilization data has enabled us to identify those portions of the software that would most reward efforts to improve their efficiency. Parts of the system thus identified are being improved to increase system speed and lower cost.

Users have encountered some difficulties because of the separation of catalog fields (for example, the excerpt and abstract fields) that might better be combined, and in understanding abbreviations, particularly in the use of coden titles in the analytic-citation field.

A common, though by no means universal complaint, was the difficulty of reading the screens on the display consoles. Of course, we were expecting some problems here since we know — as explained in previous reports — that the storage tube we are using for full-text display does not have adequate resolution to fully reproduce the original document. Also, the ARDS console tube does not have the clarity nor brightness of the Intrex-designed console which we will eventually be using. However, two factors tend to diminish the force of these complaints. In the first place, controls which affect focusing can be adjusted by the user and, in some cases, at least, it was observed that user adjustments had caused deterioration of display clarity. In the second place, many users relied on the librarian to do the typing and did not actively operate the console themselves. It was noted that active users of the system tended to sit closer to the display screen and to be less bothered by display inadequacies.

Another common complaint with the catalog console was the inability to "go back" in order to see information that had been erased from the screen. This problem, too, would be reduced by the availability of the Intrex-designed catalog console, which has a page-storage feature.

The consoles are located in the main reading room under ordinary overhead lighting and without any additional soundproofing. (See Figs. B-1 and B-2.) The system appears to satisfy most users under these conditions, although some users have found that the lighting conditions make the displays difficult for them to read because of reflections. Others have found the general noise level in the library disturbing. Conversely, some users at nearby study desks of the regular library have found the noise generated by the users of the Intrex console disturbing to them. Intrex users sometimes get quite excited by the results of a search, perhaps because of the novelty of the system. One can expect familiarity to decrease this effect. On the other hand, it may also be noted that many users of nearby study desks seemed not to be disturbed at all by the activity at the Intrex consoles.

INTREX LIBRARIAN-TRAINING PROGRAM

As mentioned above, a program to train Engineering-Library staff and ESL staff indexers as advisers to operate the Intrex retrieval system was undertaken. It is worthwhile to review this program in some detail as an indication of the problems in teaching system use, in general, and in training librarian-advisers in particular.

Roles of the Adviser. The advisers' roles are varied. At one time or another the adviser must introduce the system to new users; instruct new users if the user does not elect a self-instructional mode; guide a user; suggest or promote system alternatives; get a new or experienced user out of a search difficulty or

system difficulty; observe and comment to the system upon otherwise unmonitored behavior of serious users; act as a searcher surrogate of a user. A successful adviser requires considerable practice in system use, more than can be obtained through a training program. However, deftness in system manipulation comes from a basic understanding of the system operations. A successful training program can impart understanding, present rules of procedure, and launch practice in system operation.

Outline of Training Course. Three sections of a training course were given, each with one library-staff member and one Laboratory-staff member participating as trainee-partners, and with one or two of the Laboratory's system designers serving as instructors. The trainee pairs were thus composed of one person who was familiar with traditional library reference-service techniques and one person who was familiar with the indexing and cataloging of the Intrex data base. Each course section was composed of a number of sessions extending over several weeks (see Table B-1).

Table B-1
Training-Course Duration

	Section 1	Section 2	Section 3
Number of Sessions	16	19	20
Average Duration per Session	1/2 to 1 day	2 hr/day	2 hr/day
Calendar Spread	3-1/2 weeks	4 weeks	5 weeks

The trend is to a larger number of meetings of about two hours' duration spread over a period of several weeks. Each session consisted of either a lecture presentation, active console sessions, practice examples, or review and discussions. Each course section covered the topics outlined below, but they varied in their order of presentation, depth of coverage, and technique of presentation.

1. An active introduction, consisting of an on-line retrieval session by the trainee and guided by the instructor. This session "breaks the ice", captures the trainee in a "naive-user" state which will later enable him to better appreciate a naive user's problem, and exposes him to the spectrum of topics to be considered in detail later. A major objective of the session is to dispel awe and fear of the computing machine and its peripheral equipment. This introductory session is accomplished in about one hour.

2. A discussion of the experimental objectives of Intrex within the context of general problems in bibliographic access and text access. This topic is a direct extension of the active introduction, and both topics together can be completed within two hours. Both topics were essentially invariant among the three training sections.

3. A detailed familiarization with the operating characteristics of the Intrex system. The bulk of the training program falls into this category and it can be considered conceptually to cover two major areas: operation and manipulation of Intrex as a retrieval tool, and the construction of the Intrex system. The particular topics encompassed by system operation include: terminal mechanics; log-in and log-out procedures; general dialog structure; command language and argument structure; search commands; output formats; searching levels; search statements and strategies; search aids; instructional aids; catalog-text access cycling; and copy requests. The particular topics encompassed by system construction include data-base literature selection and coverage; catalog-record structure; catalog fields; indexing; descriptive cataloging; special formats; data input and error-correction procedures; stored catalog-record and inverted-file format; meaning of time-sharing computers; machine representation of data; computer operation; output devices and data-output formats; text-storage format; text transmission; and text-output devices. It is in the handling of the two major areas of operation and construction that the training program most differed from section to section.

4. A discussion session concerned with instructing and observing library users of the Intrex system. This seminar revolved around Intrex adviser-user dialog and user behavior before, during, and after a user's console session. These topics were related to the Intrex objective to design a retrieval system. Categories of potential points to observe in a user's behavior, some methods for soliciting a user's opinions and comments, and methods for reporting these items to system analysts were discussed.

5. A final system demonstration. Each training program concluded with each trainee demonstrating the Intrex system to a new user, with the instructor playing the role of the new user.

Information and guidance on system changes occurring after the training program are, of course, provided to each adviser. The full complement of Boolean-search capabilities is one major topic that was covered in post-training meetings.

Training Strategies. In general, trainees appear to learn best by doing, that is, by participating in active console sessions that are followed by discussions of how and why the system works. This "learn-by-doing" technique was used more and more frequently in each successive training-course section. Nothing can replace hands-on experience, and trainees were given problems to be solved at a console. These problems were a combination of general real-user search problems and contrived problems designed to instruct the trainee on particular points about searching or about the catalog.

In each successive course section, the topics from the two major conceptual areas of operation and construction were intermixed more frequently. Emphasis in describing topics concerned with the construction of the Intrex system was increasingly placed on understanding of their roles in causing (or potentially causing) a system operational failure or a system search failure. This latter technique thus relates seemingly isolated or unimportant topics (from the searcher's viewpoint) to actual system use and does so in a manner better suited to the searcher's frames of reference. Most lectures increasingly incorporated a stress on the underlying structural and symmetrical properties of the system such as those of data format and search format. The first course section emphasized, more than the others, some practical training in indexing but this is apparently of value in the larger training program only insofar as it introduces an appreciation and understanding of the methodology of index-term derivation. Imbuing searchers with a fuller understanding of indexing subtleties requires a longer time base and a different view of the training program on the part of the trainee. Since one member of the trainee team is an experienced indexer, additional points about indexing can be made to the non-indexer trainee as the occasion arises during their practical console sessions.

It is interesting to note that each trainee-pair had a distinct preference to solve problems jointly, with the members of the pair alternating at the console keyboard rather than each individual alternately working complete problems.

Training Results and System Implications. Our training program was successful to the extent that it produced Intrex advisers whom system users find helpful. The performance of the advisers who have completed the program has enabled a large number of users to engage the system successfully.

In spite of the program's outward success, it is worth noting two general areas and several specific topics of difficulty. One such area lies in the large amount of time required for the completion of the program. The second lies in the fact that many parts of the system were not thoroughly understood by the advisers

even when the program was completed. The following specific topics seemed to cause the trainees most difficulty: (1) the proper use of the commands to create, name, and otherwise manipulate lists in a Boolean manner; (2) an understanding of the nature of the indexing and its consequences, particularly with respect to the need of the user to generate synonyms; (3) the proper use of the RESTRICT command and particularly the distinction between the precise string matching used in the RESTRICT command implementation on the one-hand and the somewhat looser matching used in the search command (SUBJECT, TITLE) implementation. These difficulties are, in larg part, due to the experimental and developmental nature of the Intrex system which leads to temporary inelegancies, to frequent changes in system operating procedures, and to the inclusion of a broader spectrum of alternatives than one would find in an operational system. This suggests, however, that what we might call "learnability" is an important property of the kind of system toward which Intrex aims. Among the major components of learnability are:

Stability	A system is most easily learned if its operational features are stable and predictable. At the time of our training program, the Intrex system was in a state of flux because of intensive preparations, fixes and changes preceding its installation in the Engineering Library. Daily changes and temporary "bugs" in the system hampered our initial training, particularly during the instructional program for our first trainee group.
Conceptual and Operational Simplicity	Operational simplicity from the user's viewpoint is closely allied to a harmonious consistency among system features. The less complicated features and those with the least number of variations to the operating rule are the easiest to remember. A possible current Intrex fault, for example, is the inconsistency between the rather rigid form required in the AUTHOR search requests (last name, comma, initials) compared with the free form allowed in SUBJECT search requests.
Familiarity with Conceptual and Operational Structures	From the instructional viewpoint, the task is easier if the learner has some previous familiarity with the concepts and structure underlying the system or if such facets can be made directly analogous to a familiar concept, for example, a comparison between an inverted file and a book index. The system must be adaptable to the user's frame of reference.

In our next training course, we hope to improve the instructions by taking cognizance of the above factors. We are also interested in studying a self-instructional approach to some of the topics.

ENGINEERING-LIBRARY USAGE ANALYSIS

Delineation of Users Analyzed. The data in the catalog system serve two basic functions — it is used by the system software to select documents for the user, and it is employed by the user to select documents that are of value to him from among those that the system retrieves for him.

Nineteen serious users of the system were identified in the records of the first two months' use, and the reactions of these users to the system have been studied to evaluate the utility of the catalog data with respect to these two functions. A total of 23 sessions was studied; two of the users came to the system twice, one made three separate uses of the system.

Serious users are defined as persons who are primarily interested in engaging the system to get information about subject matter that the system covers, in contrast to the others who are either trying out the system to satisfy their curiosity or who are primarily interested in the system itself rather than the information it contains.

In order to make it possible to identify such users in an objective (and repeatable) way, we have defined a serious user as one who

a. uses at least 1000 seconds of real time and 100 seconds of computer time in a session;

b. makes at least three requests that fall within the system's general coverage area and at least two that have a closely related objective;

c. finds at least one document whose full text he wishes to examine.

The last of these requirements is based on the nature of the experimental design, described below, which requires text requests for the evaluation of the subject indexing. The other two requirements were derived from an examination of the monitor-file records; their applicability in separating out serious users will be subject to continuing analysis.

Analysis of Catalog-Output Requested. At the 23 sessions studied, a total of 308 output requests were made for data in the catalog. This is an average of a little more than 13 requests per session, but the number of output requests made varies widely from user to user, and the median number of requests is only 11 per user session. The breakdown of output requests by field is given in Table B-2.

Table B-2

Number of Requests for Fields

Field Number	Field Name	Frequency of Requests	
		Number	Percentage
75	Normal (Title, Author, Citation)	75	24
71	Abstract	61	20
74	Matching Subject Terms	38	12
90	Text Address	27	9
24	Title (only)	24	8
73	Subject Terms (all)	16	5
70	Excerpt Made by Librarian	13	4
11	Library Location	10	3
21	Author (only)	9	3
(all)	All Catalog Fields	8	3
47	Analytic Citation (only)	8	3
76	Standard (Title, Author, Subject in Standard Citation Format)	5	2
38	Series Statement	4	1
37	Language of Abstract	2	1
-	(Detected Error)	2	1
1	Document Number	2	1
12	Serial Holdings	1	0.3
25	Coden Titles	1	0.3
36	Language of Document	1	0.3
40	Contract Statement	1	0.3

The system dialog suggests to the user the output form that we call the "normal output" and offers this to him as the easiest output to request. The "normal field" is a combination of the title, author and analytic citation (location) fields. Since this pre-planned combination of information constitutes both the easiest information to ask for and the system designers' choice of the kinds of information that the user is most likely to want to see, it is interesting to note that requests for this information constitute only 24 percent of the output requests of the users. This is particularly important in view of the fact that this packet contains most of the information included in standard bibliographic references, and hence is the information the user is most accustomed to seeing.

The separate information elements comprising this field can be requested individually, and such searches comprise the following percentages of all output requests:

TITLE:	8 percent
AUTHOR:	3 percent
ANALYTIC CITATION:	3 percent

In addition, users may request the information in these fields in a slightly different format which we call "standard format". This format comprises an additional two percent of the output requests. Putting these figures together, we find that 40 percent of the requests for information require only the information in these three fields. One problem in analyzing the data is to separate the utility of these fields from each other.

The most popular field after these three "normal" fields is the abstract field. Requests for the abstract comprise 20 percent of the requests for information from the catalog. The frequency with which other fields are requested is given in Table B-2.

The data shown in Table B-2 are based on a relatively small sample (19 users) and should not be considered conclusive. No corrections have been made for users who are requesting fields only to examine them and then decide that the field is of no value to them. The relatively low number of non-fields asked for — the two detected errors — suggests that typographical errors do not contribute significantly to most of the data, but their effect cannot be ignored in the case of fields that are requested only once or twice. Lack of familiarity with fields that are indigenous to Intrex also tends to favor the traditional fields. Further studies are planned to take these factors into account.

The matching-subject-terms field is the only field in Table B-2 that requires a computation at output time to generate it. The fact that it is used as much as it is, suggests that this kind of individualized tailoring of information may have general utility.

The output of a "count" request, which informs a user about the distribution of words in his search request over the inverted files, is another example of this kind of computational output; its utilization seems to be quite widespread. (It was not counted as a field.)

It is not, of course, sufficient to evaluate a field on its utility alone. In designing a library system, one aims toward maximizing the amount of gain one gets for the amount one invests in the system. This is done by selecting those features that yield the most effectiveness for their cost.

Determining the cost of a field is a somewhat difficult matter because of the multiplicity of factors that enter into such cost and their variability. The number of characters used to represent a field is a good indication of such cost for most fields

because it correlates so closely with both inputting and storage costs. Under this measure, the abstract and subject fields, although the most used nontraditional fields, yield less per character than some other fields which require so few characters for their representation that quite light utilization suffices to justify them from an economic point of view. Thus, for example, the language of the document requires five bits in the catalog record while the median length of the abstract field is approximately 750 characters, which, even with completely successful digram encoding, require approximately 650 times as many bits per catalog record. Therefore, if one is considering only storage costs, the abstract must be used more than six hundred times as often as the language-of-document field to justify its superiority to the latter field.

Analysis of Inverted-Field Searching and Subject-Term Utility. The average user made about ten search requests per session. Subject searches predominated. The data are summarized in the following table:

Type of Search	Number of Searches	Mean per Session
AUTHOR	20	0.9
TITLE	3	0.1
SUBJECT	206	9.0
	229	10.0

In studying the effectiveness of subject term searches, we made the assumption that when the user examined the text of a document, it was relevant to his subject area. Discussions with users tend to support the validity of this assumption. The question we would now like to answer is: What was the relative utility of different subject-term ranges* in retrieving relevant documents? Call a subject term a "hit" for a given range if a relevant document was retrieved by a search matching some term with that range. Note that, for a given range, a hit will occur in more than one subject term when more than one subject term of the same range number contains the search word(s). Under such a definition of "hit" we note that more hits were made by matching on range-2 terms than on any other single range. Range numbers 0 and 4 yielded the fewest. These results are shown in Table B-3.

* In the Intrex system a range number is attached to a subject-index term to indicate the amount of a document a subject-index term describes. A range-1 term characterizes all of the document; a range-2 term, a major part of the document; and a range-3 term, a minor part. A range-4 term signifies a piece of apparatus, a mathematical derivation, or special instrumentation. Range 0 is a term which places the subject matter in a more general category of a hierarchial structure.

Table B-3

Percentage Hits for Various Subject-Term Ranges

Range	Percent of Hits
0	3.1
1	26.6
2	29.1
3	17.4
4	10.7
5 (title)	14.1

It should be noted that these figures are not normalized with respect to the number of terms indexed for a given range and the number of words in these terms. The distribution of such terms and words is shown in Table B-4. (Sample is a result of 187 "relevant" documents retrieved by matching on 430 terms. This is the result of 206 searches.)

Table B-4

Percentages of Terms and Words by Range

Range	Percent of Terms Having Given Range	Percent of Words In Terms Having Given Range
0	2.5	0.9
1	11.4	13.8
2	29.5	36.8
3	38.4	35.2
4	7.6	6.2
5 (title)	10.5	7.1

If terms having a given range number constitute n-percent of the terms describing a given set of documents, one would expect terms with this range number to contribute roughly n-percent to the total retrieval of useful documents, all other things being equal. If terms with a given range number actually contributes a different amount, say m, the ratio m/n indicates how much more, or less, than expectation one finds. The higher this number (1 is the norm), the greater the contribution. The ratio m/n is described as the recall effectiveness normalized by the relative term frequency.

One factor that was not taken into account in the above analysis is the variation in the number of words in each term. Assuming, for the moment, that there is not much variation in the number of terms for each range, we may expect that the

number of documents retrieved by terms having a given range is proportional to the total number of words in terms of that range. In a manner analogous to that above with respect to term frequency, we may then define the recall effectiveness normalized by relative word frequency to be m/n' — where m is, as before, the actual percent retrieved for a given range and n' is the percent of words in terms of that range.

We recognize that neither of these two measures alone adequately reflects the real situation, especially where multiword searches are involved. Nevertheless, they serve together in some measure to convey the inherent utility of different range numbers and we present them in Table B-5 as calculated for the particular data of this analysis.

Table B-5

Normalized Recall Effectiveness for Each Range

	Recall Effectiveness Normalized by:	
Range Number	Term Frequency	Word Frequency
0	1.3	3.7
1	2.3	1.9
2	1.0	0.8
3	0.5	0.5
4	1.4	1.7
5 (title)	1.4	2.0

In some sense the above measures overemphasize the utility of a subject term on which a document drops when there is more than one such subject term, especially if that other subject term is of broader (more important) range number and, therefore, the subject term of narrower range number has added nothing to the retrieval. To correct for this kind of situation we may count the number of additional documents dropped by a given range when the order of range importance is taken as: 5 (title), 1, 2 and 3. (Ranges 0 and 4 have been omitted from this calculation because their order of priority is unclear.) Table B-6 and Fig. B-3 show the results of calculations made on this basis. Note that the corrected effectiveness is normalized with respect to word frequency and also with respect to the omitted range values.

Table B-6

Calculation of Corrected Effectiveness

Range	Percentage of Added Drops	Normalized (for Missing 0 and 4 Ranges)	Percentage of Words	Corrected Effectiveness
5 (title)	20.1	22.5	7.7	2.9
1	26.0	29.2	14.9	2.0
2	26.1	29.3	40.0	0.7
3	17.0	19.0	38.0	0.5
	89.2	100.0		

Reviewing these results, we find that, on an effectiveness-per-word-stored, the subject terms with ranges 1 and 4 are comparable to the title. On the basis of effectiveness per term stored, range-1 terms are superior to the title. The difference between effectiveness per term and effectiveness per word is attributable, of course, to the fact that certain ranges contain more words per term than other ranges.

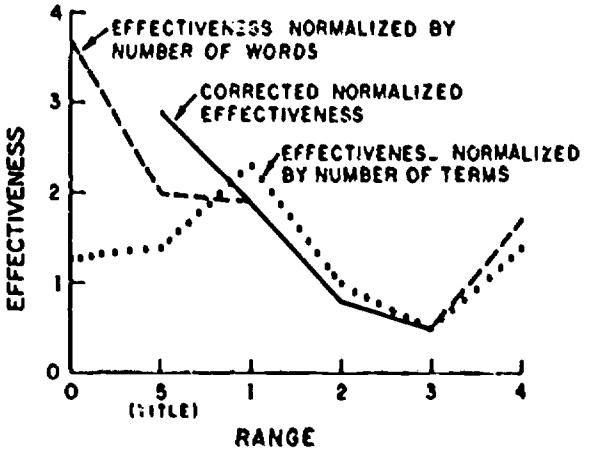


Fig. B-3 Recall Effectiveness of Subject Index Terms, by Range

We also note (Fig. B-3) that, although ranges 2 and 3 terms increase the overall effectiveness of retrieval (they return documents that users judge useful and that other ranges would not), they do so at a somewhat higher cost per useful retrieval than do the other ranges. This result is indicated in a graphic way in Fig. B-4 where the cumulative percentage of drops is plotted against the cumulate depth of subject indexing as expressed in percent of words indexed.

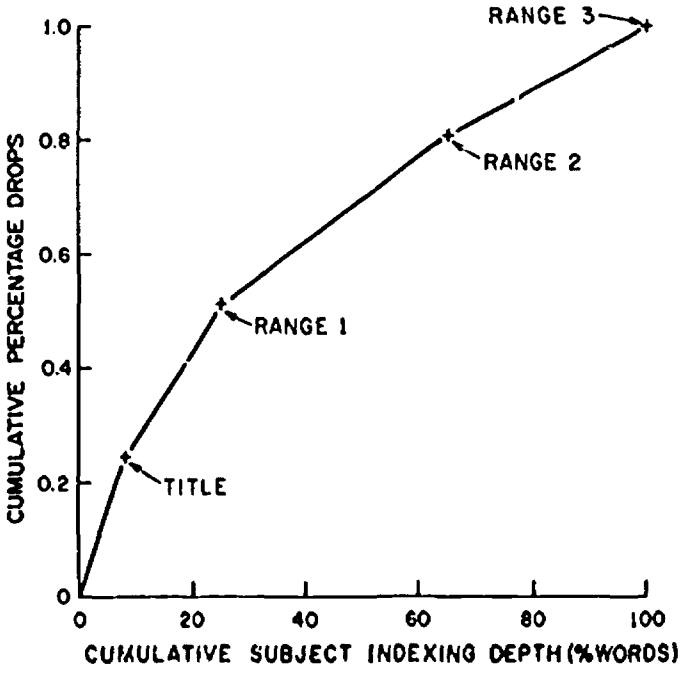


Fig. B-4 Cumulative Percentage Drops as Function of Cumulative Depth of Subject Indexing

Summary of Results. The data gathered to date are not conclusive. They cover only 23 sessions and are not yet fully analyzed. However, they suggest the following hypotheses which we are investigating further:

1. Some of the nontraditional fields, particularly those directly connected with the subject matter of the document (abstract, subject terms) justify inclusion in a computer-based catalog on the grounds that they will be frequently used.

2. Many more of the nontraditional fields justify their inclusion on a cost-effectiveness basis even when compared to the traditional fields on the grounds that, even though they are likely not to be used very frequently, they cost very little.

3. Deep indexing seems to improve performance and, although the amount of improvement per unit cost decreases as the depth increases, there is a large percentage of the recall base — about 50 percent — that is retrieved only on range 2 and 3 subject terms.

CATEGORY-I EXPERIMENTS

Category-I experiments are designed to determine the effectiveness of Intrex in terms of recall and precision by attempting to retrieve documents contained in a bibliography precompiled by conventional methods. No new Category-I experiments were begun during this reporting period. However, further analysis of the experiments already under way has begun to suggest the magnitude of the relative improvement of Intrex retrieval over abstract-journal retrieval, at least for simple one-word searches. In this section we summarize some of this analysis for the case identified in the previous semiannual report as Experimental Subject No. 4 (ES 4).

Experimental Procedure. ES 4 outlined his own special interest in the area of fatigue, and gave us a copy of a bibliography on this subject prepared by NASA Lewis Research Center. A search on the word "fatigue" was subsequently performed using Intrex. One hundred twenty-nine documents were retrieved from the data base which had approximately 8800 documents at that time. Twenty of these documents were presented to ES 4 for relevance judgments as part of a Category-II experiment (see in next section).

The abstracting journal Metals Abstracts (formerly Review of Metal Literature) was examined to ascertain its coverage of the general subject of fatigue and also to see how some of the documents in the Intrex retrieved set and the NASA bibliography had been indexed by this abstracting journal.

Comparison of Intrex Retrieval and Bibliography. There were 136 journal articles in the NASA bibliography from the common date scope with Intrex: 1967 and 1968. However, only 39 of these, or 29 percent, were from journals from which Intrex selects. Of these 39, 33 (85 percent) have been selected for the Intrex data base. Of those 33, 21 (64 percent) were in the computer data base at the time of the experiment. Of these 21, 18 (86 percent) were retrieved by Intrex; that is, a recall of 0.86.

The relevance of 20 of these documents to ES 4 was sampled in a Category-II experiment. ES 4 rated 8 of the 20 "of prime importance" to his work; 6 of the 20 "of some interest but not prime importance", and the remaining 6 were rated of no interest. However, even the 6 rated "of no interest" were considered relevant to the general topic of "fatigue". The reason for the "no interest" rating of these 6 had to do with the poor quality of the work or its approach: engineering application rather than fundamental or applied research. Thus, in terms of the sample figures, the precision of the retrieved set measured on a loose basis — as it usually is — is 100 percent. As measured on the most strict basis (only very important documents considered relevant) it is 40 percent. The precision measure for the moderately relevant documents is 70 percent.

ES 4 rated none of the three documents missed by Intrex of prime importance to him, and only one of them, of moderate importance, although again each of

them had some relevance to fatigue. Thus, assuming that the other 18 retrieved common documents divide about equally among the three levels of relevance (as did the 20 documents in the sample), we have the following recall and precision figures for Intrex:

Level of Relevance	Recall	Precision
Prime Importance	1.00(6/ 6)	0.40(8/20)
Secondary Importance	0.92(12/13)	0.70(14/20)
Any Relevance	0.86(18/21)	1.00(18/18)

It might be noted that none of the three articles missed by Intrex had the word "fatigue" in the title or indeed, in the body of the text.

Based on the above relevance sampling it may be projected that all 111 documents found by Intrex but not in the bibliography would be relevant. On the basis of at least secondary importance, the discovery set would be estimated at 111 x 0.7 = 78. On the basis of primary importance, the estimate would be 111 x 0.4 = 44. The discovery factor — defined as the ratio of newly discovered relevant documents to previously known relevant documents in the bibliography within the scope of Intrex (note that this is essentially the reciprocal of the previously defined "bibliography recall" of documents in Intrex) — is 111/21 = 5.3. This ratio would presumably remain about the same for the higher levels of relevance because the number of "relevant" documents would be decreased from 21 in similar degree as the number in the discovery set is decreased from 111.

Comparisons with Abstracting Journal. The abstracting journals considered were the ASM (American Society for Metals) Review of Metal Literature for the year 1967 and the Metals Abstract which replaced the Review of Metal Literature starting in 1968. Excerpts from the index to the 1968 volume of Metals Abstracts shows that there were nine major headings with the word "fatigue" for that year plus two see also references, four see references and 53 subheadings under the nine major headings. (The corresponding index for 1967 is very similar.) The number of documents listed under these headings for each of the years 1967 and 1968 were about 700 with very few, if any, documents listed under more than one heading.

It may be noticed that the abstracting journal has many more references per year than either the NASA bibliography or Intrex retrieved set. For example, on the basis of journal articles for one year (1967), the ratios of number of documents are: 700/108 = 6.5 for NASA and 700/82 = 8.5 for Intrex.

How many references were in Intrex or the bibliography but not in the abstracting journal? The answer to this question may be estimated by considering the different results for documents having the word "fatigue" in the title from those that do not. Of eight documents with "fatigue" in the title (seven from Intrex and one from NASA) all eight were found indexed under a "fatigue" heading in the abstracting journal.

Of 21 documents not having "fatigue" in the title (16 from the Intrex set and five from the NASA bibliography) only 2 (about 10 percent) were listed under one of the "fatigue" headings. The 19 not listed under fatigue were listed under the following headings:

Heading	Document(s) (D = Intrex No.)
Crack propagation	D 8267
Turbines	D 6961
Failure	D 9284
Aluminum base alloys, welding	D 9289
Baskets	D 8213
Fracture toughness	D 7340
Solidification	D 5274
Phase transformation	D 3548
Spacecraft, failure	D 3522
Nickel base alloys, reinforcement	D 3480
Titanium base alloys, corrosion	D 3283
Carbon steels, microstructure	D 3257
Plastic deformation	D 1366
Cyclic loads (2)	D 3746 and NASA Krempl: 1968
Creep Strength	NASA/Berkovits: 1968
Brittleness	NASA/Nicols: 1968
Fracture strength	NASA/Manjoini: 1967
Ductility, stress effects	NASA/Nishiharn: 1967

It may be noted that these 19 documents were listed under 18 different headings and many of these headings need not necessarily relate to the subject of fatigue. The results suggest strongly that the abstract-journal indexing for the subject of fatigue, at least, is very closely related to title words and very seldom goes beyond the title in identifying "fatigue" documents.

Approximately one-third of listings from a sampling of Intrex and NASA documents could not be found in the abstracting journal at all (under author name). Some of the documents were missed because the journal was not included in Metals Abstracts selection list (for example, Journal of Spacecraft and Rockets and Journal of Applied Mathematics) but most could not be explained on this basis.

The number of documents in the Intrex set without "fatigue" in the title is 33; that is, 33/129 = 25 percent. (The corresponding figure for a sample of 80 documents in the NASA bibliography showed 25/80 = 32 percent without "fatigue" in the title.) The discovery ratio for Intrex compared to the abstracting journal can now be estimated. Of the 33 titles without fatigue, about ten percent would be indexed under fatigue by the

abstracting journal. This leaves a discovery ratio of about 23 percent of the Intrex retrieved set that would be absent under "fatigue" by the abstracting journal.

This discovery ratio is quite similar to the 22 percent found for the ES 1 bibliography, the second bibliography in the Category-I series that has been analyzed in depth. However, in the ES 1 bibliography only abou. half of this total, or about 11 percent, was attributable to the deeper Intrex indexing. In the present case the full total of 23 percent is attributable to deeper Intrex indexing since the possible number of missed documents traceable to other causes has been discounted.

Does the discovery set with regard to the abstracting journal — that is, some of the documents without "fatigue" in the title — have a lower relevance rating than the other relevant documents? The answer suggested on the basis of a small sample seems to be that there is some, but not much, difference. Of 10 documents from the Intrex set without "fatigue" in the title, three documents (30 percent) were rated by ES 4 to be of prime importance, four documents (40 percent) of secondary importance, and three documents (30 percent) no interest but some relevance. The corresponding figures for a sample of 10 documents with fatigue in the title are five documents (50 percent) of prime importance, two documents (20 percent) of secondary importance, and three documents (30 percent) of no interest but some relevance. This same kind of result — that is, almost as good precision with the documents retrieved by the deeper indexed terms — has been observed also in the Engineering Library and Category-II experiments, and we are continuing to experiment to see if this may be a general result for Intrex.

Interpretation and Conclusions. The Intrex estimated recall, 0.86, and precision, 1.00, figures indicate a very good Intrex system performance, at least for this subject, a fairly broad subject which can be characterized rather well by the single term, "fatigue". The 100 percent precision indicates that the term "fatigue" is fairly specific and unambiguous, at least within the realm of the physical sciences.

Because of the relative lack of completeness of the NASA bibliography, it was found more appropriate to define the Intrex discovery set and ratio in terms of the abstracting journal. Here it was found, basically, that the Intrex discovery set consisted of 90 percent of those articles Intrex retrieved that did not have the word "fatigue" in the title. Since 25 percent of the articles retrieved by Intrex did not have "fatigue" in the title, the discovery ratio is about 23 percent.

The discovery set is thus attributable to deeper Intrex indexing. Of the 14 Intrex documents in the discovery set, the number of highest range subject terms with "fatigue" were: range 1, 1 document; range 2, 8 documents; range 3, 5 documents; range 0, 1 document. The fact that only one range 1 term appeared in this set reflects the Intrex indexing in which range 1 indexing is usually either the title itself or a somewhat expanded version of the title and the fact that the abstracting journal seemed to have indexed all articles with "fatigue" in the title under "fatigue" headings.

Comparing the results of the retrieval on the ES 1 bibliography with those on fatigue (ES 4), we have the following:

Figure of Merit	ES 4 Search	ES 1 Searches
Recall	0.86	0.87
Precision	1.00	0.90 Intrex data base 0.25 Abstract data base
Discovery Ratio (indexing only)	0.23	0.11

The figures for the ES 1 searches are based on the best Intrex strategy which was on the chemical formulas for the three chemical compounds being studied. The lowered precision figures, especially where searching the abstracting journal is considered, reflects the fact that it was only one aspect of the properties of these compounds that was desired for the Intrex bibliography. The figures for recall and discovery are quite comparable. The somewhat-lower discovery figures on ES 1, though perhaps not statistically significant, may reflect the fact that the abstracting journal employed in that search (Physics Abstracts) indexed under the chemical formulas somewhat deeper than by title only, which was roughly the depth of indexing in Metals Abstracts for "fatigue". This hypothesis is being tested through further analysis.

The relatively low figures for the discovery ratio in these two experiments appears to suggest a somewhat less value to the deep Intrex indexing than results for other Intrex experiments would indicate. A possible explanation for this discrepancy is the fact that the two situations reported here involved only single-word searches, whereas other experiments included multi-word searches in which the use of semantically related words was important. This hypothesis, too, is being further checked.

CATEGORY-II EXPERIMENTS

The Category-II class of experiments has been designed to determine the usefulness of the various fields of the Intrex catalog as a means for identifying documents that satisfy users' needs for information. Through analyses made on data derived from these experiments, evaluations are also being made of the in-depth feature of one of the fields, namely, the subject-index field.

The Category-II-A series, described in the preceding Activity Report, is a specific set of experiments and now includes a total of nine experimental subjects who have posed eleven different topical requests. A total of 356 documents enter into the results obtained thus far.

Discussion of the Category-II-A Series. As reported previously, this series tests the following fields for their ability to identify useful articles:

Field No.	Field Designation
24	Title
21	Author(s)
24, 21, 47, 74	Title, Author, Location and matching subject-index phrases, as a group
71 (or 70)	Abstract (or Excerpt)
73	All subject-index terms

The reason for concentrating on these fields is that they are likely to be most descriptive of document content. It will be valuable to know the relative effectiveness of these fields in terms of their recall potential and as identifiers of useful information. Among the questions which we aim to answer are these:

* From the user's viewpoint which among the fields tested comes closest to full text as an identifier of useful information?

* What is the relative recall potential of a computer-stored catalog whose inverted (look-up) file is constructed from title words only? From Abstract words? From title and subject-index terms, as the Intrex catalog now is constructed?

* How does the in-depth subject-indexing feature of the Intrex catalog influence the recall potential of the Intrex system?

The motivation for this study is, of course, the desire, in a computer-stored catalog, to provide the most powerful lookup system with the fewest number of stored words. Economics dictate this requirement. At present the average number of words in the fields being tested is

Field	Average Number of Words*
Title	9
Author(s)	5
Journal Location	7
Abstract	120
Subject Terms	100
Matching Subject Terms	40

* Estimates based on taking six characters per English word.

Description of the Experiments. The test procedure described in the 15 March 1970 Activities Report (p. 14) remains in force. It is repeated here with a few added comments which can be made as a result of recent experiences. In an interview the experimental subject is first asked to provide the experimenter with a verbal description of his literature need. From this description the experimenter constructs a written statement of the problem and from this statement a set of search phrases is established.

The experimenter then searches the data base and obtains hard-copy printouts of the resultant material under the various fields listed above. This material is used by the experimental subject as a basis for making the evaluations. The importance of each document based on the given item of information is ranked by the experimental subject in accordance with the following numerical-rating system:

Value Rating	Meaning
1	A document that is of prime importance to the user's needs.
2	A document that is useful, but not of prime importance.
3	No interest.
4	Cannot make a judgment.

In addition, for the two fields, 73 and 74, where subject-index terms are available for making evaluations, the experimental subject is asked to check those subject-index terms that are particularly helpful in arriving at the numerical rating given to the document.

Information is presented to the experimental subject in the following manner: A given field is presented for evaluation for all documents, all other information on each document being withheld. The order in which each field is present for evaluation is: title; author; title, author, journal location and matching subject terms; abstract; all subject terms; and full text. In order to keep value judgments among fields independent, a time lapse of a day or more is injected between evaluations of each field.

After all fields of the catalog have been examined, the subject is asked to read and evaluate the full text of each article. His evaluation of full text is taken as the reference against which evaluations of the various fields are compared.

It should be noted that the Category-II-A series is a set of off-line experiments in that the experimental subjects do not, themselves, exercise the data base. Because it is not a purpose of these experiments to measure total recall, it is not necessary to present for evaluation every possible document that might be of interest to the experimental subject. Nevertheless, it is possible to employ the documents involved in the experiments in an analysis of precision as well as incremental recall.

Our experience is that a set comprising from 20 to 40, or at most, 50 documents is the most satisfactory, all things considered. The experimental subjects have been helped substantially with sets of this size, while at the same time it has been possible to complete all the work with an experimental subject within a reasonable time span and without loss of his interest and diligence. Too few documents may result in low useful yield; too many documents may overburden the experimental subject, with the result that the experiment embraces too long a time span.

Results. Nine experimental subjects with a total of 11 search topics have been run through the Category-II-A series. The Fig. B-5 bar graphs, which are similar to those presented in the preceding Activity Report, reflect up-to-date results. A comparison with previously reported results shows that original trends continue, namely, the value ratings given abstracts and subject-index terms coincide more closely with ratings given full text than do ratings of titles and authors coincide with ratings given to full text.

Data included in Fig. B-5 are in slightly different form from those reported six months ago. In particular, data that carry the experimental subject's value rating of 4 (cannot make a judgment) have been separated from the other data that comprise Fig. B-5; data involving rating 4 are now analyzed separately (see Fig. B-4). This procedure is more representative of the facts, since the value ratings of 1, 2, and 3 are clearly in decreasing order of usefulness to the experimental subject, whereas a 4 rating is an entirely different kind of decision. Hence, the 4 should not be lumped with documents whose value lies at the 1, 2 or 3 levels.

With respect to Fig. B-6, it should be observed that the experimental subjects were most frequently unable to render a value judgment when author only was presented to them. Out of 356 documents, 149 or 41.8 percent could not be judged for their value on the basis of the author's (or authors') names alone, and hence received a 4 rating. Perhaps this is not surprising, since all but two of the experimental subjects are graduate students with very few years of experience in their research areas. Hence, it might be argued that they would not be expected to know in detail the nature of the research currently being conducted by co-researchers in other laboratories throughout the world. Two experimental subjects were senior faculty members and forefront researchers in their areas. They applied the 4 ratings to author names less frequently. However, the hypothesis that more experienced researchers are better able than less experienced researchers to make value judgments on author names alone cannot be validated until the size of our sample has been increased.

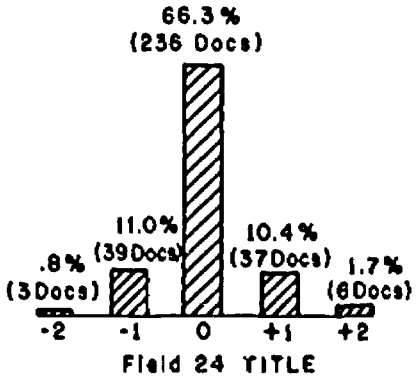
Outside the author field, the 4 rating (cannot make a judgment) was used infrequently. It is worth noting in Fig. B-6, however, that where the 4 rating was used, the highest percentage of those documents was ultimately judged to be of no interest; the lowest percentage turned out to be of prime importance.

KEY:

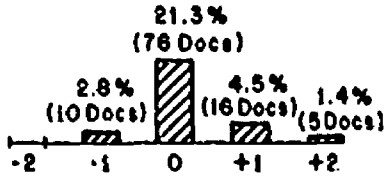
0	Indicates catalog field(s) and full text received same value rating
+1, +2	Indicates catalog field(s) received higher value rating than full text
-1, -2	Indicates catalog field(s) received lower value rating than full text

TOTAL NUMBER DOCUMENTS EVALUATED

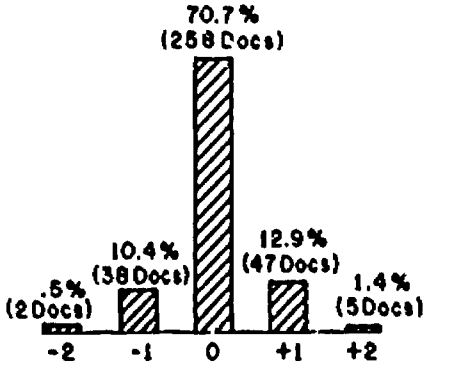
Field 24: 356 Documents
Field 21: 356 Documents
Fields 21/24/47/74: 365 Documents
Fields 70/71: 326 Documents
Fields 73: 356 Documents



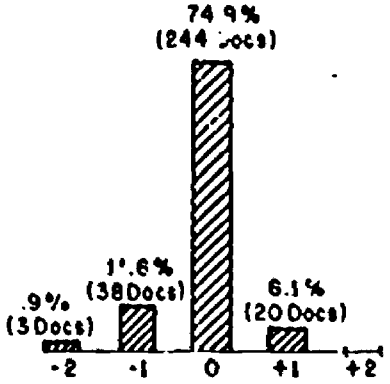
Field 24 TITLE



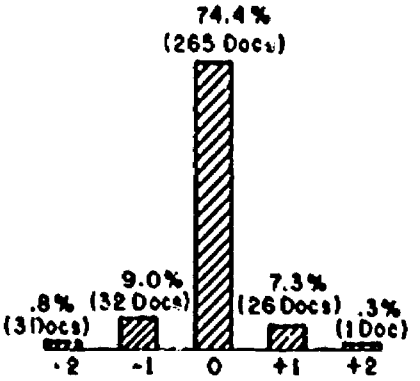
Field 21 AUTHOR



Field 24 TITLE
Field 21 AUTHOR
Field 47 JOURNAL LOCATION
Field 74 MATCHING PHRASES



Field 71 ABSTRACT



Field 73 SUBJECT TERMS

Fig. 8-5 Comparison of value ratings given by nine experimental subjects to individual fields with value ratings given to full text of the documents.

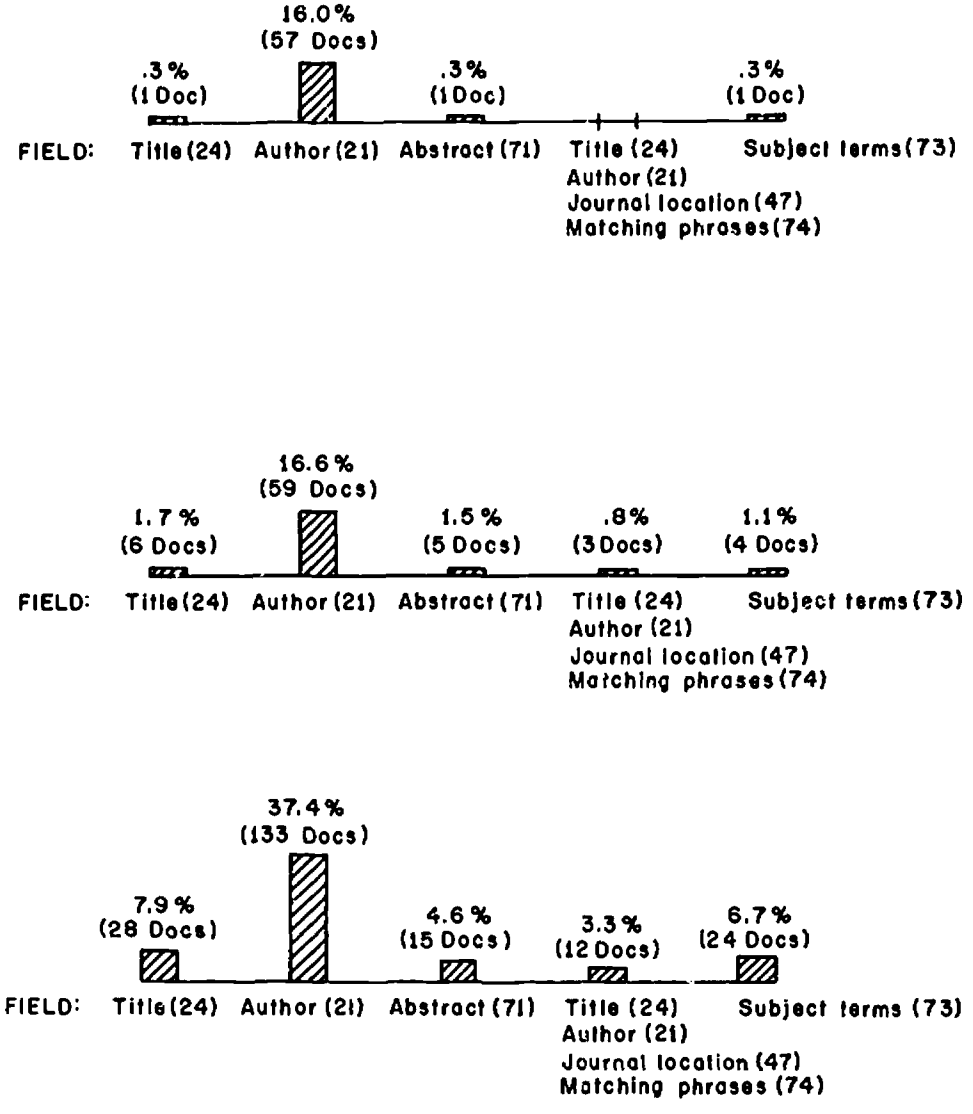


Fig. B-6 Comparison of Rating 4 (cannot make a judgment), by fields, with ratings given full text. Nine experimental subjects.

The 0-level data (meaning field and full text were rated identically) were analyzed by fields for the nine experimental subjects with their 11 research topics. Our purpose was to determine the spread of the data around the mean values for each field which had already been given in Fig. B-6. The result is shown in the scatter diagrams of Fig. B-7. From this figure it may be observed, for example, that approximately 74 percent of the 356 documents analyzed received the same rating on

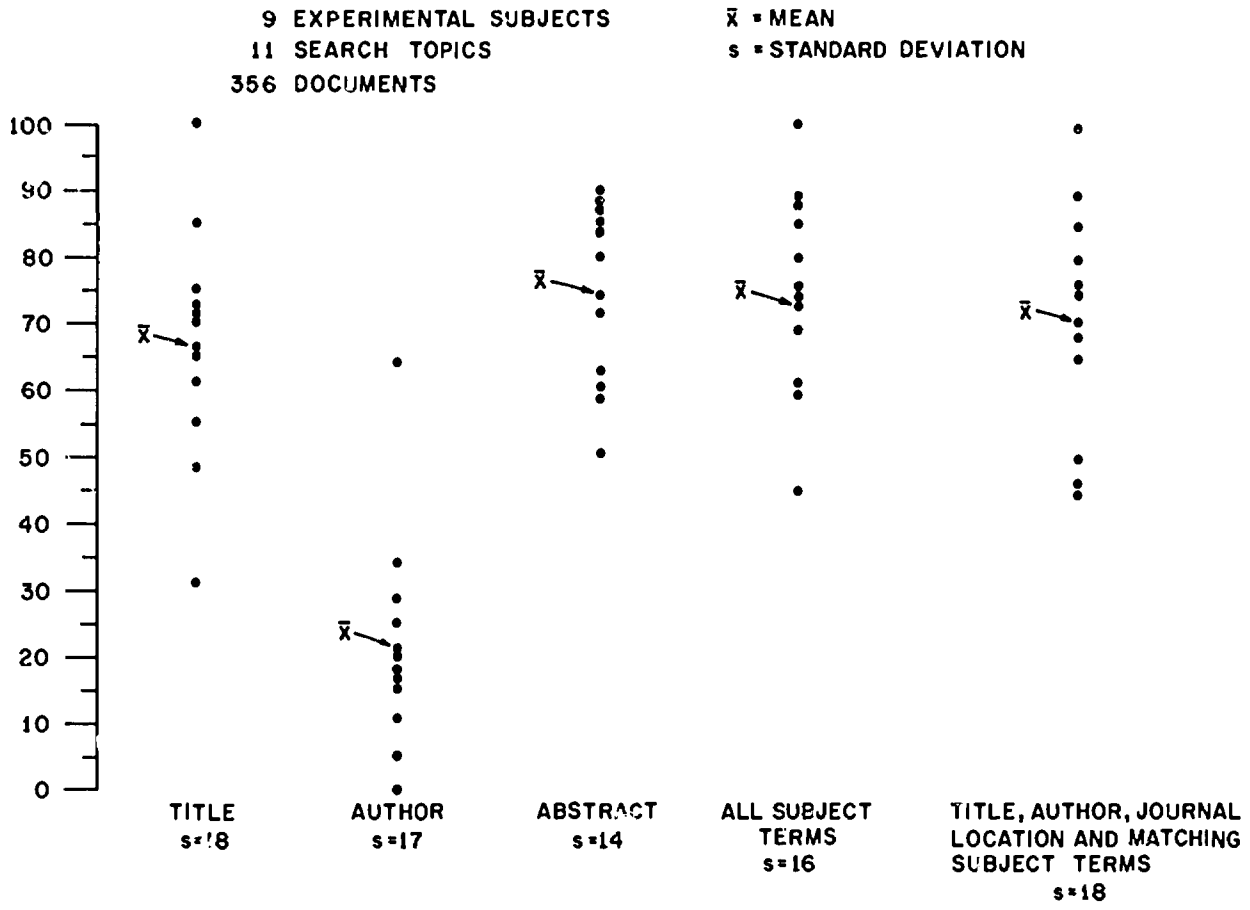


Fig. B-7 Percent of Documents with Identically Rated Field and Full Text

the basis of subject-index terms and full text. The scores of individual experimental subjects ranged, however, from 44 to 100 percent, with a standard deviation of 16 percent. Similar spreads are observed for titles, abstracts and authors when their value ratings are compared with ratings of full text.

The scatter-diagram information gives us an insight into the smoothness (or roughness) of our data and will be used as a guide for tightening controls on the

experimental procedure and for deciding when sufficient data have been gathered and it is valid to draw sound conclusions. It may be observed that there is a fairly wide variation in 0-level scores among the different experimental subjects. For that reason, it is difficult to be very precise, at present, about the ultimate levels we expect to find on these fields for large data samples. Indeed, we have some evidence that the utility of the fields as content indicators is dependent on such factors as the characteristics of the experimental subject and his research area and the overall level of importance of the documents presented to him for evaluation. However, there has been considerable consistency in the relative utilities of the different fields, and this gives us more confidence for stating the following hypothesis: Abstract and subject term utilities are rather close and both are significantly better than titles; the utility of matching subject terms seems to fall about half-way between titles and the other two fields.

On the basis of Fig. B-5 one might decide that a rather high price in word-storage costs is paid as one works with the abstract or subject-index terms in contrast with title words only. The word count goes from nine words, average, for title to 120 words, average, for abstract as the identically rated (0-level) document percentage increases from 66.3 percent to 74.8 percent. This point was examined further, therefore, from another viewpoint. Several search phrases which were used to develop the document lists for the Category-II-A series were matched, not only to the subject-index terms for each document, but also to the title and abstract of the same document in order to determine the presence or absence of the search phrases in titles and abstracts of the same documents. This procedure tells us about the "recall potential" of the subject-index terms. More specifically, it answers the following question: Given a specific search request, such as, for example, mode-locked lasers, or Brillouin scattering, or Rayleigh scattering, how many documents are recalled from Intrex (where matching is on subject-index terms), how many documents would have been recalled if the search phrases were matched against title words only, and how many would have been recalled if the search phrases were matched against words in the abstract only?

The result is shown in Fig. B-8 for five search phrases. A total of 285 documents was obtained as a result of a common occurrence of the words in the search phrases and the same words in the subject-index terms. However, the search phrases appeared in the titles of only 136 (48 percent) of the original 285 documents and in the abstracts of 157 (55 percent) of the original 285 documents. The in-depth feature of the Intrex subject indexing is thus giving us a much higher yield, on the basis of this sample, than would be obtained if we searched only titles or only abstracts. We are continuing to examine this point more exhaustively.

Another analysis which has emerged from the Category-II-A series is shown in Fig. 9. Here we analyzed 256 documents which were given to the experimental subjects from the viewpoint of the range number of the matching subject-index terms that

SEARCH PHRASE	SUBJECT TERMS	TITLE	ABSTRACT
MODE-LOCKED LASERS	22	9	17
BRILLOUIN SCATTERING	124	64	71
RAYLEIGH SCATTERING	70	29	30
THERMAL RAYLEIGH	13	6	7
STIMULATED RAMAN	56	28	32
TOTAL DOCUMENTS RETRIEVED	285	136(48%)	157(55%)

Fig. B-8 Documents Retrieved by Search-Phrase Occurrences in Title, Abstract and Subject-Index Terms

caused the documents to be retrieved. Refer to the upper curve of Fig. B-9. For 36.3 percent of the documents, retrieval was the result of the search phrases appearing in titles. An additional 14.1 percent of the 256-document sample was retrieved as the result of occurrence of the search phrases in range-1 index terms; an additional 19.5 percent of the sample was retrieved because of an occurrence of the search phrases in range-2 index terms, and so forth. The lower value of 36.3 percent for title-only retrieval for this document sample compared to the 48 percent for the 285 document sample may reflect the fact that this sample was retrieved on the basis of more complicated searches.

The documents that were retrieved as a result of matches on title words, range 1, 2, 3, 4, and 0 subject-index terms were then examined to see how many in each category were given a value rating of "prime importance" (value rating 1) and a rating of "useful" (value rating 2) by the experimental subjects. The results are shown by the lower curves of Fig. B-9. A total of 44.9 percent of the 256-document sample was in the "prime-importance" category and 17.2 percent were in the "useful" category. Of significance is the fact that index terms carrying range-numbers 2, 3, and 4 contribute substantially to the recall potential of the Intrex system. Thus we have another measure of the value of the in-depth subject indexing. In tabular form, these results are:

Search Phrases Matched to:	Increase in Cumulative Number of Documents Retrieved	Increase in Cumulative Number of Documents Retrieved Receiving Value Rating 1	Increase in Cumulative Number of Documents Retrieved Receiving Value Rating 2
Title words	93	51	16
Range-1 terms	36	14	5
Range-2 terms	50	23	8
Range-3 terms	49	20	11
Range-4 terms	21	3	3
Range-0 terms	7	4	1
	256	115	44

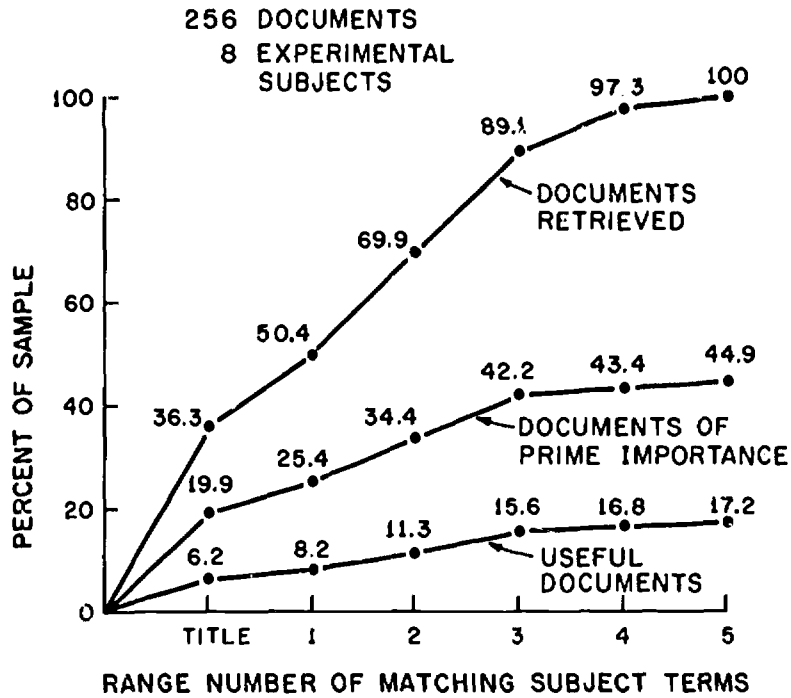


Fig. B-9 Cumulative Percent of Documents Retrieved as a Function of Range Number of Indexed Terms

These results can also be used to measure retrieval precision as a function of the range number of the subject terms providing that retrieval. Precision is defined as the percent of retrieved documents that are relevant; for this analysis relevance is equated with the value rating of "prime importance." Figure B-10 shows the precision of that subset of additional documents retrieved by extending the depth of the indexing to a given range level. Of significance is the fact that precision does not drop very much as the depth of indexing increases.

While the results given in Figs. B-9 and B-10 show the incremental value of increasing depth of subject indexing, they do not adequately portray the cost of this indexing. One good measure of the cost of indexing, both for indexer time and computer generation and storage costs, is the number of words in the subject terms. In Fig. B-11 we present the percentage of relevant documents retrieved as a function of the percentage of index words in all subject terms up to a given depth of indexing. Note that we are measuring cumulative incremental recall, as opposed to total recall, in that the base of relevant documents is taken as just those 115 documents found of prime importance among the 256 retrieved — there may be other relevant documents in the Intrex data base that were not retrieved. The percentage of index words of a given range is taken from calculations made on the whole data base, not on the 256 documents of this particular analysis and some corrections may have to be made to account for this fact.

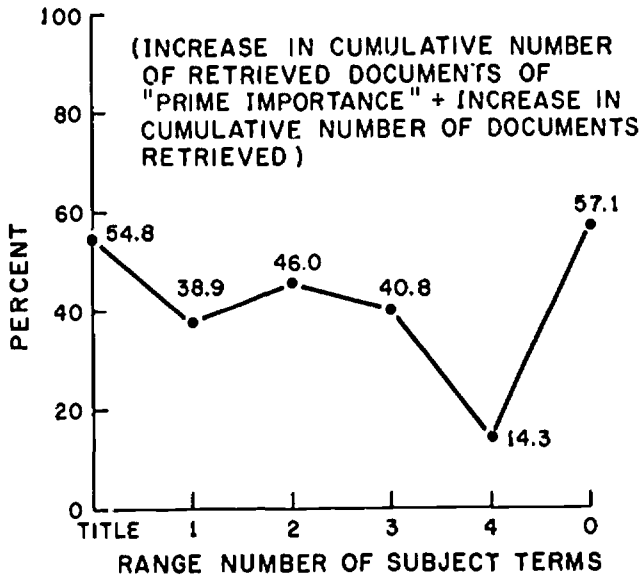


Fig. B-10 Precision of Retrieval, by Range Number

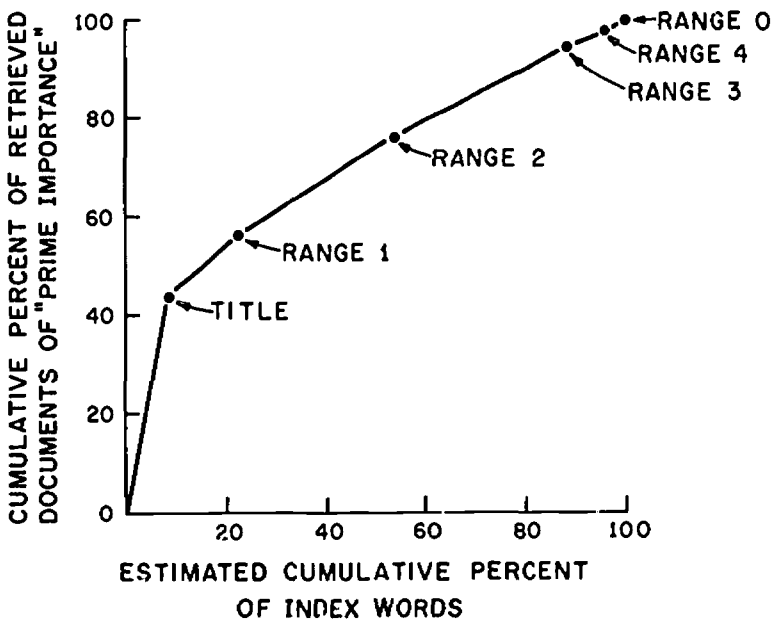


Fig. B-11 Recall as a Function of Depth of Indexing Measured in Percent of Index Words

Of significance is the quantitative verification given to "the law of diminishing returns"; namely, that increasing depth of indexing brings increased recall but at a diminished rate.

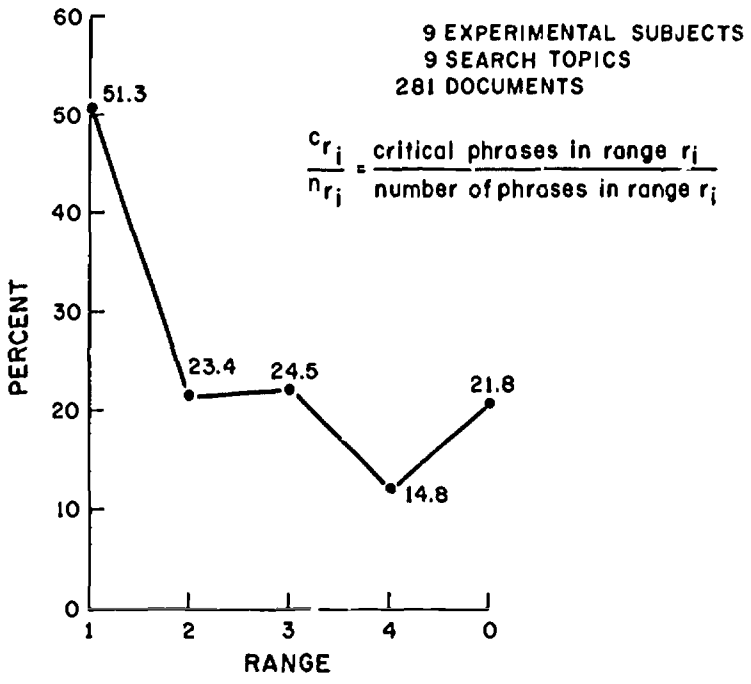
Since our last Report we have continued analysis of experimental subjects' judgments concerning the critical nature of certain subject-index terms as a means of helping them to make document evaluations in field 73 (subject-index terms). The two graphs in Fig. B-12 show the results of this analysis. Range-1 terms appear to have the greatest probability of being utilized; over half of them were checked as helpful by experimental subjects. On the other hand, although only 24.5 percent of all range-3 terms were checked as helpful, nevertheless — because of the preponderance of range-3 terms — they formed the largest fraction, 33 percent, of all subject terms marked as helpful.

CATEGORY-II EXPERIMENTS, SELECTIVE DEEP ANALYSIS

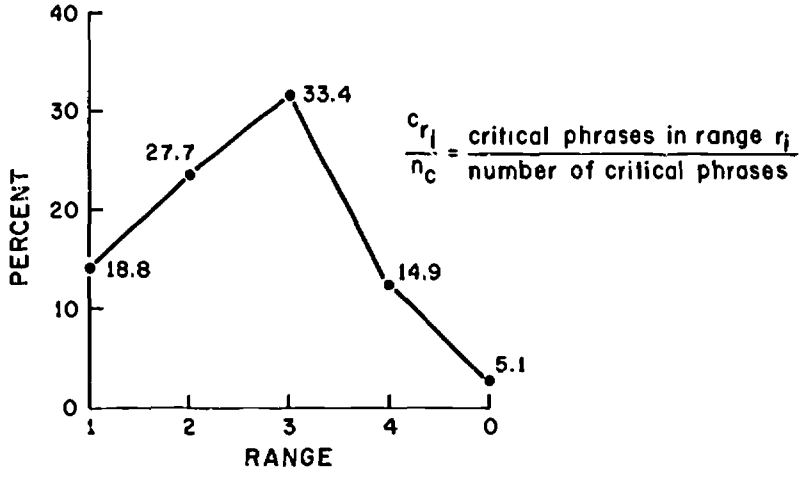
Two of the nine experimental Category-II subjects, ES 12 and ES 23, served as subjects for Category-I and Category-III experiments. This additional involvement helped to delineate the subject's problem in greater detail and, in particular, enabled us to gain additional information on why the subject made certain evaluations for the Category-II experiments. These additional data were used to extend the analysis of the experiments as summarized below.

Figures of Merit. The Category-II experiments attempt to determine how well the various catalog fields serve to indicate the usefulness or relevance of documents to the user. Therefore, we may coin the term Indicativity to express the capability of a field to serve that purpose. A good measure of Indicativity for a given catalog field is the percentage of cases in which the subject's evaluation of document relevance on the basis of that field is the same as the evaluation given to that document on the basis of seeing the full text. An Indicativity failure refers to a case where these evaluations differ.

Variability of Results. A major concern of this analysis was to analyze the variability in a subject's evaluations. In an attempt to ascertain the magnitude of this variability the analyst had ES 12 redo his evaluations on the basis of full text. Two changes in ranking were recorded. These two changes represent a 10-percent variation and if this amount of variation were regularly found, then we might consider perfection in utility of a field to indicate document relevance as 90 percent Indicativity. Actually, recognizing that there will be variability in evaluating the catalog fields as well as the full text, we might project even a lower optimum Indicativity. However, it is likely, as described below, that the documents involved in the variability are the same ones in each case. This would tend to make the optimum Indicativity closer to 90 percent, rather than 80 percent, as would otherwise be projected.



(a) Percent of Subject-Index Phrases within each Range Number Identified by Experimental Subjects as Critical to Document Ranking



(b) Percent of Total Subject-Index Phrases Identified by Experimental Subjects as Critical to Document Ranking, by Range Numbers

Fig. B-12 Analysis of Subject-Index Terms as Aids in Document Evaluations

Several causes for the variability can be cited:

(1) Borderline cases: Some documents just naturally seem to fall between a 1 and a 2 or a 2 and a 3. There were over a dozen cases in the 240 evaluations made for these two subjects where there was explicit indication that two rankings were being considered. In these cases, either one ranking was crossed out and replaced by another or both rankings were given and left standing.

(2) Interpretation of Rankings: What makes a document important or relevant to a user has many facets. Considerations here include quality of work; novelty to user; applicability to user's current, future or past needs; does it give the answer to user's problem or merely relate to it; is it relevant to problem as stated but not to user's real problem; etc. The evaluations may change as a subject emphasized one or another of these facets. Also, a subject may simply forget the instructions. ES 23 asked the analyst to re-explain the ranking system at one stage.

(3) Change of Interest of ES: A new interest, or other change in subject's interest, can occur as happened with ES 23.

In order to ascertain the extent and reasons for the variabilities, the analyst questioned the subjects and otherwise attempted to understand what was going on. In particular, if there was a discrepancy between the evaluations derived from full text and a given field, it was attempted to determine whether some objective piece of information was missing from the field that had caused the discrepancy or whether one of the variation-causing factors listed above was involved. In many cases, this determination was possible. In other cases, it was not possible from the data gathered so far. The results for each case are summarized by an "information factor". If there seemed to be some factual piece of information missing from the catalog field presented that caused the different judgment, then an information factor of 1 was assigned. If, on the other hand, one or more of the 3 non-factual causes of variation listed above seemed to be the principal cause, an information factor of 0 was assigned. Where there was mixed evidence, an information factor of one half was assigned. Thus, the information factor is an estimate of the likelihood that the lack of completeness of the catalog field in question was the principal cause for the different evaluation.

Let us summarize the results. In Table B-7 for each field we indicate the total number of cases of Indicativity failures and the weighted number of cases that could be reasonably attributed to lack of information in the catalog, that is, factual factors. This weighted number is obtained by adding the information factors. Also given are the original Indicativity for both users put together and the adjusted Indicativity considering the information factors. This adjusted Indicativity is the Indicativity assuming only factual-type failures occurred. For example, for Field 24 the sum of 16 Indicativity failures out of 40 document evaluations gives an Indicativity score of 0.60;

however, since only 11 of these failures have a factual basis, the adjusted Indicativity is (40-11)/40 = 0.73.

The results show that the adjusted Indicativity is about 12 to 14 percent higher than the original Indicativity. While Indicativity went up consistently when both users were averaged, in the case of the abstracts, the results were improved much more dramatically for ES 12 than for ES 23.

Table B-7

Factual Indicativity Failures as Fraction of all Indicativity Failures in Field Evaluations

Parameter Measured	Experimental Subject	Field				
		24	74/75	71	73	ALL
Failures	ES 12	5/7	2/4	.5/3	1.5/4	9/18
	ES 23	6/9	4/7	3/5	1.5/5	14.5/26
	Sum	11/16	6/11	3.5/8	3/9	23.5/44
Indicativity	Average for both Subjects	24/40 =.60	29/40 =.73	29/37 =.78	31/40 =.78	
Adjusted Indicativity		29/40 =.73	34/40 =.85	33.5/37 =.90	37/40 =.92	

Utility as Function of Number of Words. A basic hypothesis to be accepted, rejected, or modified is the "law of diminishing returns" or length hypothesis for catalog data. The law says that the more data that is included, the better the evaluations on it will be, but at a decreasing rate. In particular, we need to determine quantitatively the shape of this curve for our particular kind of data and whether there are any significant exceptions to it. For example, we know that subject terms are roughly as indicative as abstracts. But do the subject terms provide greater Indicativity per word of catalog data stored? This would seem to be the case, since abstracts make up about 40 percent of the bulk of catalog data while subject terms are only about 30 percent. These figures need to be refined, especially with regard to the actual Indicativity and counts for the documents in the sample with particular reference to those cases where Indicativity failures occur and taking the information factors into account. Thus we may try to answer the question of how much the deeper coverage and greater selectivity (in terms of emphasizing important information-bearing words) of the subject terms outweighs the greater naturalness and comprehensibility of the abstracts. By trying to detect the operative factors in each case we may even hope to quantify the importance of each factor and thus make justifiable decisions about not only whether to use abstracts and/or subject terms but how to specify what makes good abstracts and/or subject terms.

When we consider the particular field-length statistics for the two given experiments, the above comments can be amplified. Looking at the summary information of Table B-8, we see that the general nature of the length hypothesis is well verified, especially when non-factual factors are screened out to give adjusted indicativity. The supposition that subject terms are giving better indications per word is sidestepped by observing that for this particular sample of 40 documents, the subject terms are longer than the abstracts. These trends are indicated graphically in Fig. B-13 which plots the results given in Table B-8.

Table B-8

Field Length (in Words) and Indicativity

Subject	Parameter	Field			
		24	74*	73	71
ES 12	Field Length (Words)	10.0	45	116	132
	Indicativity	.65	.80	.80	.84
	Indicativity (Adjusted)	.75	.90	.93	.97
ES 23	Field Length (Words)	10.6	42	171	109
	Indicativity	.55	.65	.75	.72
	Indicativity (Adjusted)	.70	.80	.93	.83
Average ES	Field Length (Words)	10.3	43	144	120
	Indicativity	.68	.73	.78	.78
	Indicativity (Adjusted)	.73	.85	.93	.90

*Note: Accuracy scores are actually based on the combination of fields 74 (matching subject terms) and 75 (author, title and journal location) and for the latter an additional equivalent of about 20 words should be added.

This latter observation bears out the importance of analyzing the individual cases and not just taking averages over the whole data base or set of experiments. In particular, in this case the further we go into the details the more obvious the situation becomes. That is, the figures for the individual experiments and the individual documents seem to emphasize the point. In five out of eight cases where there was a difference of evaluation between Fields 73 and 71, and the information factor was not zero, the field with the correct evaluation was significantly longer; in two cases the length was about the same; and in only one case was it significantly shorter. Furthermore, in three of the five first-mentioned cases the field with the Indicativity failure was much shorter than the average length of that field and there was a definite indication that this was the reason for the failure.

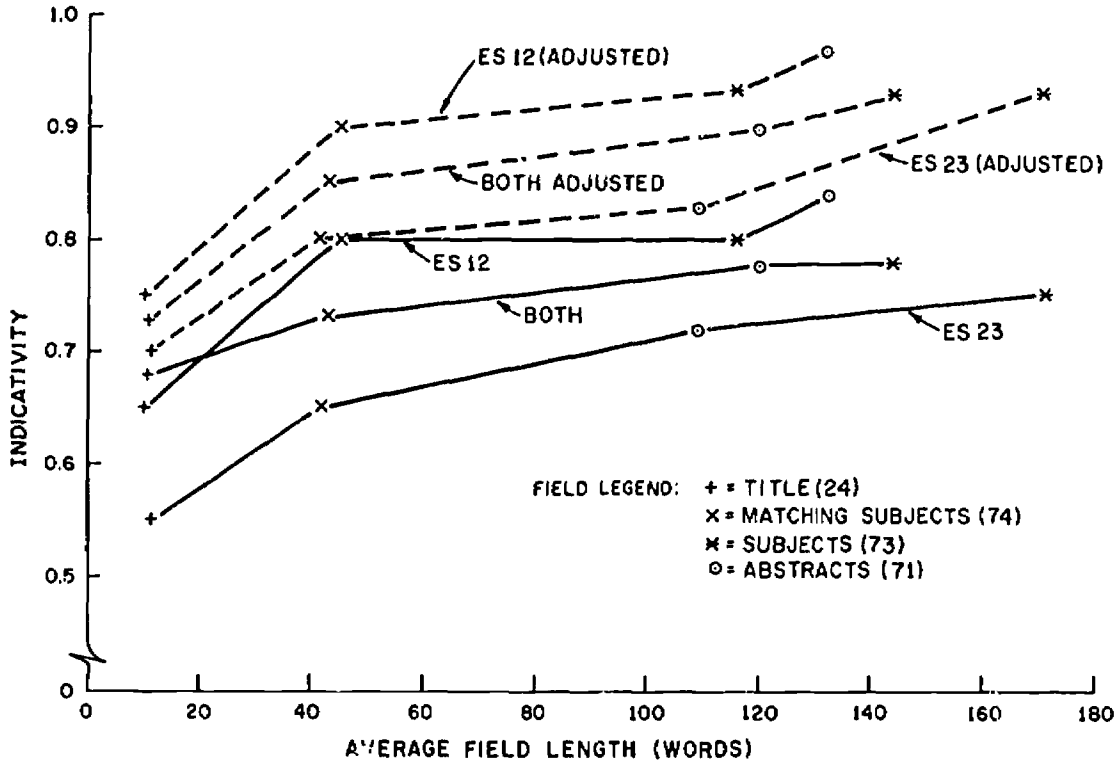


Fig. B-13 Indicativity as Function of Average Field Length

Thus we have some support for the hypothesis that the length rule is a first-order parameter to be considered in affecting field Indicativity.

Combinations of Fields. Two fields when considered together may provide a better indication of relevance than either field alone. Thus if the Indicativity for subject terms is 81 percent and for abstracts is 82 percent, then that for showing both fields together could conceivably be, say, 90 percent. As an upper bound to the possible improvement for these two particular Experimental Subjects (ES) we may consider the situation in which the ES, on seeing the fields together, would always choose that evaluation he made on the individual field that was closer to the full-text choice. In that situation the Indicativity would be 95 percent as compared to the average Indicativity of the individual fields of 79 percent. Perhaps a more reasonable approximation would be to take the situation where the ES, when faced with a different evaluation for the individual fields, would simply flip a coin to make his choice. In this situation the Indicativity would be raised only eight percentage points instead of 16, that is, to 87 percent. These rather large increases in Indicativity reflect the fact that only rarely in these two experiments (two times out of 40) were both the Field 73 and Field 71 evaluations wrong on the same document.

Summary. It has been shown how a deeper analysis of Category-II experiments can be employed to answer detailed questions about utility of catalog fields. In particular, preliminary analysis indicates that about half of the Indicativity failures observed in Category-II experiments are the result of variational factors not related to the explicit information content of the field itself. It also appears that Indicativity is strongly correlated to the length of the field in words.

These results are not yet conclusive; many other observations and hypotheses raised in this analysis remain to be tested in detail. As far as the methodology of deep analysis, it is proposed that results can be obtained for manyfold fewer experiments than otherwise at the cost of greater effort in experimental observation, analysis, and the need of the analyst to guard against forcing the analysis to conform to any preconceived ideas he might have.

CATEGORY-III EXPERIMENTS

Category-III experiments investigate the online, interactive features of Intrex and explore the need for improvements in system functions. Two additional Category-III experiments were begun during this reported period and their status is described below. In addition, further analysis of one of the previous Category-III experiments was accomplished.

Experimental Subject No. 4. This subject was a Professor of Metallurgy who had previously participated in Category-I and Category-II experiments. His problem for this experiment was to get background material for a paper he was writing on "the effect of processing variables on fatigue". His first Category-III session on this problem was run in the guided mode with the Intrex analyst doing the typing on the Intrex console. In this session, which lasted about 90 minutes, various catalog outputs from approximately 50 documents was viewed. Of these, about 20 were marked as especially interesting and the experimental subject requested full-text hard copy of them. The subject expressed high praise for these results and the Intrex system.

Several initial observations may be made. These observations are in no way conclusive but raise hypotheses to be tested by further analysis. In the first place, it was apparent that the original statement of the problem was useless as a search request: "processing variables of fatigue" yielded no documents; "processing variables/ and / fatigue" yielded one document that was judged "not useful" by ES 4.* It was necessary to use more specific terms for "processing variables". The direct interaction of the experimental subject was important in arriving at these specific terms because the closest

* See Section D for explanation of the less restrictive nature of the matching algorithm implied by "and" and how the statement is actually made in the present Intrex retrieval system.

entries found in our thesauri (NASA and DOD), "processing" and "process variables", had many related terms but none of interest to our subject. Also the subject was able to modify his initial terms to make them more useful. For example, the term "grain size" was modified to just "grain" — a useful modification that a surrogate searcher might believe too broadening.

A second observation is that the less restrictive matching algorithm implied by the AND command did not result in any lowering of precision. For example, "grain fatigue" found 13 documents of which seven (54 percent) were judged as likely to be useful by the user, whereas "grain and fatigue" found 29 with 10 of the additional 16 (62 percent) being judged useful.

Additional sessions and analysis are planned for this class of experiment.

Experimental Subject No. 5. This subject was a chemist working in industry. He was interested in learning certain properties of laser dyes. While the Intrex data base does not fully cover this problem area, it was felt that there was enough of an overlap to justify using the system. The experimental subject was run in a guided mode on the Intrex console but with the subject doing his own typing.

About 10 documents were viewed, both from catalog information and full text — in general, the catalog information did not seem sufficient to determine usefulness to this subject although the set of documents seemed perfectly relevant to the stated search request. While none of the documents themselves seemed useful, the subject was not surprised since he had previously found a very low percentage of useful material in searching for this very specific information using standard bibliographic reference tools.

This subject expressed enthusiasm with the speed and relative ease of Intrex retrieval compared to the standard bibliographic reference tools he had been using. Nevertheless, certain features like output delays and the manual paging operations required on the Intrex console bothered him. There are, of course, ways to alleviate these problems and as we gather statistics on just how frequently and how many users are bothered by each problem, we shall have a better idea as to how much it is worth to resolve any one of them. Further analysis needs to be done on the specific information problem presented here to see whether Intrex type indexing would allow for specific retrieval if the proper search request were made and if relevant documents were in the data base.

C. AUGMENTED-CATALOG INPUTTING

Staff Members

Mr. A. R. Bencnfeld
Miss M. A. Jackson
Miss L. T. Lee
Miss V. A. Miethe
Miss L. Rossin

Cataloger Assistants

Miss L. A. Langille
Mrs. E. T. Raska
Miss R. L. Seegal

Graduate Students

Mr. D. R. Cherry
Mr. N. A. Clark

Undergraduate Students

Mr. L. E. Bergmann
Mr. C. R. Davis
Mr. P. L. Martin
Mr. T. A. Skotheim

SUMMARY

As of 15 September 1970, 14,800 documents have been indexed, and 12,589 catalog records have been completely processed into the computer-stored data base. All active processing files have been rearranged by semiannual dates of publication so that the most current material awaiting any processing action is always processed first. A more equitable distribution of work loads among the input staff has been introduced. For the years 1967-1970, the distribution of the data base by type of document, and a ranking of the largest contributing source journals are given. Preliminary data concerned with a reinvestigation of average processing characteristics is reported and comparisons are made with data previously reported. The initial results of a pilot study of the number, distribution, and cause of errors in data preparation are presented.

LITERATURE SELECTION

Shown in Table C-1 are the distribution by date of publication of: (1) journal articles in the computer-stored data base or in some stage of input processing; (2) an estimated number of journal articles remaining to be selected; and (3) conference papers, including those papers in proceedings published as an entire journal issue and which are not treated by Intrex as journal articles. Approximately 150 other data-base documents, primarily journals, reports, books, and isolated pre-1967 articles, are not included. The estimated number of journal articles to be selected was derived by proportionate extrapolations from the number of articles already selected as of June 1970 from each of the current 72 Intrex source journals. The larger number of 1967 journal articles is attributed to the fact that the Project's initial document-selection procedure

was performed by librarians who selected more documents peripheral to the prime subject concerns of the data base (see the Semiannual Activity Report for 15 March 1968). Under current document-selection procedures for the subject interests of the current data base, approximately 4600 journal articles are selected per year, along with approximately 400 conference papers.

Table C-1

Journal-Article and Conference-Paper Composition of Data Base

	Date of Publication					
	Pre 1967	1967	1968	1969	1970	Total
Journal Articles in Stored Data Base or in Processing		4686	3371	4518	1239	13814
Estimated Journal Articles still to be Selected		45	51	92	3392	3580
Expected Total Journal Articles as of Dec. 31, 1970		4731	3422	4610	4631	17394
Percent of Total Number of Journal Articles		27.2	19.7	26.5	26.6	100.
Conference Paper in Stored Data Base or in Processing	1982	386	551	400	269	3588

The ten source journals providing the most articles selected for the Intrex data base for each of the years 1967 to 1970 are ranked in Table C-2. The figures for 1970 are estimates. Physical Review ranks first for each of the years. Fluctuations from year to year are generally small and are not considered significant. Of the ten highest ranking journals for 1968 to 1970, nine are primarily oriented to the physics literature and only one is primarily oriented to metallurgy.

PROCESSING

The workflow remains essentially as described in the Semiannual Activity Reports of 15 March 1970, 15 September 1969, and 15 March 1968. The 15 March 1970 report described two major changes that were incorporated into the workflow. In one change, frozen processing buffer stacks were created to smooth fluctuations in input to each major operation; these buffers held the oldest documents so that more recent material would always be processed first. This latter aspect has been further advanced by

Table C-2

Article Contributions from Major Intrex Source Journals

Journal	1970		1969		1968		1967	
	Rank	Articles (estimated)	Rank	Articles	Rank	Articles	Rank	Articles
Physical Review	1	500	1	468	1	351	1	595
Journal of Applied Physics	2	297	4	220	2	316	4	345
Metallurgical Transactions	3	280		*		*		*
Metallurgical Soc. of AIME Trans.		*	5	202	3	254	7	176
Soviet Physics — Solid State	4	253	3	309	4	220	6	267
Physics Letters	5	229	2	357	5	203	2	454
Physical Review Letters	6	224	6	195	6	192	5	315
Journal of Chemical Physics	7	198	8	160	7	137	2	454
Physical Society of Japan Journal	8	193	7	192		**		**
Soviet Physics — JETP	9	158	9	156	8	117	8	154
Physica Status Solidi	10	138	11	139		**		**
Applied Physics Letters	11	128	12	126	9	98	10	137
Solid State Communications	12	126	10	153	15	57	21	49
Philosophical Magazine	16	84	18	79	10	95	13	83
Acta Metallurgica	18	83	17	80	12	84	9	151

* The Metallurgical Society of AIME Transactions merged with the American Society for Metals Transactions Quarterly to form in 1970 the Metallurgical Transactions.

** Not an Intrex source journal for that year.

arranging all active processing files by publication date in semiannual intervals. Thus, the most current material awaiting any processing action is now always processed first. In the second major change described in the previous report, all correction-loop processing was placed on a monthly cycle. This has significantly smoothed out the keying, proofreading, and editing sequence. Regular monthly additions to the computer-stored data base are now commonplace.

During the reporting period, changes were made in the apportionment of work loads among the input staff. This has contributed to a more equitable distribution factor, particularly with respect to the reviewing of indexing, and to the proofreading of the first printout of a file of ten records. The reviewing process is now paced such that all individual records indexed in one week are reviewed no later than the end of the following week and sent to the typists. In addition the reviewing of each experienced student-indexer's work is now apportioned among all the staff indexers; previously, the work of any one student was assigned to one staff member for review. Inexperienced students are still responsible to a staff member for review of their work and further training.

Modifications were also introduced into the process of second proofreading of a file printout. This process is performed by a typist and it consists of proofreading the printout of a file that has been previously edited once. Greater emphasis is now devoted to checking for correct editing of errors caught during the first proofreading process, and for correct field tags and end-of-record delimiters. No additional checks of field data are made except for the title, index-term and abstract fields.

Work is proceeding on drawing up specifications for an expanded set of computer checks for errors in the input-data stream. These specifications will also allow us to better coordinate manual and computer checks for errors. The preliminary evaluation work on error analysis discussed below is related to the error-checking specifications.

Work is currently underway on the preparation of revised and expanded sets of instructions for initial keyboarding of catalog records and for on-line editing of catalog records.

As of 15 September, 1970,

14,800	documents have been indexed;
14,670	records have been reviewed;
13,450	records have been keyed;
13,300	records have passed through the first proofreading and editing process;
12,589	records have been completely processed into the computer-stored data base.

EVALUATIONS

Preliminary work has been completed on a reinvestigation of input performance and processing characteristics. A comparison is made here between data reported earlier which reflected methods of indexing and processing applicable to the first 4200 documents processed by Intrex, and initial data from recent studies which reflect, among other things, a major change in indexing methods applicable to all documents processed after April 15, 1968. Descriptions of both earlier data and the major indexing change are given in the Semiannual Activity Report of 15 September 1968. Briefly, the indexing process was modified to reduce the extent of word redundancy by eliminating dual-weighted terms, by eliminating full index terms also appearing as a string within another term, and by combining similarly worded and logically equivalent terms into a single term with compound-subject or object sentence elements. For example, given the three separate terms "barium titanate (4)", "optical-dispersion curves for barium titanate (3)", and "optical-dispersion curves for strontium titanate (3)", the first term would be eliminated and the last two terms would be combined into the single term "optical-dispersion curves for barium titanate and strontium titanate (3)".

The upper half of Table C-3 gives a comparison of the average time to index, descriptive catalog, review, and key a catalog record. The earlier data are from a 120-document sample of records processed from January to April, 1968; the more recent data are based on a 185-document sample systematically selected from catalog records processed during the last two years. The data from the recent sample are also displayed by type of document, that is, regular journal article, letters-type article, and conference paper. The average time to index a document has decreased about 29 percent from 28.3 minutes to 20.2 minutes per record; this is statistically significant at better than the 0.01 level. The ratio of average review time to average indexing time is about 0.4 for both samples; it is also about 0.4 for each type of document in the later sample. This ratio is apparently independent of indexing-method changes, and the type of document indexed.

The lower half of Table C-3 gives additional information about changes in average document indexing time per document-page, derived by the mean-of-the-ratios method, and changes in the average page length of documents. Data are from two samples and each is analyzed by type of document. The 1546 documents in the earlier sample come from a complete survey of essentially all of the first 3000 documents processed but excluding all of the documents indexed during an indexer's learning period. The more recent data are from the systematically selected 185-document sample. The average document length has increased during the last few years for each type of document; for all documents, the length has increased 22 percent from 4.2 pages to 6.1 pages. The average indexing time per page has decreased 42 percent for all types of documents from 6.7 minutes per page to 3.9 minutes per page; this is in agreement with the increased

Table C-3

Average Time Required to Perform Various Unit Operations of the Indexing Process, and Associated Average Characteristics

Operation of Characteristics	Date of Analysis	Type of Document			
		Regular Journal Article	Letters-Type Article	Conference Paper	All Types
Subject Indexing (average minutes)	January 1968-April 1968				28.3
	May 1968-June 1970	22.0	12.3	26.3	20.2
Descriptive Cataloging by Librarian (average minutes)	January 1968-April 1968				5.4
	May 1968-June 1970	4.4	2.7	3.8	4.7
Review of Subject-Indexing and Descriptive Cataloging (average minutes)	January 1968-April 1968				10.4
	May 1968-June 1970	8.9	4.6	11.1	8.0
Keying (inc. Desc. Cat. by Typist) (average minutes)	January 1968-April 1968				17.1
	May 1968-June 1970	21.2	11.9	17.0	18.4
Indexing Time per Page (average minutes/page	March 1967-March 1968	6.8	9.4	5.1	6.7
	May 1968-June 1970	3.1	4.7	5.8	3.9
Average Document Length (pages)	March 1967-March 1968	6.4	2.6	2.9	4.2
	May 1968-June 1970	7.7	2.8	5.4	6.1
Sample Size (documents)	March 1967-March 1968	605	344	597	1546
	January 1968-April 1968				120
	May 1968-June 1970	114	50	21	185

average document length and decreased average indexing time per document. The increase in average indexing time per page for conference papers is probably a function of the small sample size. An examination of the average indexing time per page by document type shows that the rank ordering of article types that are easiest to index has reversed, so that currently the subject indexing of conference papers consume the greatest time and subject indexing of full journal articles consume the least time. The 50-percent decrease in average indexing time per page for letters articles without any significant change in page length is tentatively attributed, at least in part, to an increase in the number of letters articles being published that have an accompanying abstract. An abstract appearing with document text is a prime content clue-point and an initial source of index terms.

The preliminary data on average processing times and characteristics discussed above indicate a number of changes in processing characteristics which can be attributed to changes in document characteristics and to changes in indexing and other processing procedures. A more comprehensive study of performance characteristics with a larger sample of the data base is being planned.

Presented here are the initial findings of a pilot study undertaken to determine the number, major contributory cause, and field distribution of errors discovered during the proofreading of printout files of catalog records; the study did not cover residual errors, that is, those errors still undetected in the catalog records. The feasibility of an extended study of data-preparation error analysis is shown; the role such an analysis can have in designing and operating the data preparation processes supporting an information retrieval facility is indicated.

The analysis considered errors at two general levels, that of data within a field, and that of data at a level affecting an entire field, an entire catalog record, or an entire file of ten catalog records.* Data in error at the latter, larger level includes entire fields that are omitted, mistagged, or repeated, and entire records or files whose names are omitted or misspelled, or whose end-of-record or end-of-file tags are omitted, mistagged, or repeated. At the smaller level, data within a field were considered as being

---

* A catalog record is composed of a set of fields, each containing a specific type of data; each data field is explicitly labelled (or tagged) for machine identification. During keyboarding and through all subsequent pre-formatting processing operations, ten catalog records are batched to form a file. A file is assigned a numbered name of its own, and specific tags separate catalog records within a file, and indicate the end of a file.

composed of words, interword spacing or punctuation, and delimiters. Words were further characterized as being either linguistic or symbolic. Linguistic words are the normal language words or their abbreviations bounded by a space or graphic. Symbolic words were arbitrarily defined for this pilot study as character strings representing an explicit unique data code, and sequences of special-character representations and mathematical and chemical symbols. Thus, the major possible error types that were distinguished were errors in entire fields, in linguistic words or symbolic words, in interword spacing or punctuation, and in internal field delimiters, that are either omitted, misspelled (or analogously, mistagged), or extraneous.

Each error was traced to a cause through an examination of transmittal sheets and successive generations of printouts. The causes of error are briefly defined below, but the category names used in this pilot study are not completely indicative of the contents of the category:

1.	Typist Oversight	Exclusion or extraneous inclusion of a field or field data when such data is present on the transmittal records or is data which the typist must generate. For example, a field present on a transmittal sheet is omitted from the keyed record, or a line in a printed text abstract is either skipped or keyed twice.
2.	Cataloger Oversight	Exclusion or extraneous inclusion of a field or field data when such data must be cataloger generated. It includes errors from improper indexing review procedures. For example, the author's purpose is omitted from the transmittal sheet and the omission was not caught during review.
3.	Recognition	An error resulting from illegible writing, illegible text copy, or ambiguously interpreted characters, each meeting a criterion of reasonable resemblance. For example, the letter "k" is keyed instead of the similar looking Greek "kappa".
4.	Miskeying	An error in misspelling of words or graphics not attributable to any other cause. For example, wrong case, or transposition of letters, or keying "eror" for "error".
5.	Edit	An error arising from any kind of incorrect error marking or editing step. For example, illegible proofer markings, or specifying a non-unique string such that each occurrence, correct and incorrect, is changed during context editing.
6.	Policy	An error arising from ambiguous, unclear or changing policy, provided that the keyed data is a reasonable interpretation of a rule.
7.	Mechanical	A machine-generated error, or an error generated by a program bug.

Each error was thus analyzed by type of error, attributable cause, field in which the error occurred, and the process during which the error was discovered. If a word contained more than one error, only the first occurring one was analyzed but provision was made to tabulate the number of words with multiple errors. In addition the word length was noted of omitted or extraneous multiword phrases. Mistakes in linguistic words were also noted as to misspellings in the stem or in the ending for those words in fields (subject, author, title) where inverted-file stemming procedures would be affected. The pilot study was done on a sample of 14 files (each containing 10 catalog records) with 2 files randomly selected from each block of 100 files over the last 700 files processed; this sampling excluded all records indexed before the major change in indexing methodology discussed earlier.

Before a discussion of the tabulations is given, it may be helpful to consider some processing characteristics of the files comprising the sample. These correction-loop characteristics are reported in Table C-4; as such, this table is an extension of the data given in Table C-3, although the documents and files sampled in each case are different. The average keying time for the error-analysis sample is 16.0 minutes per record, the average record size is 1846 characters (or 308 six-character English words), and the time for first proofreading is 48.4 minutes/file. These sample figures are each lower than previously reported. This reduction can only be partially attributed to a smaller sized record which, in turn, is partially attributed to less redundant indexing. Additional operative factors may become known through examination (other than for errors) of the particular records comprising the files selected for study. The sampling method itself may be partially responsible; the files were randomly selected within blocks, but the records contained within a file potentially are runs of records for similar-type documents. Consequently more definitive results than initially reported here need to be based upon a larger sample. While the number of errors is expected to vary directly with the size of the keyed record, the error rates and error distribution is not expected to change markedly, other things being equal. Therefore, it is of some value to consider the pilot study results; the discussion given below is brief because our evaluations of the study are still in progress.

The nonresidual error analysis data are presented in condensed format. Table C-5 shows the number of nonresidual errors by type of error and proofreading process in which they were discovered for errors at the larger level of a file, record, or entire field; the attributed cause of these errors is shown in Table C-6. Table C-7 shows the number of nonresidual errors by type of error for errors at the level of data within a field; a breakdown by proofreading process in which these errors were discovered is not detailed, but the percentage of the total errors discovered during the first proofreading is indicated. Table C-8 shows the attributed cause of errors at the level of data within a field. In Tables C-5 and C-7, the data is differentiated by primary fields and non-primary fields only, and not by the specific field in which an error occurred. A

Table C-4

Correction Loop Processing Characteristics
for Error Analysis Sample Files

Processing Characteristic	Sample Size	Average Characteristic
Keying Time	140 records	16.0 minutes/record
Character Count	140 records	1846 characters
Proofreading Time		
- First Proof	14 files	48.4 minutes/file
- Second Proof	14 files	14.1 minutes/file
- Third Proof	3 files	2.3 minutes/file
Console Edit Time		
- First Edit	13 files	23.9 minutes/file
- Second Edit	8 files	7.6 minutes/file
- Third Edit	1 file	1.0 minutes/file
Computer Edit Time		
- Processing Time		
- First Edit	14 files	14.6 seconds/file
- Second Edit	6 files	11.7 seconds/file
- Third Edit	2 files	9.4 seconds/file
- Swap Time		
- First Edit	14 files	16.5 seconds/file
- Second Edit	6 files	7.3 seconds/file
- Third Edit	2 files	4.7 seconds/file
Number of Files Correct After		
- First Edit		5 files
- Second Edit		7 files
- Third Edit		2 files

Table C-5

Number of Non-Residual Errors, by Type, at the Level of Entire Files and Entire Fields (Error Analysis Sample of 14 Files)

Error Level and Proofreading Stage in which Errors were Discovered	Type of Error		
	Omission	Mistagging	Repetition
Entire-File and Entire-Records within a File			
- First Proofreading	3	3	0
- Second Proofreading	3	1	1
- Third Proofreading	0	0	0
Entire Primary Fields within a Record			
- First Proofreading	2	0	0
- Second Proofreading	0	0	0
- Third Proofreading	0	1	0
Entire Non-Primary Fields within a Record			
- First Proofreading	11	5	5
- Second Proofreading	2	1	1
- Third Proofreading	1	1	0
Total Non-Residual Errors by Type (= 41)	22	12	7

Table C-6

Number of Non-Residual Errors, by Cause, at the Entire-File and Entire-Field Levels

Type of Error	Cause of Error							Total Non-Residual Errors by Type
	Typist Oversight	Cataloger Oversight	Mis-keying	Recog-nition	Edit	Policy	Mechan-ical	
Field Omitted	13	1			2	6		22
Field Mistagged	7	1	1				3	12
Field Repeated	4			3				7
Total Non-Residual Errors by Cause	24	2	1	3	2	6	3	41
Percent of Total by Cause	58.5	4.9	2.4	7.3	4.9	14.6	7.3	100

Table C-7

Number of Non-Residual Errors, by Type, at the Level of Data Within a Field (Error Analysis Sample of 14 Files)

Type of Error	Primary Fields	Non-Primary Fields	Total Non-Residual Errors by Type	Percent of Total Discovered in First Proof-reading
Linguistic Word				
- Omitted	90	53	143	100
- Misspelled	61	69	130	69.2
- Repeated	38	39	77	97.5
Symbolic Word				
- Omitted	9	21	30	96.6
- Misspelled	12	36	48	89.6
- Repeated	1	7	8	87.5
Interword Spacing/ Punctuation				
- Omitted	14	12	26	80 8
- Misspelled	6	12	18	16.7
- Repeated	1	0	1	0
Delimiters				
- Omitted	22	14	36	97.3
- Misspelled	0	1	1	100
- Repeated	6	4	10	100
Total Non-Residual Errors, by Field	260	268	528	86.6

Table C-8

Number of Non-Residual Errors, by Cause,
at the Level of Data Within a Field
(Error Analysis Sample of 14 Files)

Type of Error	Cause of Error							Total Non-Residual Errors by Type
	Typist Oversight	Cataloger Oversight	Mis-keying	Recog-nition	Edit	Policy	Mechanical	
Linguistic Word								
- Omitted	135	8						143
- Misspelled	30	22	45	17	8		8	130
- Repeated	37	40						77
Symbolic Word								
- Omitted	28	1				1		30
- Misspelled	18	3	14	8	4	1		48
- Repeated	7					1		8
Interword Spacing/ Punctuation								
- Omitted	14	2	4	2	4			26
- Miskeyed	3		3		9	2	1	18
- Repeated					1			1
Delimiters								
- Omitted	30	5				1		36
- Misspelled			1					1
- Repeated	4	6						10
Total Non-Residual Errors by Cause	306	87	67	27	26	6	9	528
Percent of Total by Cause	58.0	16.5	12.7	5.1	4.9	1.1	1.7	100

primary field is defined here as being any field which affects the creation of the inverted files and directories to the computer-stored data base, and all other data fields are considered non-primary. The primary fields are author, title, subject terms, document number, and microfiche number.

An examination of Tables C-5 and C-7 clearly shows that most errors are caught during the first proofreading process. At the level of file and entire field, 70.7 percent of the 41 nonresidual errors were caught during first proofreading and an additional 22 percent were caught during a second proofreading. At the level of data within a field 86.6 percent of the 528 nonresidual errors were caught during first proofreading and an additional 12.3 percent were caught during a second proofreading. For total error analysis, 85.5 percent of the 569 nonresidual errors were caught in a first proofreading. The primary fields affecting inverted file or directory creation contained 46.2 percent of the total number of errors. The subject and title fields alone accounted for 42.2 percent of the total errors, a figure roughly proportional to the contribution of these fields to the catalog record size. Only 22.3 percent of the total errors are found in the abstract or excerpt fields, a figure lower than their expected size contribution. The total average number of nonresidual errors per record is 4.1.

An examination in Tables C-5 and C-7 of the ratios among the error types of omission, misspelling or mistagging, and repetition shows that, except for delimiters, repetitions of data occur least often, and that omissions are generally more prevalent than either misspelling or mistagging. For linguistic words, omission of words and misspellings of words occur about equally often, but the more difficult symbolic words are misspelled at least half again as often as they are omitted.

When we look in Tables C-6 and C-8 at the causes for the known errors, only a small percentage (about 10 percent) can be attributed to ordinary keyboarding mistakes, that is, automatically striking another key in place of the correct key. Presumably this represents for a given speed/accuracy quality typing level an error threshold, first, because any keyboarding effort will normally contain errors of this type, and second, because errors attributed to other causes are at least theoretically subject to reduction, if not elimination, through procedural changes and training.

The majority of errors, nearly 60 percent, are attributable to the category "typist oversight". Procedural check controls could probably eliminate most of these errors. Only about 6 percent of the errors could be attributed to a reasonable problem in recognition or interpretation of written and printed data; this is a surprisingly low number considering that a large block of data appears on transmittal sheets in handwritten form and that a fair number of technical symbols require special coding, but on the other hand, several procedural checks have long been employed in this area of our data preparation. Because the recognition category was open to varying interpretation in this pilot analysis, some recognition errors may have been attributed to miskeying or to typist oversight and vice versa. A more complete study will need stronger

definitions of these categories. Cataloger oversight errors are large in number but procedural checks, particularly during pre-keying review, would be expected to reduce errors of this type. A larger number of errors than expected were the result of improper correction-loop procedures; some of these errors arose because a non-unique string was specified in the context editing, and others arose because of illegible or buried proof markings. Some procedural checks are now being introduced into our processing to reduce edit errors. Errors attributable to unclear or changing policy should gradually be eliminated over time; indeed, the large number of field level policy errors in the sample was specifically caused by one ambiguity on when to include a text excerpt in the catalog record. Transient mechanical errors are nearly impossible to correctly categorize as such but their presence is extremely small; non-transient mechanical errors are more obvious and indicative of a machine maintenance problem. The 7 percent mechanical errors at the field level are due entirely to a program bug which inserts extra characters at the end of a file.

A more detailed breakdown of error types by specific field, although not presented here in tabular form, does show that in the subject, title, and author fields, where words are stemmed for the inverted file, 61 linguistic words were misspelled with 85 percent of the misspellings occurring in the stem and 15 percent of the misspellings occurring either in the word endings or in common words. These percentages probably reflect the proportional number of characters occurring in the stem or in the ending of words. Further, of 87 linguistic words or symbolic words omitted from the subject field, 3 were single word strings only; the remaining 84 represented the total number of words in 11 omitted multiword phrases whose average length was 7.6 words. All 22 words omitted from the abstract field constituted a single 22-word phrase. Thus nearly all word omissions in these fields really represent skips over lengthy phrases.

The non-residual errors that formed the basis of this pilot study refer to those errors discovered and corrected before a catalog record enters the public data-base. Therefore, knowledge about the residual errors is required in order to calculate the true total number of errors and the true efficiency of each proofreading step. Although exhaustive proofreading was not part of this analysis, at least 21 residual errors are known to be contained in the records of this sample. This comprises 3.7 percent of the nonresidual errors and we can consider this percentage as a rough indicator of the lower bound of residual errors.

Data preparation is a significant part of a computer-based information system. Highly error-free data are a necessary requisite to the success of the system, most importantly for retrieval (both catalog record and text), and secondarily for data integrity and aesthetics. This pilot-error analysis study indicates the potential utility of such an analysis as an aid in the design and improvement of data processing functions. A more comprehensive sampling and analysis will require several refinements in the sampling

techniques, in the definitions of error types and causes, in the word basis location of an error, and in normalization criteria. The pilot study has shown the feasibility of a larger study.

D. STORAGE AND RETRIEVAL

Staff Members

Mr. C. E. Hurlburt
Mr. P. Kugel
Mr. R. S. Marcus
Mr. M. K. Molnar
Professor J. F. Reintjes

Graduate Students

Mr. R. Goldschmidt
Mr. N. Goto

Undergraduate Students

Mr. D. Griffin
Mr. R. Prakken

SUMMARY

In the area of computer programming, in addition to maintaining data-base generation we have introduced several improvements into the retrieval program. One improvement involves the ARDS combined catalog/text-access console. The retrieval program was modified so that a request for full text by a user is reformulated by the system into the proper form and transmitted for direct interpretation by the text-access subsystem. A second improvement has been to implement a general Boolean-function capability by which the user can combine results from simple searches. In conjunction with this facility, lists of references derived from searches of a given session can now be saved and used at a later date by the same, or different, user. Through a third improvement, the encoding of superscripts, subscripts and special symbols (Greek characters, for example) is decoded for proper presentation on the Intrex console. Also, certain fields which have only encoded information are now decoded before presentation to the user.

In the field of computer-systems analysis, we have begun an in-depth study of the storage and operating-time costs of our present system. This study will provide a basis for making improvements in system efficiency, and will give us additional insights into the design of future, large-scale systems.

RETRIEVAL-PROGRAM IMPROVEMENTS

Boolean Operations. A general facility was introduced into Intrex for combining previously retrieved and named document sets through use of a full set of Boolean pairwise operators. It was decided to limit the Boolean capability to pairwise combinations of previously named lists because of the greater simplicity of the parsing problem and because more complicated multilist operations would often take so much computer execution time that real response time would be sluggish. For example, the command, IRON AND OXIDE will perform an intersection of the document numbers found in the lists named IRON and OXIDE and create a new list containing only those reference words found to exist in both the IRON and OXIDE lists. The WITH command performs the same kind of function but in addition insists that the references match

on subject-term number as well as document number. This is, in fact, the same operation that is performed by the search module on the several words of a SUBJECT search request.

The OR command appearing between two list names performs a disjoining of the document numbers of those lists. The resulting list will contain all references (merged to preserve sort order) found to exist in either list.

The negation function of culling one list from another can be performed by the NOT or AND NOT commands. L1 AND NOT L2 will produce an output list containing those references of L1 whose document numbers were not found in L2. L1 NOT L2 will produce a list of references from L1 none of which come from the same subject-term of any of the L2 references. Resulting lists may also be named if further operations are to be performed on them, although this is not necessary if the next operation is performed immediately. In fact, Boolean commands may be concatenated on a single command line.

Saving Document References. The new SAVE command allows a named list to be entered into a "SAVE" file on the disk where it will be preserved for future restoration and use, even on a different day or by a different user. Thus, for example, bibliographies may be saved, updated, and shared.

A SAVE file must be first established by the user and assigned a name via the command:

SAVE FILE sfname

where sfname is any six-letter or shorter mnemonic the user wishes to assign to his file. A directory of SAVE file names is kept and checked when names are added or used. To re-activate the lists contained in a SAVE file, the user issues the command USE sfname which makes the lists in the file named sfname the current group of named lists and, therefore, available for output or further manipulation.

If the user only wants to add new lists to the SAVE file without a complete reactivation of the existing lists, he types USE ADD sfname.

Any current named lists are then undisturbed and may be deposited into file sfname via the SAVE command.

Automatic Text Access. The output-command software has been modified to take advantage of the recent hardware improvements which allow automatic switching from catalog mode to text mode. The command "output text" sends a special signal to the hardware to cause the switch in modes and to display the first page of the text of the first document in the current list of catalog items. Repeated output text commands will produce the text of each document in sequence until reset by a different kind of command or the end of the list is reached.

On the ARDS combined CATALOG/TEXT-ACCESS console, switching back to catalog mode is accomplished automatically by the typing of a command other than

"output text". On the Intrex console reversion to the catalog mode is accomplished by manually pushing the catalog mode button.

The previous output from the "output text" command was the fiche number of the document. Fiche number is now obtained by requesting "output 5" or "output fiche".

The fiche-directory format has been modified to handle fiche numbers up to 2048 fiche. Formerly, the limit was 1024 fiche.

Decoding of Special Symbols and Fields. Certain catalog fields which are coded in fixed binary format are now passed through decoding logic to print more meaningful output. For example, the language of the document would now be output as "English" instead of "e" or "French" instead of "f".

As a Master's thesis, Mr. N. Goto has generated a procedure for translating the special-character string representations into the appropriate font-shift codes and characters for the Intrex console. His procedure includes shifts for superscripts and subscripts. See Fig. D-1 for sample output showing how special characters appear on the console screen after decoding.

d 11266/0 73

1. D,11266-
(73) ferromagnetic resonance linewidth and relaxation in the system $Mn_xFe_3 - {}_xO_4 + \gamma$ for $x < 1$ measured at frequency 9.2 GHz in temperature range from 80 ° to 300 °K (1);
contents of Fe^{2+} ions effect in positions of pecks in linewidth in $Mn_xFe_3 - {}_xO_4 + \gamma$ system (2);
temperature dependence of the linewidth of $Mn_xFe_3 - {}_xO_4 + \gamma$ is principal directions in a (110) plane at 9.2 GHz (2);
Clogston's mode for hopping mechanism (3);
Clarke measurements on the ferromagnetic relaxations in the system $Mn_xFe_3 - {}_xO_4 + \gamma$ (2);
single crystals $Mn_xFe_3 - {}_xO_4 + \gamma$ prepared by Bridgman's method and by Verneuil's method (1);

Fig. D-1 Photograph of Display-Console screen showing catalog output with special characters.

Off-line Output Options. Offline output heretofore required a special version of Intrex. This capability has now been integrated with the regular Intrex as an option requested by a special argument to the "resume Intrex" command.

Console Control Procedures. When the Intrex retrieval system became available on an open basis in the Engineering Library it became important to prevent the system from being logged out because of inactivity and to control the exit from and resumption of the system by different users. A new command, BEGIN, was added so that a new user could begin with a re-initialized version of the system and the monitor files for the old user would be properly closed out. In order to close out an entire Intrex session and return to CTSS command level it is necessary to issue the command QUIT followed by an appropriate password. The requirements of a password prevents unauthorized use of Intrex consoles for CTSS use other than Intrex retrieval.

COMPUTER-SYSTEMS STUDIES

On the topic of computer-systems analysis we have begun an in-depth study of the storage and operating-time costs of our present system. This analysis will enable us to make the present system more efficient as well as help us to better specify how to design future systems. Of course, in any future system design recognition must be given to inherent differences that may exist among the computing systems upon which Intrex may be placed. A few comments from our recent studies are given below. They are made in the context of our present computing system, CTSS.

About 25 percent of the time consumed by the retrieval system goes into the reading and writing of various files on disk. Some disk Input/Output (I/O) is overlapped with computation and much more could be except for the limitations of CTSS in this respect. However, a number of things have been done to ensure that these I/O routines operate at their maximum efficiency. The catalog and inverted files are divided into segments and the length of a segment is chosen so that it can be easily handled by the file system. The file directory is sorted to minimize search time. The catalog uses a diagram encoding scheme which, by virtue of its compactness, improves access time.

A study is being made to determine how much each of the 314 subroutines in the Intrex system contribute to the system's operating cost. We have learned that five percent of these routines account for 90 percent of the subroutine calls made within Intrex. Most of the routines included in this five percent have already been reprogrammed in machine language. Our sophisticated console output package, TYPEIT, will be a prime target in our future efforts to increase the speed of the system. TYPEIT presently accounts for 30 to 40 percent of the operating time. It seems likely that this can be reduced to 20 percent.

The overlay mechanism adds about 10 percent to the operating execution time of the system. We have attempted to minimize this burden by watching the allocation of storage very closely. Several routines have been re-written to make them faster and more compact. Also, an elaborate scheme of assignment of areas for data is used so that areas of core do not remain idle.

A few statistics are given below to give an idea of how the retrieval system

performed for a sample set of 20 user sessions. These figures are only approximate and additional analysis must be done to refine them.

Allocation of time according to function

Performance of search requests	16 percent
Retrieval from catalog	26 percent
Interpretation of all user commands plus execution of commands other than search and output	44 percent
Miscellaneous operations (initialization, segment switching, etc.)	14 percent
	100 percent

Allocation of time according to internal operations

Disk I/O (input or writing and output or reading)	25 percent (approx.)
Console I/O (includes TYPEIT)	35 percent
Computation	40 percent (approx.)
	100 percent

Allocation of time according to specific task

Single access to inverted files	0.7 sec
Single access to catalog file	1.0 sec
Output of characters by TYPEIT	0.003 sec/character

E. DISPLAY CONSOLES

Staff Members

Mr. J. Bosco
Mr. J. E. Kehr
Mr. D. R. Knudson
Professor J. K. Roberge

Graduate Student

Mr. R. Goldschmidt

SUMMARY

The buffer/controller software for the Intrex display console has been modified to simplify the transition from catalog searching to full-text display. Previously three separate user inputs were needed to request text. These have been replaced by a single user command, which results in computer look-up of the fiche access number, a change from catalog to text-access modes in the buffer/controller, and an automatic transmission of the text request to the microfiche central store.

Program overlays for the catalog and text-access software have been developed to reduce the number of drum trucks read into core during mode changes. The software for pre-programmed messages used to instruct the user and display-program status has been modified to reduce the required drum storage and improve the message processing. Also, programs have been developed to write, read, and check records on a magnetic-tape cassette unit used as back-up storage for the buffer/controller.

Means for providing hard copy from both the Intrex display console and from an Advanced Remote Display Station (ARDS) are being implemented. On the Intrex console, a 35-mm camera has been mounted for recording images directly from the display. Processing and copying equipment available in the text-access system is used to develop the images and produce full-size copy. The ARDS hard-copy device is being constructed from an experimental unit developed previously by another research group at MIT. Dry-process silver paper is used to record images from a storage-tube display which is driven in parallel with an ARDS terminal. Further work on the thermal processor and paper transport is required before the device becomes operational. It is expected that a processor and paper transport unit be made available to the project for experimental use by an industrial manufacturer.

A second Intrex display console is being assembled with essentially the same design as the present one. A few minor revisions are being included primarily to exploit the progress in integrated-circuit components since the initial design was made. The buffer/controller is designed for a multi-console system, and the second display will demonstrate this capability, as well as provide another terminal for experimentation.

BUFFER/CONTROLLER SOFTWARE

Modifications and additions to the buffer/controller software during this report period fall into four areas: 1) Automatic entry into text-access mode of operation from catalog mode; 2) Two-track overlay for text/catalog modes of operation; 3) Improved processing of stored system messages; 4) Magnetic tape read/write/check programs for the cassette tape unit.

Automatic Mode Change. The buffer/controller monitor program for the Varian 620i computer controls three primary modes of operation: catalog, text-access, and edit. The catalog mode allows the user of the Intrex CRT display console to conduct catalog searches by issuing commands to the Intrex retrieval programs residing within the Central Time-Shared System (CTSS). The text-access mode allows the user to request a display image or a film copy from the microfiche document-storage subsystem. The edit mode allows local editing of a user's catalog material that he has stored on the magnetic-drum tracks assigned to him.

The 620i software has been modified to change automatically from catalog to text-access modes whenever a text request is issued by the user. Previously, the user obtained the fiche-access number from the catalog data base by typing the command "output fiche"; then he entered the text-access mode via a function switch. The mode change required reading the four drum tracks containing the text-access program into core. The 620i then instructed the user how to enter the fiche number through the keyboard and processed his request. Thus, three user actions were required to obtain text: request the fiche-access number, enter text-access mode, and re-enter the fiche-access number.

These three separate actions often resulted in user confusion and frustration, so that a single-step procedure was desirable. When the combined ARDS catalog and text-access display was designed (as described in the Text-Access Section of this report), the CTSS software was modified to transmit a specially formatted message in response to text requests. The 620i software has been revised to use this message for automatic mode changing and to permit the Intrex CRT console to respond in the same manner as does the combined console.

For both consoles the user conducts his catalog search in the usual manner. To request text he types the command 'output text'. CTSS then returns the access number of the first document on his current list along with certain control characters to the display. The receipt of these control characters causes the display hardware or software to relay the request, properly formatted, to the text-access subsystem. At this point the user becomes aware of the difference between the two consoles. For the combined display terminal the catalog information is erased and the text image is presented on the storage screen of the ARDS terminal. On the other hand, at the Intrex console a separate storage tube is located adjacent to the CRT screen for displaying

text images and the catalog dialog remains intact on the CRT display. The user then directs his attention to the separate text display. Thus, the 620i automatically enters the text-access mode and the user is not required to concern himself with the access number, although it does appear on the CRT display and is available for later use if the need arises. This availability of access number is particularly useful to experienced users if the text-access subsystem happens to be unavailable at that moment or if the user wishes to view the text at a later time without entering the CTSS retrieval program.

Both the Intrex CRT console and the ARDS combined catalog/text terminal have associated pushbutton panels to allow paging forward and backward in the text and magnification of page sectors. These operations are local to the display terminal and the text-access subsystem; CTSS is not utilized in any way once the initial request has been activated.

To return to the catalog for additional searches or to obtain the text of a new document, the user of the combined terminal simply types his next Intrex command which automatically puts him into communication with CTSS. Currently, the Intrex CRT console does not operate in the same manner; all keyboard actions when in the text-access mode are interpreted as text commands. To return to catalog mode in the Intrex terminal the user must press the catalog function switch on the keyboard which causes the 620i to recall the catalog programs from the drum into core.

One advantage of separate catalog and text displays is that both images remain intact and the user can view the catalog information and text at the same time. However, the availability of the catalog display when in the text-access mode sometimes causes users to forget the necessity to switch modes before resuming communications with CTSS. The manual-mode change is necessary in the present system because the keyboard is used for local data entry during the text-access mode, and any keyboard action cannot be used to resume catalog mode automatically, as in the combined terminal.

Both the Intrex CRT display and the combined ARDS/Text display can be used to display text images independently of CTSS if the fiche access number for a document is known. At the combined display the access number is entered via a set of selector switches, but at the Intrex CRT console it must be entered via the keyboard.

Text-Access Overlay. Some effort has been expended in rearrangement and modification of the text-access/catalog 620i software to isolate functions performed by both text-access and catalog modes. These common functions are incorporated into a new monitor program and associated subroutines. The monitor occupies one-half the 4-K core. All routines unique to the text-access and the catalog modes were organized into an overlay system so that the 2-K core unused by the monitor was used either by the text or the catalog programs at a given time. Thus, instead of bringing four drum tracks into core when changing modes, only a 2-track overlay is required. This results in a reduction of operating time and drum space. Implementing these modifications was

facilitated by the assembly-program improvements described in the 15 March 1970 Semiannual Report.

Stored-Message Processor. The 620i software provides messages and switch labels to inform the user of the program status and to instruct him in operation of the system. The programs which locate, format, and display these messages were modified so that messages may be packed on the magnetic drum and to allow messages to be compounded from status text in core and stored messages on the drum. Previously compound messages were written on the display character-by-character. Now they are compounded in core and written on the display a line at a time. This reduces screen blanking and greatly reduces operating time which is essential for multiple-console operation. Previously, all messages stored on the magnetic drum required a complete drum line regardless of the length of the message. In the modified program, messages may be packed several to a line. No message exceeds a drum line in length.

Magnetic-Tape Cassette Unit. Software to write, read, and check magnetic-tape files and records using a magnetic-tape cassette unit have been developed and put into operation. The tape is used in two ways. The current operating 620i monitor and text-access/catalog overlays occupy six drum tracks. An additional two tracks contain stored messages and programmable-switch labels. Copies of the eight tracks are stored as an eight-record file on a tape cassette. In the event of a system crash all or any portion of these tracks may be restored rapidly from magnetic tape in approximately one-tenth the time required to load from paper tape. Also stored in cassette files is the four-track assembly program and several special display pages retained for demonstration purposes.

The second use of the magnetic-tape unit is to store 620i source programs that were formerly on paper tape. The source programs are read into core and onto the magnetic drum. The 620i assembly program was previously modified to take the source input from drum tracks, as described in the last Semiannual Report.

As the 620i programs evolve, there is always a current 'working' system and a 'developmental' system. This usually means an additional six or eight drum tracks of code and text. A back-up for the developmental system is also maintained as a cassette-tape file.

CATALOG HARD COPY

As noted before, two types of cathode-ray-tube (CRT) terminals are used to access the Intrex catalog — the Intrex Display Console and the ARDS — and both lack a hard-copy capability. Since the end result of a catalog search might be a bibliography of relevant material, a permanent record would be highly desirable. In the past, the only alternatives were to copy manually the information from the CRT or to repeat the search at a teletype terminal. Currently two different means for producing hard copy directly from the CRT displays are being implemented.

Hard Copy from Intrex Console. A simple, inexpensive device has been assembled for recording images directly from the Intrex Console CRT. It utilizes a 35-mm single-lens reflex camera with a 28-mm lens mounted on pivoted slides, as shown in Fig. E-1.

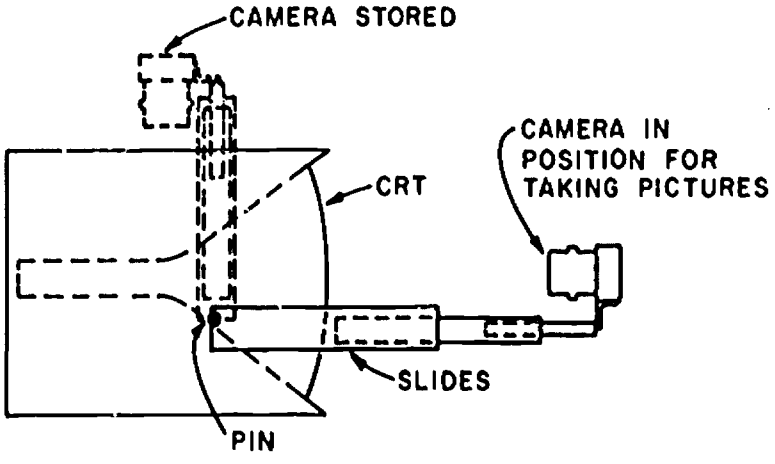


Fig. E-1 Side View of Camera Mounted on Intrex Console

With the CRT intensity set at a reference mark, an exposure of f/8 at 1/8 sec is used with Kodak No. 2498 recording film. After exposure, the film is developed and dried in approximately 1-1/2 minutes in the high-temperature liquid processor associated with the text-access film terminal. The developed film can then be used to make 8-1/2 x 11-inch copies on the modified electrostatic copier available in the text-access system. (The copier described in the 15 September 1969 Intrex Semiannual Activities Report.) The cost for the film plus the hard copy is estimated to be six to eight cents per frame. A sample page produced by the 35-mm camera and hard-copy device is shown in Fig. C-1.

Although this device is a bit cumbersome to use, in that it requires several manual operations to generate a full-size copy, it does produce legible copies quickly at reasonable cost. Because the film processor and 35-mm-to-8-1/2 x 11-inch copier were already available as part of the text-access system, the additional parts cost for this device was less than $200. From the user viewpoint a simpler solution for a multi-console system would be a line printer connected to the buffer/controller and shared by the displays. Then a single command could be used to generate a copy of any page stored in the buffer/controller.

Hard Copy from an ARDS Terminal. Development of a hard-copy device for the ARDS was initiated a few years ago under Project MAC at M.I.T.[1] An experimental unit demonstrated the feasibility of furnishing full-size copy quickly and conveniently from a storage-tube display using dry-process silver paper. The configuration of this unit is illustrated in Fig. E-2.

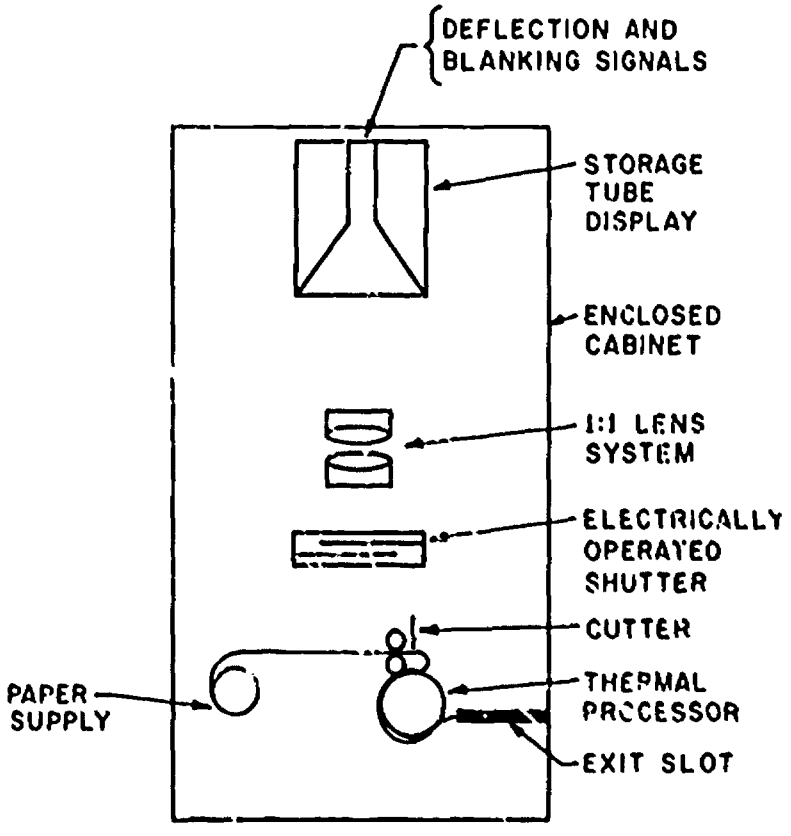


Fig. E-2 Diagram of Experimental ARDS Hard-Copy Unit

With Kodak type 1991 dry silver paper and an f/2.2 lens aperture, an exposure of approximately 10 seconds is required to record an image from a Tektronix-611 Storage Display Unit. After exposure the paper is stabilized by heating it to 230°C for about 5

[1] Vezza, Albert, "Hard Copy from a Direct View Storage Tube Remote Display Terminal", Proceedings of the 1970 IEEE International Computer Group Conference, June 1970.

seconds. Then the paper is transported through an exit slot and the image is developed by exposure to ultraviolet light. Development requires 20 to 30 seconds under an ordinary fluorescent lamp, but a light source with a higher ultraviolet content can reduce the development time to a few seconds. Once stabilized and developed the image is maintained for several months in normal room lighting.

This device was never refined beyond the experimental phase. It has recently been made available to Intrex for refinement and use in an operational environment. Some modifications and improvements are necessary, particularly to the thermal processor and paper transport, before making the unit available to users. The Eastman Kodak Corporation has offered to make available on loan an improved thermal processor which is a result of considerable design effort on their part. This greatly simplifies our development effort because the processor is the most critical component. While awaiting the Kodak processor, a circuit for timing the exposure automatically has been designed and special ultraviolet fluorescent lamps have been ordered. When operational, the storage tube in this device will be slaved to an ARDS display and the user may request hard copy of any page appearing on his display simply by actuating a button.

INTREX-CONSOLE HARDWARE

Considerable time has been spent in refinements and modifications to the present Intrex console in order to improve its reliability and picture quality. Some of the more significant modifications are listed below:

1. Construction of a 5-volt regulator board for the phase-locked loop. This unit replaces a temporary power supply which had been installed previously to reduce the effect of transients in the console power supplies.

2. Modifications to the phase-locked loop, horizontal deflection, and power-supply circuitry to provide more stable operation.

3. Additional circuitry to furnish the light-pen address to the buffer/controller. The light-pen capability has not yet been fully implemented because of higher priorities given to other tasks.

Operating experience with the Intrex console has provided sufficient confidence in its design to begin assembling a second console. The basic design will remain unchanged although certain refinements will be added to improve reliability and performance, and to reduce cost. These include:

1. A solid-state keyboard for improved reliability.

2. Advanced integrated circuits to reduce the number of packages required to perform a given function.

3. An improved interface between the display console and the buffer/controller.

F. FULL-TEXT STORAGE AND RETRIEVAL

Staff Members

Mr. P. Campoli
Mr. R. Chin
Mr. J.E. Kehr
Mr. D.R. Knudson
Professor J.K. Roberge

Student Employee

R. Prakken

SUMMARY

A display terminal that provides access to both the computer-stored catalog and the microfilm-stored text of the Intrex data base by means of a single display tube is now operational. This combined terminal merges the Advanced Remote Display Station (ARDS) and the Stand-Alone Text-Access Terminal into a common storage-tube display. Upon receipt of a user's request for text, the computer furnishes the microfiche access number which is automatically transmitted through the terminal to the text-access system for retrieval and display of the first page. Subsequently, the terminal functions as a stand-alone display. Any message transmitted by the computer or a key actuated at the keyboard causes the terminal to revert to the ARDS mode for computer-user interaction.

Minor modifications and improvements have been installed in the 35-mm film terminal in preparation for moving it into a user environment. The flying-spot scanner cathode-ray tube failed and was replaced during this report period.

COMBINED ARDS/STAND-ALONE TERMINAL

A terminal has been designed, fabricated and put into operation which displays catalog information and full text on the same cathode-ray tube. The terminal combines an ARDS computer terminal and a stand-alone text-access terminal. The ARDS is a commercially-available graphics-display terminal using a direct-view storage tube. The stand-alone text-access terminal also uses a storage-tube display as described in the 15 September 1969 Semiannual Report. Through combination of these units, only one storage tube is required and the user can view catalog information and full text on the same display.

The combined unit is operated initially from the ARDS keyboard. During catalog searches the terminal communicates in its usual manner with the time-shared computer and the catalog information is displayed on the storage tube. To request the text of a document the user issues an "output text" command. The computer responds by looking-up the document's microfiche access number and sending a formatted message to the display terminal. The format includes special control characters which allow the terminal to recognize it as a text request. The

fiche-access data are transferred to the stand-alone terminal's shift register and the terminal enters the stand-alone mode. The data are transmitted to the central station during the next polling service and the fiche is retrieved and scanned. Solid-state switches in the display terminal are used to transfer control of the storage-tube erase, write-through, and x, y, and z axes between the ARDS and the stand-alone electronics.

After the first page of text is retrieved and displayed, subsequent requests are entered through the stand-alone pushbutton panel. This allows requests such as next page, previous page, magnify, film copy, and so forth, to be transmitted directly through the text-access system without imposing any delays from the time-shared computer. However, the terminal is transferred automatically out of the stand-alone mode into the ARDS mode whenever any new message is received from the computer or a key is actuated on the ARDS keyboard. Thus, computer messages are always displayed and computer communication can be resumed for additional catalog searching by simply typing on the keyboard. A pushbutton on the stand-alone panel provides a manual override of the mode control and permits the terminal to be operated as a stand-alone text-access display when the computer is not available.

One combined terminal is presently in operation and the availability of both catalog information and text on the same display has proved to be quite effective. Because the combined terminal was designed with a minimum of changes to the standard ARDS and stand-alone electronics, any existing pair of terminals can be converted readily by adding the interface electronics.

FILM PROCESSOR

The 35-mm film terminal will be moved in the near future from the Laboratory to the Materials Sciences Building where it will be more accessible to users. Some modifications and refinements have been added to facilitate routine operation in a user environment and to improve image quality.

The electronics were revised to simplify the turn-on procedure by assuring the proper logic states when power is applied. This allows the terminal to be turned on-and-off with a single switch and eliminates the initial cycling and resetting formerly required. An "out-of-film" indicator was installed on the control panel.

Some minor modifications were made to the film-processor transport to improve its reliability. Also, a device was added for loading film from 35-mm cassettes into the processor. This is used to develop film from the Intrex-Console camera described in the console section of this report.

An amplifier with logarithmic-gain characteristics has been installed in series with the film-terminal video amplifier to study the effectiveness of non-linear amplification of the video signal. The intent is to improve image quality by emphasizing the small video signals associated with fine detail lines. The logarithmic amplifier provides higher gain for low-level signals. Initial evaluations indicate more

consistent image quality with less dependence on the font of the scanned document. The signal-to-noise ratio in the video signal limits the degree of image enhancement that can be achieved.

CONTROL-STATION SCANNER

The cathode-ray tube of the central-station flying-spot scanner failed and was replaced during this report period. The failure was due to cathode degeneration which may have been aggravated by high-voltage arcing during its initial operation approximately 1-1/2 years ago. It was replaced with a different tube model, the Litton L4123, which provided good performance in the scanner during the early text-access experiments and in the film terminal for several years. A P-37 phosphor is used again in the new tube because of the compatibility of its spectral output with the scanner optics.

III. MODEL-LIBRARY PROJECT (Model-Library Staff)

A. STATUS OF THE PROJECT

Mr. C. H. Stevens

In the period covered by this document, the Model-Library Project personnel have taken significant steps toward the goal of providing the necessary transition elements from the traditional library to the information transfer system that could be the outgrowth of the experimental programs described in Section II of this report. Emphasis since March has been on production of Pathfinders and point-of-use instruction scripts and visual materials. Secondary attention has been given to the use of these items in the Barker Engineering Library because students are not present in large numbers from mid-May until mid-September. Work has gone forward on user preference studies for microfiche or hard copy and on preparation for visits by librarians during the second year of the Project. There is nothing new to report on use of data bases prepared by NASA or Engineering Index.

To accomplish the stepped-up level of activity in the Project, three additional professional persons were employed for the summer months. Two were librarians, and one an audio-visual specialist. When this report is distributed, these people will have left us to return to former duties; one has been replaced on a temporary basis.

Equipment needed to begin point-of-use instruction in the fall is on order and scheduled for delivery in time for use at the beginning of the school year.

B. POINT-OF-USE INSTRUCTION

Mr. C. H. Stevens
Miss M. P. Canfield
Miss T. E. R. Carsten
Mr. J. J. Gardner

Before beginning production of point-of-use materials, two basic decisions had to be made. We had to choose which library tools we would include in our program and which media best suited our material.

The procedure for selection of library tools consisted of a library user questionnaire (mentioned in the March 1970 report), in-depth interviews with members of the Barker Engineering Library reference staff, and observation of library users. The criteria were: frequency of use and difficulty in use. Application of these criteria would assure us of producing instructional aids with a large body of potential users, and the best chance for significant feedback.

We selected for our first five programs the author-title catalog; the subject catalog; the Engineering Index, NASA STAR; and Science Abstracts, Section B: Electrical and Electronics Abstracts.

The selection of media and media programs to be used was based on the following criteria:

1. The media material must be suitable for individual instruction.

2. The programs must be brief and to the point. We will not have a captive audience.

3. The media used should be those in which we can get a maximum of information to the user in a lively manner in a short period of time.

4. The programs must be simple enough to be produced in-house at low cost.

5. The programs should be exportable to other institutions. The media used should, therefore, be non-esoteric.

With these basic criteria in mind, we began with a comprehensive study of the literature. This study indicated that various approaches to the library orientation problem had been tried with uncertain success. The most frequent approach had been the general orientation program — usually either a tour, a lecture, or, in a few instances, a film. The inadequacy of this approach seems to be that it can only successfully serve as a general introduction to the range of services available in a library. Were it to detail specific reference tools, it would bombard the listeners with much that is irrelevant to his current needs. Were he to research a subject some months, or even weeks later, he would in all probability have to learn how to use a reference tool anew.

Other approaches were closer to ours. Mt. San Antonio Junior College in Walnut, California, is developing a single location, sound-filmstrip viewer capable of holding four programs on different aspects of using their library. Monsanto Chemical Company Research Center in St. Louis has two sound-slide programs describing their library and their technical reports library. The University of Toronto has done one individualized listener sound-slide presentation as a general introduction to their library and tentatively plans more.

Other libraries have done audio presentations for certain tools — most often the card catalogs, but scripts have tended to be lengthy and pedantic.

With this background we began talking with people in the audio-visual field — commercially and/or educationally. Discussions with various manufacturers of audio-visual equipment and with Dr. Philip Sleeman at Boston University's Graduate School of Education led us toward an audio-visual rather than audio only approach, and commercial production costs led us to our present focus on in-house production.

Currently, we are producing two types of audio-visual materials: synchronized sound-slide and synchronized sound-filmstrip. Each has advantages and drawbacks. Sound-slide productions have the advantage of being easily exportable, requiring only a Carousel projector and standard 1/4" tape player. (For permanent individual use, more elaborate equipment is required.) In addition, sound-slide presentations are simple to revise and relatively simple to produce.

Sound-filmstrips are not as easy to produce or revise as sound-slide presentations, and their individual equipment requirements make them less acceptable to other libraries. Filmstrips are more compact than slides thus making the projection equipment more compact. This is significant in practically any library environment.

Our next step was the production of written scripts for sound-slide and sound-filmstrip programs. We considered three factors during this stage: (1) Give the user only what he is likely to want to know at an early stage of using a library tool. For example, it is unlikely that a library user cares a great deal about who published a book, or where it was published or how many centimeters high it is. We want him to understand how to get into a library tool. Much of the detail will come with his working with it. (2) Keep each script as short as possible. The judicious use of visuals should cut down the verbiage. (3) The tone should not be pedantic; but rather conversational and lively. The largest number of our users are students, and we have tried to talk in their vernacular. The visuals should reflect this tone.

The scripts have been prepared by the Model Library Project staff with valuable consultation with Howard Canning, a professional technical writer and editor at M.I.T.'s Lincoln Laboratory. The first draft of the first script — on the subject card catalog — had a listening time of just over 8 minutes. Extensive editing and rewriting brought the finished product to a listening time of just over two minutes. Subsequent scripts have gone through much the same process, and the completed scripts have a listening time of from two to four minutes. It is expected that final addition of visuals will add from 30 to 60 seconds to each script. Our future tests with users will tell us if these times are acceptable.

Below are listed the completed scripts with their recording times without visuals:

1.	Subject Card Catalog	2 min/40 sec
2.	Author-Title Catalog	3 min/15 sec
3.	Engineering Index	3 min
4.	NASA STAR	4 min
5.	Science Abstracts: Section B	4 min

In addition, we are preparing one script in Chinese for M.I.T.'s sizable Chinese student population and will try to determine whether their language problems are significant enough to warrant dual programs. The library reference staff has

indicated this as a possibility and Miss Lucy Lee, an Intrex cataloger, has translated and recorded the subject catalog script. The following is a sample script with sample slides (Figs. III-1, III-2, III-3, III-4).

Engineering Index

It sometimes seems as though libraries have been arranged into a kind of psychological experiment. You wander through the maze, rewarded at each correct turn with bits and pieces of information — some of it useful, some of it not so useful. But, hope springs eternal in the librarian's breast, and it's our current hope that this two-minute tape will decrease your frustration while it increases your reward.

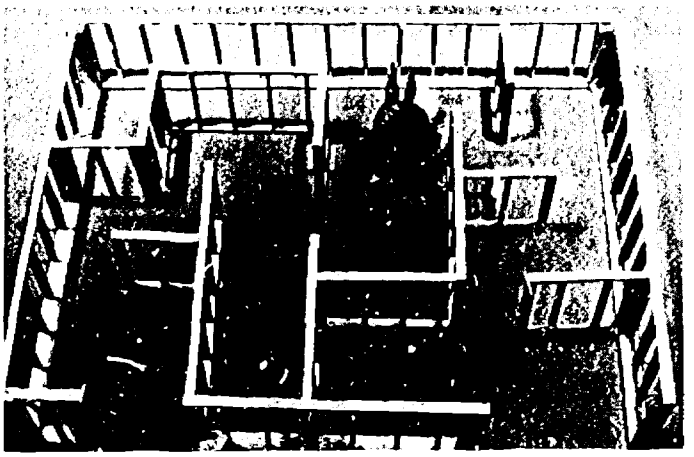


Fig. III-1 Black and white reproduction of color slide from Engineering Index program. "You wander through the maze, rewarded at each correct turn ..."

The Engineering Index is an annual publication with monthly supplements for the current year. Each volume or supplement presents summaries of published material in the engineering and related sciences. It lists journal articles, books, professional papers and report literature. They're listed alphabetically by subject and they appear in Engineering Index some 6 to 12 months after they have appeared in print.

Although Engineering Index is a classic reference tool, it's not as easy to use as it might seem. The generality of its subject headings will probably be your biggest hangup. Having found the relevant subject headings, you're likely to feel as though you're at a college mixer, sifting through a multitude of entries to find the few chicks

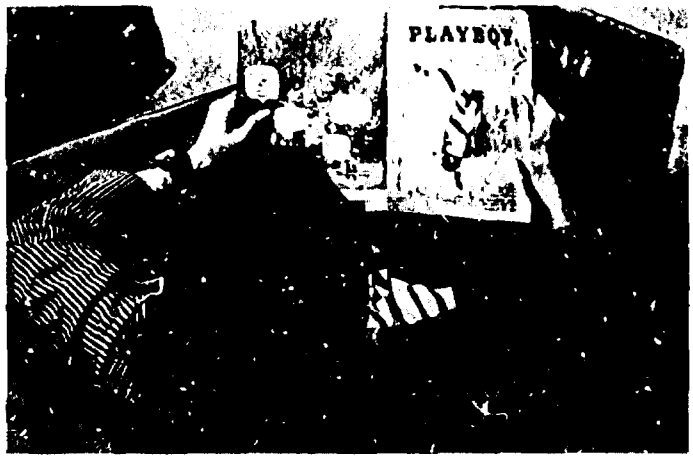


Fig. III-2 Black and white reproduction of color slide from Engineering Index program. "... it's our current hope that this two-minute tape will decrease your frustration while it increases your reward."

or even the one chick with the best potential. The main subjects, however, are often broken down into more specific subjects and sometimes this will make life easier for you. If, however, you're unable to locate a relevant subject heading, try broader, more specific, and synonymous terms.

Occasionally, you'll want to locate material by a specific author. If the author is represented in the author index in the back of any issue of Engineering Index, you'll be referred to a specific page. Turning to that page, scan it for the author's name, which is always printed in all capital letters.

When you find an item of interest, you'll note that the listing includes:

the title
the author
where and when the item was published
a brief summary or abstract of the item

Listings of journal articles also include:

the journal title
the volume number
the date of the journal
the relevant pages

To find the full text of an item, jot down all the information except the summary. If abbreviations of journal titles or engineering terms hassle you, check in the front of any of the annual volumes to decipher them. Then check the library's author-title card catalog to learn if the material you want is in our collection. The wall directories throughout the library will point you in the right direction. If you get hung up anywhere along the line, ask for help.

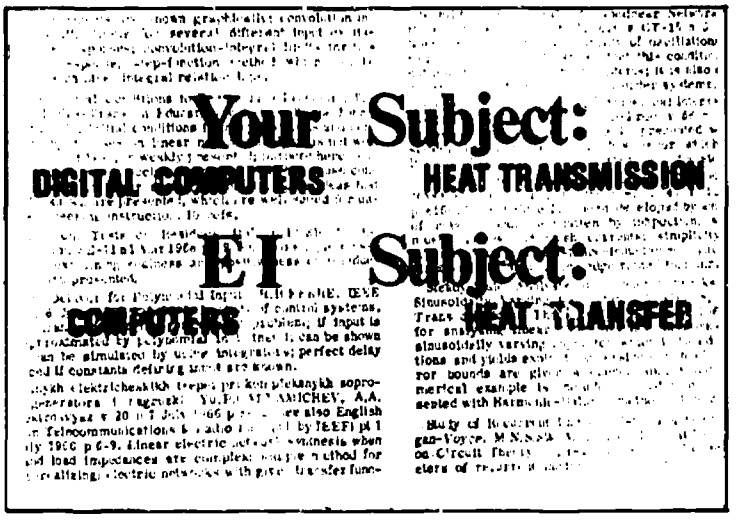


Fig. III-3 Black and white reproduction of color slide from Engineering Index program. "Having found the relevant subject headings ..."

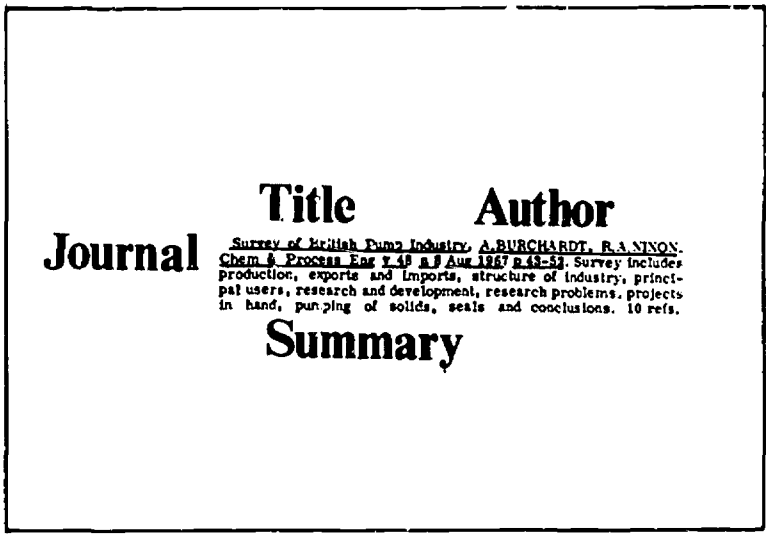


Fig. III-4 Black and white reproduction of color slide from Engineering Index program. "Listings of journal articles also include the journal's title, the volume number, the date of the journal, and the relevant pages."

We then began preparation of the visual material for our five working scripts. A staff member with audio-visual educational background and broad experience with graphics and photography has undertaken the entire scope of production except film processing (which is being done commercially) and final narration which will be done at either the campus radio station or in a professional recording studio.

The production of visuals has consisted of the following steps:

1. Preliminary layout with sketches.
2. Minor script revisions to work in with visuals.
3. The production of the art work. Most of this has been on acetate sheets for copy work.
4. Actual photography — both slides and filmstrips.
5. Processing of the film by Kodak.
6. Checking processed film with scripts.
7. Retakes when necessary.
8. Processing of retakes by Kodak.

The photography for the first five scripts has been completed. It is ready now for final editing, final narration, synchronization of audio and visual, and adaptation to rear screen projection equipment.

The selection of projection equipment has been a major problem. Our equipment must be rear screen for individual use; must be small enough to fit into the Barker Engineering Library environment; must be portable; simple to use; lockable; and must be economical for use in university libraries.

Of these criteria, perhaps the most important is ease of operation. There is little question that the typical M.I.T. user will not choose to spend five minutes figuring out how to use a self-instruction device. Ideally, the user should be faced with one button, which turns on the power and begins the program. At the program's end it should automatically rewind and shut down, ready for the next user.

These features are available, with minor alteration, in various sound-filmstrip projectors, and the one we are working with initially is the LaBelle Sentinel which utilizes a continuous loop tape cartridge and continuous loop film magazine. It is easily adapted to single button operation; has headphone plug; is portable; and can easily be adapted to shut down at a program's conclusion. The continuous loop arrangement does away with the need for a rewind. The LaBelle Sentinel retails for $365. Its dimensions are 18-1/4 x 12-1/4 x 15" and its screen size is 8-3/4 x 12". Minor adaptations are being carried out by their Boston dealer, and two will be in use in the Barker Library this fall.

Sound-slide equipment raises more problems. We have found no rear screen, individualized use sound-slide projection units suited to our needs and we have been led to design and fabricate our own.

The major design problem is to combine a rear projection screen, carousel slide projector, and synchronized sound playback unit in a cabinet small enough to fit into our library testing environment. Since we hope to be able to place at least some of the units on bookshelves near the reference tools being discussed, size restrictions are dictated by bookshelf size. Working with maximum dimensions of 30 x 11 x 12", we have completed a design which is being used by New England Film Service in Waltham, Massachusetts, to fabricate a test device. This unit is a horizontal projection unit, adapted to the shape of a standard bookshelf.

In addition, Laurence Associates of Boston is designing for us a vertical sound-slide projection unit for table-top use, which will take less table space but be too tall for use on bookshelves. These units will be tested with our catalog instruction productions.

Our work in the future will fall roughly into three areas. We will continue to produce programs, some for audio use only, and some super 8 sound cartridge productions on the U. S. Patent Gazette; Science Citation Index; math tables; and handbooks.

More important though will be the installation of programs into the Barker Engineering Library and the testing of their effectiveness. This testing will rely heavily on subjective evaluation — that is, an active solicitation of comments from actual users. In addition, we are considering the possibility of controlled experiments to test the learning that takes place through the programs. A defect of many library orientation programs has been a tendency to assume that they are actually helpful with no solid attempt to test this assumption. Within the scope of our objectives — to introduce users to specific library research tools — we will test our level of success.

Finally, we plan to share our results. We have available, now, written scripts and cassette recordings of our scripts for loan and/or retention by other interested libraries and will have available for loan this fall complete sound-slide and sound-filmstrip productions. In most cases the sound-slide productions will be best for this purpose since the equipment needed is more commonly available.

C. LIBRARY PATHFINDERS

Mr. C. H. Stevens
Miss M. P. Canfield
Mr. J. J. Gardner
Miss A. Jackson
Mrs. E. King

In addition to the initial group of mechanical engineering Pathfinders that were listed in the last semiannual report, fifty-six titles have been prepared for computer technology, civil engineering, naval architecture and marine engineering, and pollution. Titles now available are:

Heat Transfer

Bubbles
Film Boiling
Heat Conduction
Heat Convection
Heat Transfer - Absorptivity
Heat Transfer - Emissivity
Nucleate Boiling
Plate-Fin Heat Exchangers
Pool Boiling
Radiation Heat Transfer
Shell-and-Tube Heat Exchangers
Thermal Contact Resistance
Thermal Pollution
Thermal Regenerators
Thermal Stresses
Two Phase Flow

Fluid Mechanics

Boundary Layer Control
Boundary Layer Flow
Boundary Layer Separation
Cavitation
Flow-Induced Vibrations
Fluidics
Laminar Boundary Layer
Noise Attenuation
Thermal Boundary Layer
Turbulent Boundary Layer

Computer Technology

Analog Simulation
Analog-to-Digital Converters
Artificial Intelligence
Automata Theory
Cathode-Ray-Tube Display Devices
Cybernetics
Digital Simulation
Electronic Analog Computers

Computer Technology (cont.)

Electronic Digital Computers
Heuristic Programming
Holography
Hybrid Computers
Logic Design
Optical Character Recognition
Queuing Theory
Time Sharing

Civil Engineering

Airport Design
Coastal Engineering - Erosion
Earthquake Engineering
Foundation Engineering
Ground Water Seepage
Harbor Design
Highway Engineering
Rock Fracture/Failure
Salinity Intrusion
Sedimentation Transport
Soil Freezing
Soil Instrumentation
Soil Stabilization
Thermal Stratification
Traffic Flow
Tunnels
Water Distribution Systems
Water Drainage Systems

Pollution

Air Pollution - Carbon Monoxide
Air Pollution - Chimneys/Stacks
Air Pollution - Cyclones
Air Pollution - Inversion Layers
Air Pollution - Plumes
Air Pollution - Radioactive Materials
Air Pollution - Scrubbers/Scrubbing
Air Pollution - Smog

Pollution (cont.)

Air Pollution - Sulfur Compounds
Solid Waste Disposal - Composting
Solid Waste Disposal - Incineration
Solid Waste Disposal - Sanitary Land Fill
Water Pollution - Detergents
Water Pollution - Pesticides
Water Pollution - Phosphates

Naval Architecture/Marine Engineering

Air Cushion Vehicles
Deep Sea Submergence Vehicles
Free Surface Hydrodynamics
Hydrofoil Vehicles
Marine Power Systems
Marine Sonar Systems
Oil Pollution

At present, Pathfinders are being completed on topics relating to materials science and engineering. Some of these topics are:

Ferroelectrics
Ferromagnetism
Fiber Composite Materials
Mössbauer Effect
Raman Effect
Semiconductors
Superconductivity

In addition, topics are being considered for projected series of Pathfinders in other areas of interest, e.g., bio-engineering and laser/maser technology.

The preparation of Pathfinders has been given high priority during this last reporting period. The goal has been to have available by mid-September a series of topics relevant to the interests of the five engineering departments which the Barker Engineering Library serves. To this end, two members were added to the Model-Library staff to work exclusively on Pathfinders during the summer period. The additional staff members had library school training and some library experience. They did not have engineering subject backgrounds or experience in technical library work.

Initially, each of the new staff members was assigned to prepare a set of Pathfinders in a specific area. One worked on computer technology; the other, civil engineering. Because computer technology was a more difficult area to approach, these Pathfinders took somewhat more time to prepare than those in civil engineering. After a preliminary two- to three-week training period, the average time involved in compiling a Pathfinder was seventeen hours.

Pathfinder compilation includes library catalog searching, inspection and selection of material, preparation of handwritten copy for typing, and proofreading of final typed copy. Detailed procedures and guidelines for the generation and production of Pathfinders have been prepared by the Model-Library staff and are available from Project Intrex.

The mechanics of compiling and typing a Pathfinder involve six steps:

1. Bibliographic and location information notes on materials to be included on the Pathfinder are made on individual 5 x 8-inch index cards.

2. The information on cards is transferred to a five-page worksheet. The worksheet is preprinted with the introductory statements to the Pathfinder sections arranged and adequately spaced in an order corresponding to the final Pathfinder form (Fig. III-5).

3. The worksheet entries, with the exception of location information, are typed on an oversize master in a columnar arrangement; however, spaces are allotted for call numbers and locations. This typing procedure takes 1-1/2 hours when done by an experienced typist who is familiar with the format.

4. Fifty oversize masters without location information are offset-printed from the typing master (Fig. III-6).

5. Barker Engineering Library location information from the worksheet is typed on one of the oversize printed masters (Fig. III-7). The remainder of the printed masters are made available to other libraries for use with their collections. The insertion of location information takes 1/2 hour typing time.

6. The printed master with call numbers is photoreduced and one hundred 8-1/2 x 11-inch Pathfinders are printed on notebook-punched cover stock for use in the Barker Engineering Library (Figs. III-8 and III-9).

A cooperative program of Pathfinder compilation has developed between the Model-Library staff and the Simmons College Graduate School of Library Science. Students enrolled in the Literature of Science and Technology course, conducted by James Matarazzo of the Simmons faculty, were assigned to prepare Pathfinders on topics supplied by the Model-Library staff. The topics dealt with aspects of pollution and were selected for their relevance to M.I.T. curriculum and research interests. Because of the proximity of Simmons to M.I.T., the students were directed to work with the Barker Engineering Library collection. Each of the final student-compiled Pathfinders required an average of five hours' editing time by the Model-Library staff. To date, fifteen Pathfinders in the area of pollution have been printed in final form. This cooperative program with Simmons will continue and it is projected that fifty additional Pathfinders will be prepared during the academic year.

Within M.I.T., the Model-Library staff has publicized the Pathfinders in a variety of ways. The staff has contacted faculty subject specialists to explain the purpose of the Pathfinders and to submit to them finished products for comment. Illustrative, informative posters have been placed on bulletin boards in the Barker Engineering Library, in departmental office areas, and in laboratory areas. Sets of Pathfinders have been placed in office and laboratory areas. Notices announcing the availability of Pathfinders have been printed in the Engineering Library Bulletin and in departmental Newsletters. These publicity methods will continue as new groups of Pathfinders are prepared.

1. SCOPE

2. An introduction to this topic appears in

3. BOOKS dealing with are listed in the subject card catalog. Look for the subjects:

4. Frequently mentioned texts include:

5. Other books including material on are shelved under call numbers:

6. HANDBOOKS, ENCYCLOPEDIAS, and DICTIONARIES which contain information on are:

7. BIBLIOGRAPHIES which contain material on include:

8. JOURNAL ARTICLES and other literature on are indexed primarily in the guides listed. The quoted subject headings are those in use since 1965 unless other dates are given.

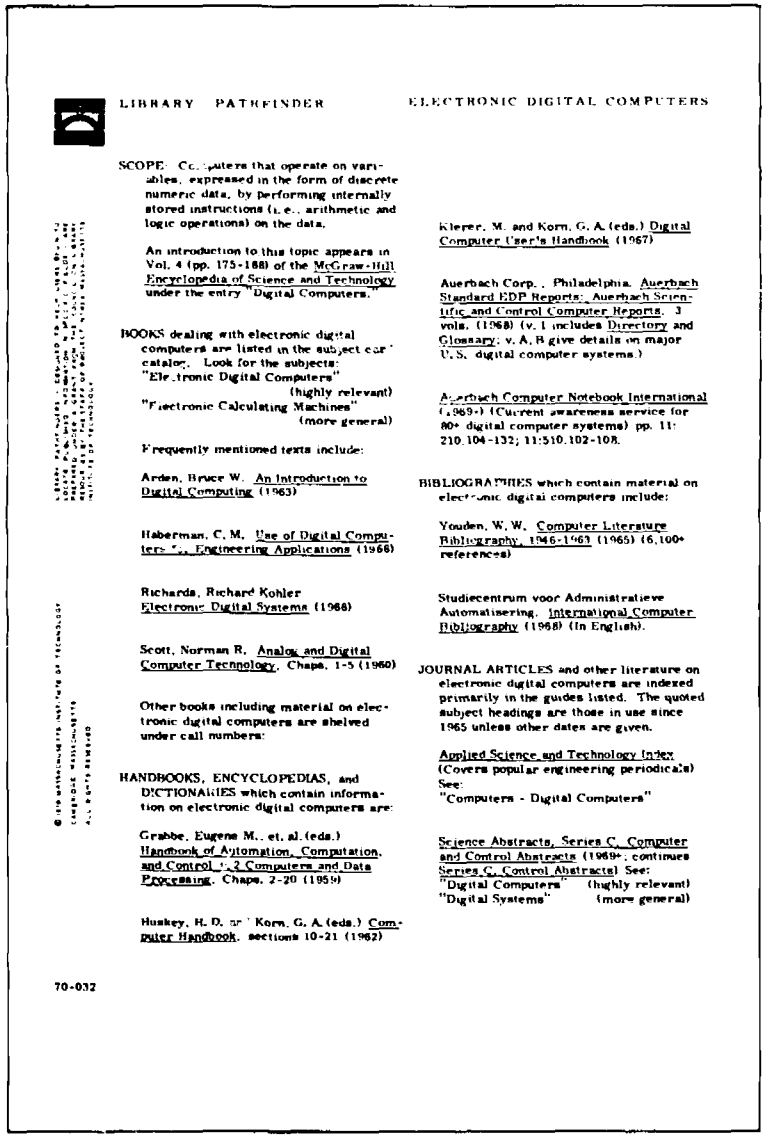
9. Other indexes, listed here, should be used for an exhaustive search. Only a limited return can be expected for the time spent. Directions are generally given in the front of each issue.

10. JOURNALS that often contain articles relevant to are:

11. STATE-OF-THE-ART REVIEWS and CONFERENCE PROCEEDINGS containing material on include:

12. REPORTS and other types of literature are indexed in these guides:

Fig. III-5 Five-page worksheet, with sectional introductory statements, condensed here to one page.



LIBRARY PATHFINDER — ELECTRONIC DIGITAL COMPUTERS

SCOPE: Computers that operate on variables, expressed in the form of discrete numeric data, by performing internally stored instructions (i.e., arithmetic and logic operations) on the data.

An introduction to this topic appears in Vol. 4 (pp. 175-188) of the McGraw-Hill Encyclopedia of Science and Technology under the entry "Digital Computers."

BOOKS dealing with electronic digital computers are listed in the subject card catalog. Look for the subjects:
"Electronic Digital Computers" (highly relevant)
"Electronic Calculating Machines" (more general)

Frequently mentioned texts include:

Arden, Bruce W. An Introduction to Digital Computing (1963)

Haberman, C. M. Use of Digital Computers for Engineering Applications (1966)

Richards, Richard Kohler Electronic Digital Systems (1966)

Scott, Norman R. Analog and Digital Computer Technology. Chaps. 1-5 (1960)

Other books including material on electronic digital computers are shelved under call numbers:

HANDBOOKS, ENCYCLOPEDIAS, and DICTIONARIES which contain information on electronic digital computers are:

Grabbe, Eugene M., et. al. (eds.) Handbook of Automation, Computation, and Control. v. 2 Computers and Data Processing. Chaps. 2-20 (1959)

Huskey, H. D. and Korn, G. A. (eds.) Computer Handbook. sections 10-21 (1962)

Kiener, M. and Korn, G. A. (eds.) Digital Computer User's Handbook (1967)

Auerbach Corp., Philadelphia. Auerbach Standard EDP Reports; Auerbach Scientific and Control Computer Reports. 3 vols. (1968) (v. 1 includes Directory and Glossary; v. A, B give details on major U.S. digital computer systems.)

Auerbach Computer Notebook International (1969-) (Current awareness service for 80+ digital computer systems) pp. 11: 210.104-132; 11:510.102-108.

BIBLIOGRAPHIES which contain material on electronic digital computers include:

Youden, W. W. Computer Literature Bibliography, 1946-1963 (1965) (6,100+ references)

Studiecentrum voor Administratieve Automatisering. International Computer Bibliography (1968) (In English).

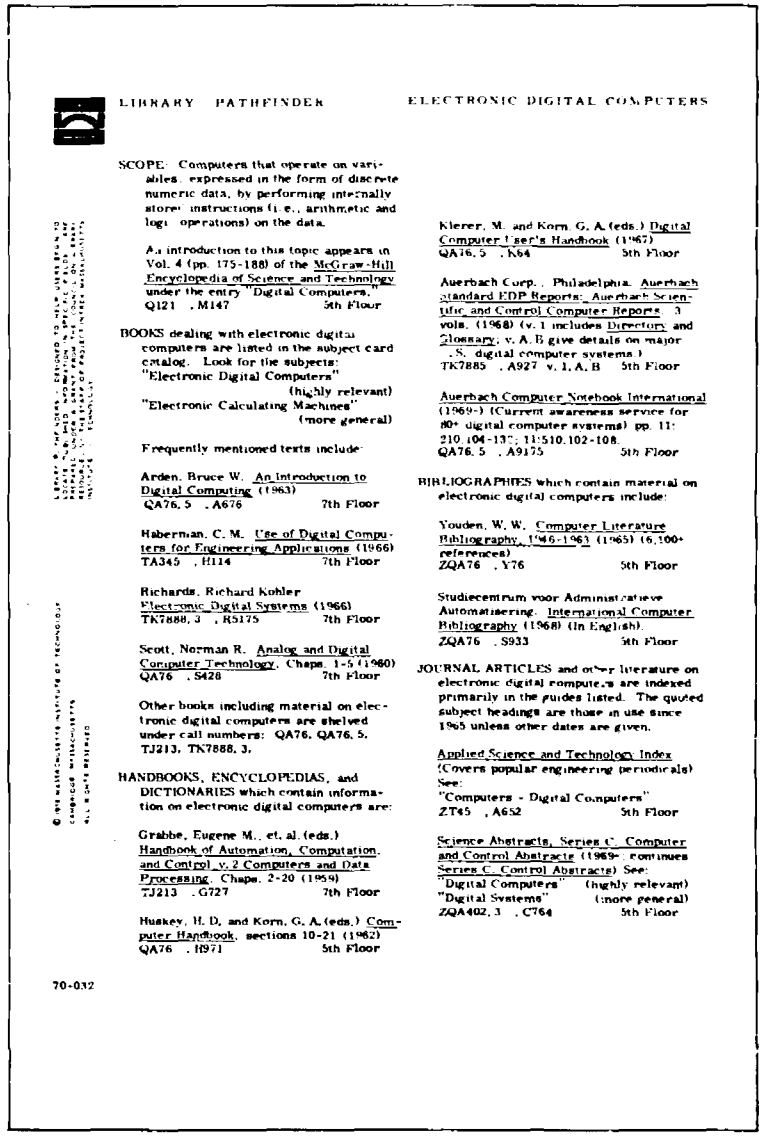
JOURNAL ARTICLES and other literature on electronic digital computers are indexed primarily in the guides listed. The quoted subject headings are those in use since 1965 unless other dates are given.

Applied Science and Technology Index (Covers popular engineering periodicals) See:
"Computers - Digital Computers"

Science Abstracts, Series C, Computer and Control Abstracts (1969-, continues Series C, Control Abstracts) See:
"Digital Computers" (highly relevant)
"Digital Systems" (more general)

70-032

Fig. III-6 Printed master without call numbers and floor locations.



LIBRARY PATHFINDER — ELECTRONIC DIGITAL COMPUTERS

SCOPE: Computers that operate on variables, expressed in the form of discrete numeric data, by performing internally stored instructions (i.e., arithmetic and logic operations) on the data.

An introduction to this topic appears in Vol. 4 (pp. 175-188) of the McGraw-Hill Encyclopedia of Science and Technology under the entry "Digital Computers."
Q121 .M147 5th Floor

BOOKS dealing with electronic digital computers are listed in the subject card catalog. Look for the subjects:
"Electronic Digital Computers" (highly relevant)
"Electronic Calculating Machines" (more general)

Frequently mentioned texts include:

Arden, Bruce W. An Introduction to Digital Computing (1963)
QA76.5 .A676 7th Floor

Haberman, C. M. Use of Digital Computers for Engineering Applications (1966)
TA345 .H114 7th Floor

Richards, Richard Kohler Electronic Digital Systems (1966)
TK7888.3 .R5175 7th Floor

Scott, Norman R. Analog and Digital Computer Technology. Chaps. 1-5 (1960)
QA76 .S428 7th Floor

Other books including material on electronic digital computers are shelved under call numbers: QA76, QA76.5, TJ213, TK7888.3.

HANDBOOKS, ENCYCLOPEDIAS, and DICTIONARIES which contain information on electronic digital computers are:

Grabbe, Eugene M., et. al. (eds.) Handbook of Automation, Computation, and Control. v. 2 Computers and Data Processing. Chaps. 2-20 (1959)
TJ213 .G727 7th Floor

Huskey, H. D. and Korn, G. A. (eds.) Computer Handbook. sections 10-21 (1962)
QA76 .H971 5th Floor

Kiener, M. and Korn, G. A. (eds.) Digital Computer User's Handbook (1967)
QA76.5 .K64 5th Floor

Auerbach Corp., Philadelphia. Auerbach Standard EDP Reports; Auerbach Scientific and Control Computer Reports. 3 vols. (1968) (v. 1 includes Directory and Glossary; v. A, B give details on major U.S. digital computer systems.)
TK7885 .A927 v. 1, A, B 5th Floor

Auerbach Computer Notebook International (1969-) (Current awareness service for 80+ digital computer systems) pp. 11: 210.104-132; 11:510.102-108.
QA76.5 .A9175 5th Floor

BIBLIOGRAPHIES which contain material on electronic digital computers include:

Youden, W. W. Computer Literature Bibliography, 1946-1963 (1965) (6,100+ references)
ZQA76 .Y76 5th Floor

Studiecentrum voor Administratieve Automatisering. International Computer Bibliography (1968) (In English).
ZQA76 .S933 5th Floor

JOURNAL ARTICLES and other literature on electronic digital computers are indexed primarily in the guides listed. The quoted subject headings are those in use since 1965 unless other dates are given.

Applied Science and Technology Index (Covers popular engineering periodicals) See:
"Computers - Digital Computers"
ZT45 .A652 5th Floor

Science Abstracts, Series C, Computer and Control Abstracts (1969-, continues Series C, Control Abstracts) See:
"Digital Computers" (highly relevant)
"Digital Systems" (more general)
ZQA402.3 .C764 5th Floor

70-032

Fig. III-7 Printed master with Barker Engineering Library call numbers and floor locations.

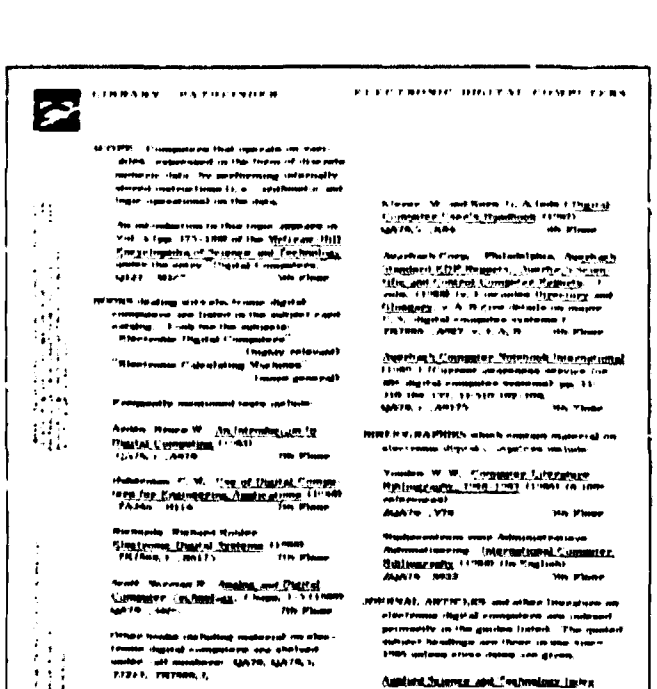


Fig. III-8 Final printed Library Pathfinder.

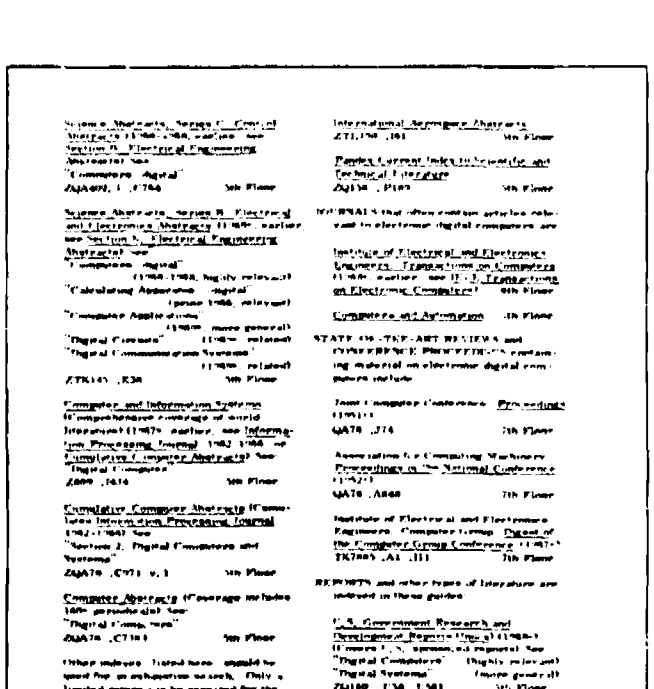


Fig. III-9 Final printed Library Pathfinder.

In the Barker Engineering Library, users are made aware of the existence and availability of individual Pathfinders through entries in the Subject Catalog and through notebooks at the public service desks.

In the Subject Catalog, Pathfinder cards are entered under their titles when printed copies are available for distribution. These cards are enclosed in distinctive, transparent plastic sleeves to increase their visibility. In the left-hand margin of the sleeve, the word Pathfinder is printed vertically in large white type against a blue background (Fig. III-10).

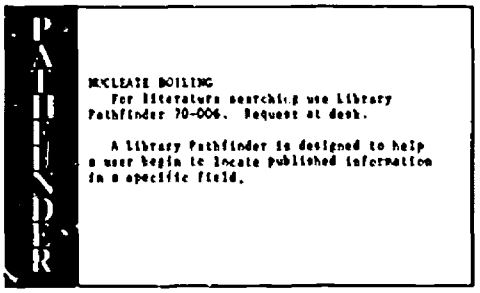


Fig. III-10 Library Pathfinder catalog card enclosed in color-coded plastic sleeve.

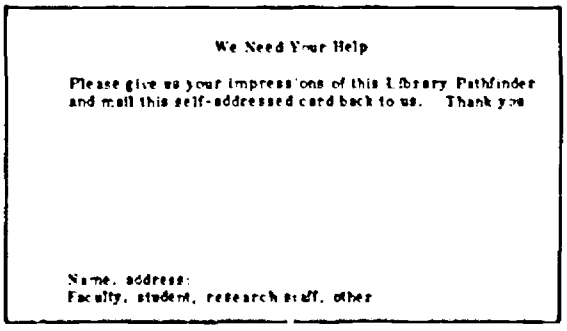


Fig. III-11 Self-addressed reply card for user comments.

Notebooks containing an up-to-date list of Pathfinders and samples of available titles are maintained at the circulation desk, the catalog information desk, and the reference desk. Library users can browse through the notebooks, select titles, and obtain copies for retention — free of charge — from the circulation desk file. Each Pathfinder furnished to the user has attached a self-addressed reply card on which the user is requested to record comments (Fig. III-11). During the period April through July, 258 mechanical engineering Pathfinders were requested.

Although the response of M.I.T. users to the Pathfinders has been limited to date, most of the comments have been favorable:

> "Excellent. . .a great convenience in starting research."
> "Great idea. . .saves research time."

One user raised the objection that the Pathfinders should be compiled by practicing subject specialists and not by librarians.

> "The Pathfinders. . .have to be built up by experts in the fields. . .
> You should not make Pathfinders."

With the beginning of the fall semester we expect to obtain more response to Pathfinders than we have during this last reporting period. The reasons are: (1) Pathfinders of interest to five engineering departments will be available and publicized; (2) a larger potential audience of serious users with a need for these tools will be using the library; viz., degree candidates doing initial literature searching; and (3) Pathfinders will be incorporated into the orientation program that the Barker Engineering Library is planning for the members of the M.I.T. engineering community. The Model-Library staff will actively pursue a program to obtain and evaluate user response.

From our earliest talks with librarians about Pathfinders, we have known that the idea would attract interest and attention. Engineering and science libraries in academic institutions and corporations would be able to use our first Pathfinders without substantial change. Other libraries could use our work as a model for Pathfinders in other disciplines. We have encouraged this idea to spread by providing Pathfinders without call numbers to requesting libraries and by showing and explaining Pathfinders to groups of librarians and information specialists during local and national professional meetings. Meetings during which Pathfinders were discussed in open forum included those of the American Library Association, American Society for Engineering Education, National Microfilm Association, and Special Libraries Association. Federal librarians have been contacted through the Committee on Scientific and Technical Information (COSATI) and through the newsletter of the Federal Library Committee. In the discussions with other librarians we have sought to expand the Pathfinder set by asking other librarians to plan a group of Pathfinders in areas of their competence and collect the information necessary for them. Since samples with worksheets and guidelines have just become available, it is too early to know whether this opportunity for cooperation will be seized or spurned. Several librarians have indicated a willingness to compile Pathfinders; a few have indicated the titles they would undertake, but none have reached us except those done by Simmons' students as described above.

Pathfinder revision plans are a concern for us and apparently for those who have seen Pathfinders outside of M.I.T. We have opened the possibility for change by

asking users for "additions, comments and corrections" in the printed text of each Pathfinder. As time passes we expect that our mail will ask for new items to be included and old ones deleted. Since the major effort for a single addition or deletion would be in the retyping and proofreading, we are taking action to minimize this portion of the work. On an experimental basis, Pathfinders are being stored on computer tape together with an editing and typesetting program that simplifies the routines of change and eliminates retyping. As the changes are required and made, we expect to keep careful records of cost, time delay, and copy quality. Should the computer edit program be too costly for use, the corrections will be made manually by retyping an entire side of a Pathfinder. It is our expectation that, with few exceptions, Pathfinders will not require revision more often than once in two to three years.

D. USER PREFERENCE STUDY

Mr. C. H. Stevens
Miss M. P. Canfield
Mr. J. J. Gardner

In March the Barker Engineering Library opened its Microform Service Area, a department which makes available microfiche copies and/or xerographic copies of some of the library's materials. The collection consists of the documents which comprise the Intrex data base; M.I.T. School of Engineering theses from 1945 to date; selected report literature; Thomas Micro-catalogs; and selected journal titles from 1960 to date. The collection has grown from a few hundred titles to 15,000 since March and will continue to grow very rapidly during the next year.

Service to users is provided through use of this array of commercially supplied equipment:

1. Xerox Microprinter capable of producing hard copy from negative microfiche. (It has an adapter for negative microfilm which is used only for reader supplied film.)

2. Bell & Howell Microfiche Printer and Bell & Howell Microfiche Processor which in combination provide on-demand microfiche-to-microfiche copy service in approximately 30 seconds.

3. Five microfiche readers: three Recordak Easamatic, and two IBM Document Viewer IIs.

4. Remington Rand Kard-Veyer, a power driven rotary microfiche file with capacity for 50,000 titles.

The service area stresses the concept of guaranteed text access. It provides each user with a duplicate microfiche copy of the item requested, a process that takes approximately 30 seconds. The master fiche is then returned to the file, and the user is given an option of retaining the fiche copy with no charge or of having hard copy produced at a charge of ten cents per page. It is this choice which the Model Library staff is studying.

We are interested in determining users' preferences between microfiche and hard copy and how variables affect those preferences. The primary variables which we are testing are: user status; type of material; length of material; and perhaps most important, the relative cost to the user of fiche and hard copy. In addition, we are soliciting comments on why users select one form of copy over the other.

Information is gathered on forms which serve as combination order-study forms and which are reproduced here as Figs. III-12 and III-13. The blanks are completed at the time the user makes his choice of text format.

A major element of the study will be to test preferences at varying costs. At the present time, duplicate fiche are given free of charge and hard copy for ten cents per page. We plan to shift these costs upward and downward for selected periods of time to determine their effect. We will also test users' time demands by offering fiche copies immediately and hard copy in 24 hours for selected periods.

The charts printed below (Figs. III-14 and III-15) give an indication of users' preferences. The figures are for 63 random Microform Service Area users and show that 53 or 84 percent selected fiche while 10 or 16 percent selected hard copy. No generalizations are possible from this small sample.

Users' comments on why they selected one form over the other are interesting and, in general, not surprising. Thirty-two of those who selected fiche responded to this question. Their reasons for choosing fiche are listed below:

1.	Low cost	13
2.	Curiosity	7
3.	Ease of storage	7
4.	No need for hard copy	2
5.	Low usage frequency	1
6.	Ease in scanning fiche	1
7.	Owns fiche reader	1

The nine who selected hard copy and responded to this question gave the following reasons:

1.	No fiche equipment readily available outside library	8
2.	Ability to write notes on hard copy	1

During the next reporting period we expect our sample to increase enough to provide meaningful statistics. The Microform Service Area will be in full operation from the start of the academic year, and the fiche collection will expand beyond 20,000 items. Also, with the availability of the Intrex console in the Barker Library, demands for documents from the Intrex data base will increase. We should then be able to determine the effect that our selected variables have on user preference of microfiche vs. hard copy.

MIT ENGINEERING LIBRARY - MICROFORM SERVICE AREA

MICROFICHE ORDER FORM (Use pink form for Hard Copy)

PLEASE PRINT

_____
Date

Name _____ Phone _____

Room Number or Address _____

Status: ☐ MIT Undergraduate ☐ MIT Faculty ☐ MIT Graduate Student
☐ MIT Staff/Employee ☐ Other (specify): _____

MATERIAL TO BE COPIED:

INTREX FICHE: Number _____ Pages (inclusive) _____

THESIS: Author _____ Department _____
Degree _____ Year _____

REPORT: Number _____

JOURNAL: Title _____
Date _____ Volume _____ Pages (inclusive) _____

Number of copies of each item: _____

****

THIS INFORMATION IS FOR A STUDY OF USER PREFERENCES OF TYPES OF COPY SERVICE

If there were a charge for this order, it would be paid for:
☐ personally - not reimbursed
☐ reimbursed business expense or charged to MIT account

Microfiche reading equipment is: ☐ available in your office ☐ available in your home
☐ available in your department ☐ not readily available outside library

Are you satisfied with the quality of microfiche you have used in the past?
☐ yes ☐ no ☐ never used any

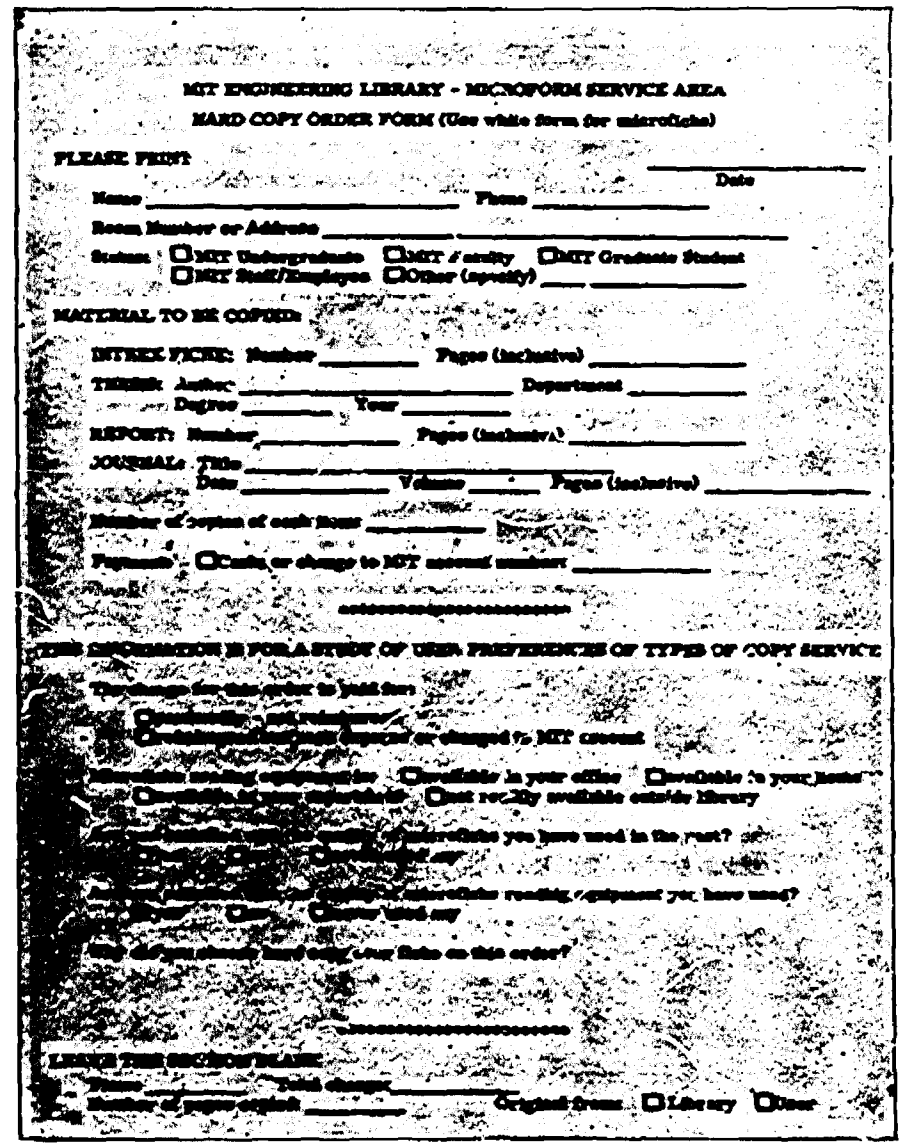
Are you satisfied with the quality of microfiche reading equipment you have used?
☐ yes ☐ no ☐ never used any

Why did you choose fiche over hard copy on this order?

****

LEAVE THIS SECTION BLANK
Time: _____
Number of fiche _____ Original from: ☐ Library ☐ User

Fig. III-12 User preference order-study form for copy service: microfiche to microfiche.


MIT ENGINEERING LIBRARY - MICROFORM SERVICE AREA

HARD COPY ORDER FORM (Use white form for microfiche)

PLEASE PRINT

Date

Name _____ Phone _____

Room Number or Address _____

Status: ☐ MIT Undergraduate ☐ MIT Faculty ☐ MIT Graduate Student
☐ MIT Staff/Employee ☐ Other (specify) _____

MATERIAL TO BE COPIED:

INTREX FICHE: Number _____ Pages (inclusive) _____

THESIS: Author _____ Department _____
Degree _____ Year _____

REPORT: Number _____ Pages (inclusive) _____

JOURNAL: Title _____
Date _____ Volume _____ Pages (inclusive) _____

Number of copies of each item _____

Payment: ☐ Cash, or charge to MIT account number _____

****

THIS INFORMATION IS FOR A STUDY OF USER PREFERENCES OF TYPES OF COPY SERVICE

The charge for this order is paid for:

☐ personally - not reimbursed
☐ reimbursed business expense or charged to MIT account

Microfiche reading equipment is: ☐ available in your office ☐ available in your home
☐ available in your department ☐ not readily available outside library

Are you satisfied with the quality of microfiche you have used in the past?
☐ yes ☐ no ☐ never used any

Are you satisfied with the quality of microfiche reading equipment you have used?
☐ yes ☐ no ☐ never used any

Why did you choose hard copy over fiche on this order?

****

LEAVE THIS SECTION BLANK
Time: _____ Total charge _____
Number of pages copied _____ Original from: ☐ Library ☐ User

Fig. III-13 User preference order-study form for copy service: microfiche to hard copy.

User Preference

Type of Material	No. of Orders for Microfiche	No. of Orders for Hard Copy	Total
Documents from Intrex Data Base	29	5	34
Technical Reports	4	1	5
M. I. T. Engineering Theses	9	0	9
Journals	2	0	2
Thomas Micro-catalogs	9	4	13
Total	53	10	63

Fig. III-14 Number of orders for microfiche vs. number of orders for hard copy according to type of material ordered.

Type of Material	% of Orders for Microfiche	% of Orders for Hard Copy
Documents from Intrex Data Base	85.3%	14.7%
Technical Reports	80%	20%
M. I. T. Engineering Theses	100%	-
Journals	100%	-
Thomas Micro-catalogs	69.2%	30.8%
Total	84.1%	15.9%

Fig. III-15 Percent of orders for microfiche vs. number of orders for hard copy according to type of material ordered.

IV. PROJECT INTREX STAFF

A. PROJECT OFFICE

Professor Carl F. J. Overhage, Director

Mr. Charles H. Stevens

B. ELECTRONIC SYSTEMS LABORATORY

Professor J. Francis Reintjes
Mr. Alan R. Benenfeld
Mr. Larry E. Bergmann
Mr. Joseph Bosco
Mrs. Susan Brown
Professor Lynwood S. Bryant
Mr. Peter Campoli
Mr. Daniel R. Cherry
Mr. Richard Chin
Mr. Noel A. Clark
Mr. Craig R. Davis
Mr. Robert Goldschmidt
Mr. Nobuyuki Goto
Mr. Daniel J. Griffin
Dr. Paul Holland
Mr. Charles E. Hurlburt
Miss Margaret A. Jackson
Mr. James E. Kehr
Mr. Donald R. Knudson
Mr. Peter Kugel
Miss Linda A. Langille
Miss Lucy T. Lee
Mr. Richard S. Marcus
Mr. Patrick L. Martin
Miss Virginia A. Miethe
Mr. Michael K. Molnar
Mr. Randy L. Prakken
Mrs. Elsie T. Raska
Professor James K. Roberge
Miss Leslie Rossin
Miss Rhonda L. Seegal
Mr. Terje A. Skotheim

C. BARKER ENGINEERING LIBRARY

Miss Rebecca L. Taggart, Head
Mrs. Marjorie Chryssostomidis
Miss Barbara C. Darling
Miss Carol L. Keator
Mr. James M. Kyed
Miss Helen Magedson
Miss Susan Nutter
Miss Mary Pensyl
Mr. David C. Van Hoy

D. MODEL LIBRARY PROGRAM

Mr. Jeffrey J. Gardner
Miss Marie P. Canfield
Miss Teresa E. R. Carsten
Miss Arlyne Jackson
Mrs. Elizabeth King
Miss Katherine C. Todd

V. CURRENT PUBLICATIONS

A. REPORTS

Goto, Nobuyuki, "A Translator Program for Displaying a Computer Stored Set of Special Characters." ESL-R-429, July, 1970.

B. BOOK CHAPTERS, JOURNAL ARTICLES AND CONFERENCE PAPERS

Knudson, D. R. and Vezza, A., "Remote Computer Display Terminals." Conference on Computer Handling of Graphical Information sponsored by SPSE, NMA, and SID, Newton, Mass., July 9-10, 1970, Proceedings, pp. 249-268.

Lovins, J. B., "Development of a Stemming Algorithm." Mechanical Translation Computational Linguistics, Vol. 11, Nos. 1 and 2, March and June, 1968, pp.22-31.

Overhage, C. F. J and Reintjes, J. F., "Computers in Libraries, Servant or Savant?" Presented at American Society for Information Science, New England Chapter Meeting, March 25, 1970.

Reintjes, J. F., "Hardware," as related to "Issues and Problems in Designing a National Program of Library Automation." Library Trends, Vol. 18, No. 4, April, 1970, pp.503-519.

Reintjes, J. F., "Recent Experiments with the Project Intrex Information Storage and Retrieval System." Gordon Conferences, New London, New Hampshire, July 16, 1970.

Roberge, J. K. and King, P. A., Jr., "An Economical Approach to High-Speed Character Generation and Display." 1970 Society for Information Display Symposium, New York, N. Y., May 26-28, 1970, Digest of Papers, pp.104-105.

Stevens, C. H., "Experiments with Microfiche in an Academic Library." Presented at the National Microfilm Conference, San Francisco, California, April 27, 1970.

Stevens, C. H., "Destination Shangri-La, First Stop Erewhon." Presented at American Society for Engineering Education National Conference, Columbus, Ohio, June 25, 1970.

Stevens, C. H., "Point-of-use-Instruction in Libraries." Presented at American Library Association National Convention, Detroit, Michigan, June 29, 1970.

Stevens, C. H., "New Whine in Olde Bottles." Presented at American Library Association National Convention, Detroit, Michigan, July 2, 1970.

C. THESES

Goto, Nobuyuki, "A Translator Program for Displaying a Computer Stored Set of Special Characters." M.S. thesis, Electrical Engineering Department, Massachusetts Institute of Technology, July, 1970. Also Electronic Systems Laboratory Report ESL-R-429.

VI. PAST PUBLICATIONS -- October, 1969 through 15 March, 1970

A. REPORTS

Haring, D. R., "The Augmented-Catalog Console for Project Intrex (Part II)." ESL-TM-410, December, 1969.

Lovins, J. B., "Error Evaluation for Stemming Algorithms as Clustering Algorithms." ESL-R-411, December, 1969.

Kusik, R. L., "A File Organization for the Intrex Information Retrieval System on the 360/67 CP/CMS Time-Sharing System." ESL-TM-415, January, 1970.

Project Intrex Staff, Semiannual Activity Report, 15 March 1970.

B. BOOK CHAPTERS, JOURNAL ARTICLES AND CONFERENCE PAPERS

Knudson, D. R., "Image Storage and Transmission for Project Intrex", Conference on Image Storage and Transmission for Libraries, National Bureau of Standards, Gaithersburg, Maryland, December 1-2, 1969.

Overhage, Carl F. J., "Information Networks", Chapter 11 in Annual Review of Information Science and Technology, Vol. 4, Carlos A. Cuadra, Editor. Encyclopedia Britannica, Inc. Chicago, 1969.

C. THESES

Kusik, R. L., "A File Organization for the Intrex Information Retrieval System on the 360/67 CP/CMS Time-Sharing System." M.S. Thesis, Electrical Engineering Department, M.I.T., November, 1969. Also Electronic Systems Laboratory Report ESL-TM-415.

VII. PAST PUBLICATIONS — 1966 through September, 1969

A. REPORTS

Jagodnik, A. J., Jr., "Performance Evaluation of Image Storage and Transmission Systems." ESL-R-391, June, 1969.

King, P. A., Jr., "A Novel Solid State Character Generator." ESL-TM-386, June, 1969.

Lufkin, R. C., "Determination and Analysis of Some Parameters Affecting the Subject Indexing Process." ESL-R-364, September, 1968.

Kampe, W. R., "Pre-Indexing by Machine." ESL-R-355, July, 1968.

Lovins, J. B., "Development of a Stemming Algorithm." ESL-TM-353, June, 1968.

Benenfeld, A. R., "Generation and Encoding of the Project Intrex Augmented Catalog Data Base." ESL-R-360, August, 1968. This report is based upon a paper presented at the 6th Annual Clinic on Library Applications of Data Processing, University of Illinois, Urbana, Illinois, May 7, 1968.

Haring, D. R., and Roberge, J. K., "The Augmented Catalog Console for Project Intrex." ESL-TM-323, October, 1967.

Gronemann, U. F., Knudson, D. R., and Teicher, S. N., "Remote Text Access for Project Intrex." ESL-TM-312, July, 1967. This report is based upon a conference paper presented at the National Microfilm Association Convention, Miami Beach, Florida, April 26-28, 1967.

Benenfeld, A. R., Gurley, E. J., and Rust, J. E., "Cataloging Manual." ESL-TM-303, February, 1967.

Project Intrex Staff, Semiannual Activity Report, 15 September, 1969.

Project Intrex Staff, Semiannual Activity Report, 15 March, 1969.

Project Intrex Staff, Semiannual Activity Report, 15 September, 1968.

Project Intrex Staff, Semiannual Activity Report, 15 March, 1968.

Project Intrex Staff, Semiannual Activity Report, 15 September, 1967.

Project Intrex Staff, Semiannual Activity Report, 15 March, 1967.

Project Intrex Staff, Semiannual Activity Report, 15 September, 1966.

Project Intrex Staff, Semiannual Activity Report, 15 March, 1966.

B. JOURNAL ARTICLES AND CONFERENCE PAPERS

Stevens, C. H., "Will the Library Sink or Swim — A Buoyant Response." Presented at the American Chemical Society National Meeting, New York, New York, September 8, 1969.

Reintjes, J. F., "The Use of Multi-access Computers for the Management and Control of Professional Literature," Fourth Congress of the International Federation of Automatic Control, Warsaw, Poland, June 16-21, 1969.

Knudson, D. R., and Teicher, S. N., "Remote Text Access in a Computerized Library Information Retrieval System," 1969 Spring Joint Computer Conference, Boston, Mass., May 14-16, 1969, AFIPS Conference Proceedings, Vol.34, pp. 475-481.

Marcus, R. S., Kugel, P., and Kusik, R. L., "An Experimental Computer-Stored, Augmented Catalog of Professional Literature," 1969 Spring Joint Computer Conference, Boston, Mass., May 14-16, 1969, AFIPS Conference Proceedings, Vol. 34, pp.461-472.

Reintjes, J. F., "System Characteristics of Intrex," 1969 Spring Joint Computer Conference, Boston, Mass., May 14-16, 1969, AFIPS Conference Proceedings, Vol. 34, pp. 457-459.

Stevens, C. H., "Book Publishing, A Look to the Future." Presented at the Association of American University Presses Conference on Information Science and Publishing, University of Chicago, April 22, 1969.

Benenfeld, A. R., "Generation and Encoding of the Project Intrex Augmented Catalog Data Base." Proceedings of the May 6, 1968 Clinic on Library Applications of Data Processing, Urbana, Illinois, University of Illinois, Graduate School of Library Science, 1969, pp. 155-198. (Also issued as Electronic Systems Laboratory Report ESL-R-360, August, 1968.)

Haring, D. R., "Computer-Driven Display Facilities for an Experimental Computer-Based Library." 1968 Fall Joint Computer Conference, San Francisco, California, December 9-11, 1968. AFIPS Conference Proceedings, Vol. 33, pp. 255-265.

Haring, D. R., and Roberge, J. K., "Display Techniques for an Experimental Computer-Based Library." UAIDE Conference, San Francisco, California, October 28-31, 1968.

Knudson, D. R., Teicher, S. N., Reintjes, J. F., Gronemann, U. F., "Experimental Evaluation of the Resolution Capabilities of Image-Transmission Systems." Information Display - September/October, 1968.

Stevens, C. H., "Then Beggars Would Ride." Engineering Joint Management Council Conference, Philadelphia, Pennsylvania, September 30, 1968.

Haring, D. R., "A Display Console for an Experimental Computer-Based Augmented Library Catalog." 1968 ACM National Conference and Exposition, Las Vegas, Nevada, August 27-29, 1968. (To Appear in Conference Proceedings.)

Reintjes, J. F., "Project Intrex, (A Research Project under Development)." Presented at the Conference on Computers and the University, Technical University of Berlin, July 22- August 2, 1968.

Benenfeld, A. R., "Data Encoding for the Project Intrex Augmented-Catalog Experiments." Presented at the American Documentation Institute Convention, User Discussion Group on Emerging Machine-Readable Tape Services, New York, October 25, 1967.

Haring, D. R., "A Terminal for an On-Line Interactive Retrieval System." Presented at the IEEE Workshop on Advanced Computer Peripherals, Lake Arrowhead, California, August 25-27, 1967.

Overhage, Carl F. J., "Science Libraries: Prospects and Problems." Science, 155, 17 February, 1967, pp. 802-806.

Overhage, Carl F. J., "Plans for Project Intrex." Science, 1952, 20 May, 1966, pp. 1032-1937.

C. THESES

King, P. A., Jr., "A Novel Solid State Character Generator," M.S. Thesis, Electrical Engineering Department, Massachusetts Institute of Technology, June, 1969. Also Electronic Systems Laboratory Report ESL-TM-386, June, 1969.

Jagodnik, A. J., "Performance Evaluation of Image Storage and Transmission Systems." M.S. Thesis, Massachusetts Institute of Technology, February, 1969.

McKenzie, P. F., "A Flying-Spot-Scanner Character Generator." M.S. Thesis, Massachusetts Institute of Technology, February, 1969.

Kampe, W. R., "Pre-Indexing by Machine." M.S. Thesis, Electrical Engineering Department, Massachusetts Institute of Technology, June, 1968. Also Electronic Systems Laboratory Report ESL-R-355, published July, 1968.

Lufkin, R. C., "Determination and Analysis of Some Parameters Affecting the Subject Indexing Process." B.S. Thesis, Massachusetts Institute of Technology, June, 1968. Also Electronic Systems Laboratory Report ESL-R-3[illegible], published September, 1968.

Domercq, R. J., "A Machine-Aided Thesaurus Generation System." M.S. Thesis, Massachusetts Institute of Technology, September, 1967.

Stuntz, S. C., "Inputting Process to an Augmented Library Catalog -- An Efficiency Study." B.S. Thesis, Massachusetts Institute of Technology, June, 1967.

D. MISCELLANEOUS PRESENTATIONS

Alan R. Benenfeld

"The Project Intrex Augmented Catalog." Syracuse University, Graduate School of Library Science, May 22, 1969.

Gary L. Benton

Panel on New Hardware Developments, ADI Annual Meeting, Santa Monica, California, October 5, 1966.

"Project Intrex-Information Transfer Experiments at M.I.T." Raytheon Institute of Bedford, Bedford, Mass., February 15, 1966.

"Project Intrex: Plans and Progress." Massachusetts Association of Medical Record Librarians, Boston, Mass., January 26, 1966.

Carl F. J. Overhage

"Information Transfer Experiments at M.I.T." International Federation for Information Processing, IFIP 68 Congress, Edinburgh, Scotland, August 6, 1968.

"Information Transfer in the Library of the Future." Panel of the International Communications Conference on "The Evolving Relationship Between the Publisher and the Electronic Firm--Software and Hardware," IEEE, Philadelphia, Penn., June 12, 1968.

"Plans for Project Intrex." Symposium on Mechanized Information Services in the Library, UCLA, March 29-31, 1968.

"Project Intrex-Plans and Progress." Harvard University Librarian's Seminar, Cambridge, Mass., May 17, 1967.

"Project Intrex and the Future of the Book." American Book Publishers' Association, Arden House, Harriman, New York, March 16, 1967.

M.I.T. Center for Advanced Engineering Study, Cambridge, Mass., November 29, 1966.

American Cyanamid Company, Librarians and Information Specialists Meeting, Princeton, New Jersey, October 20, 1966.

M.I.T. Sloan School of Management, Cambridge, Mass., October 13, 1966.

"Plans for Project Intrex at M.I.T."

Panel on National Information Issues and Trends, ADI Annual Meeting, Santa Monica, California, October 6, 1966.

Panel on Information Systems Applications, ADI Annual Meeting, Santa Monica, California, October 5, 1966.

Association of Research Libraries, New York, New York, July 9, 1966.

AFIPS Spring Joint Computer Conference, Boston, Mass., April 28, 1966.

Rochester Computer Systems Association, Rochester, New York, April 26, 1966.

Optical Society of America and American Chemical Society, (Rochester sections) Rochester, New York, March 1, 1966.

M.I.T. Lincoln Laboratory, Lexington, Mass., February 8, 1966.

American Physical Society, New York, New York, January 29, 1966.

American Library Association, Chicago, Illinois, January 26, 1966.

American Association for the Advancement of Science, Berkeley, California, December 29, 1965.

American Documentation Institute, Cambridge, Mass., October 28, 1965.

"Plans for Information Transfer Experiments at M.I.T."

Charles H. Stevens

"Project Intrex - A Progress Report." MITRE Corporation Information Systems Seminar. Bedford, Mass., November 5, 1968.

"Console Development at Project Intrex." Special Libraries Association, Los Angeles, California, June 5, 1968.

"Books and Other Paraphernalia - A Forward Look." National Endowment for the Humanities, Boston, Mass., May 10, 1968.

"Engineering Libraries and Information Systems." Center for Advanced Engineering Study, M.I.T., Cambridge, Mass., April 16, 1968.

"Project Intrex - Plans and Progress." New York Chapter, Special Libraries Association, New York, New York, February 6, 1968.

"Project Intrex." Gordon Research Conference on Scientific Information, New London, New Hampshire, July 19, 1967.

"New Developments in Libraries." State University of New York (Albany), April 4, 1967.

"Libraries in the Age of Electronics." Library Symposium, Boston University, Boston, Mass., November 19, 1966.

"Project Intrex: Plans and Progress." Information Sciences Colloquium, Lehigh University, Bethlehem, Pa., October 27, 1966.

"Project Intrex and the Future of the Academic Library." Science and Technology Information Group, Washington, D.C., October 25, 1966.

"Project Intrex and Information Transfer." Sloan Fellows Seminar, M.I.T., Cambridge, Mass., October 13, 1966.

"Project Intrex: Plans and Progress." Special Libraries Association - Texas Chapter, Austin, Texas, September 24, 1966.

"Special Libraries in the Future." John Cotton Dana Lecture, University of Texas, Austin, Texas, September 23, 1966.

"Project Intrex and the Special Library." Special Libraries Association, Minneapolis, Minn., May 30, 1966.

"Future Trends in Academic Libraries." Simmons College Library School, Boston, Mass., April 22, 1966.

"Project Intrex at M.I.T." Simmons College Library School Alumni Association, Boston, Mass., April 12, 1966.

"The Future of the Book." Boston Bookbuilders Workshop, Boston, Mass., March 16, 1966.

"Project Intrex." Special Libraries Association (Boston Chapter), Boston, Mass., January 17, 1966.