DOCUMENT RESUME

ED 076 211 LI 004 316

AUTHOR	Schumacher, H. H.; And Others
TITLE	The Use of Selected Portions of Technical Documents as Sources of Index Terms and Effect on Input Costs and Retrieval Effectiveness.
INSTITUTION	Air Force Materials Lab., Wright-Patterson AFB, Ohio.; Dayton Univ., Ohio. Research Inst.
REPORT NO	AFML-TR-73-53
PUB DATE	May 73
NOTE	87p.; (14 References)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Automation; *Cost Effectiveness; Data Bases; Information Centers; Information Processing; Information Retrieval; *Input Output Analysis; *Relevance (Information Retrieval); *Subject Index Terms; Technical Reports
IDENTIFIERS	*Aerospace Materials Information Center; AMIC

ABSTRACT

A correlation between recall (the retrieval of all available relevant documents) and quantity of text which served as a source of index terms on input can reasonably be expected. Specifically, recall should decrease as the quantity of text serving as a source of index terms is restricted. On the other hand, the time for indexing and therefore the input cost should be less, thus establishing a tradeoff between input cost and retrieval effectiveness. It was desired to quantify the effect of restricting the source text on both retrieval effectiveness and input cost. An experiment was designed in which the full technical document text was divided into five categories: (1) title; (2) abstract; (3) table of contents and lists of figures and tables; (4) author-assigned keywords; and (5) the body. An experimental data base of technical documents was created, for which the index term source category and the time required for indexing by category was recorded. Sets of Selective Dissemination of Information (SDI) and retrospective searches were run against the data base, and retrievals were analyzed. Based on the results, it was decided that the body of the document could be excluded as a source of index terms. This decision was translated into a reduction of unit cost from $10 to $8.25. (Author/SJ)

AFML-TR-73-53

THE USE OF SELECTED PORTIONS OF TECHNICAL DOCUMENTS AS SOURCES OF INDEX TERMS AND EFFECT ON INPUT COSTS AND RETRIEVAL EFFECTIVENESS

H. H. SCHUMACHER
J. F. MARCH
F. L. SCHEFFLER

UNIVERSITY OF DAYTON RESEARCH INSTITUTE
DAYTON, OHIO

TECHNICAL REPORT AFML-TR-73-53

AIR FORCE MATERIALS LABORATORY
AIR FORCE SYSTEMS COMMAND
WRIGHT PATTERSON AIR FORCE BASE, OHIO

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated furnished, or in any way supplied the said drawings specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

AIR FORCE/56780/31 May 1973 — 400

THE USE OF SELECTED PORTIONS OF TECHNICAL DOCUMENTS AS SOURCES OF INDEX TERMS AND EFFECT ON INPUT COSTS AND RETRIEVAL EFFECTIVENESS

H. H. SCHUMACHER
J. F. MARCH
F. L. SCHEFFLER

FOREWORD

This report was prepared by the University of Dayton Research Institute, Dayton, Ohio, under Air Force Contract F33615-71-C-1069. The work described herein was accomplished under Project 7381 "Materials Application" and Task 738103 "Materials Information Development, Collection, and Processing." The effort was administered under the direction of the Operations Office of the Air Force Materials Laboratory with Mr. Edward Dugger (AFML/DO) as Project Monitor.

This is a final summary report and covers the work accomplished from 1 December 1971 through 30 November 1972.

The authors acknowledge the efforts and contributions of Mr. Eugene R. Egan, Mr. Mark S. Klug, Mr. John T. Logan, Mr. Donald L. Wannemacher and Miss Virginia C. Weber in providing indexing input and assistance in compiling statistical data.

This technical report has been reviewed and is approved.

EDWARD DUGGER

Operations Office

Air Force Materials Laboratory

ABSTRACT

A correlation between recall (the retrieval of all available relevant documents) and quantity of text which served as a source of index terms on input can reasonably be expected. Specifically, recall should decrease as the quantity of text serving as a source of index terms is restricted. On the other hand, the time for indexing and therefore the input cost should be less, thus establishing a tradeoff between input cost and retrieval effectiveness. It was desired to quantify the effect of restricting the source text on both retrieval effectiveness and input cost. An experiment was designed in which the full technical document text was divided into five categories: 1, title; 2, abstract; 3, table of contents and lists of figures and tables; 4, author-assigned keywords; and 5, the body. An experimental data base of technical documents was created, for which the index term source category and the time required for indexing by category was recorded. Sets of Selective Disemination of Information (SDI) and retrospective searches were run against the data base, and retrievals were analyzed by category in terms of retrieval response, S; relevant document response, R; categorical relevance, $\overline{R}$; indexing time, T; and retrieval efficiency, E and $\overline{E}$. It was found for the subset of documents retrieved for all searches, that 81% of the available relevant documents were retrieved from categories 1-4, whereas the indexing time required for these four categories was only 53% of the total indexing time, as compared to the time for all five categories. For the entire set of documents input into the experimental data base, the portion of indexing time for the first four categories was 60%. Based on these results, it was decided that the body of the document could be excluded as a source of index terms. This decision was translated into a reduction of unit cost from $10 to $8.25.

TABLE OF CONTENTS

TABLE OF CONTENTS continued

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES CONTINUED

SECTION I

INTRODUCTION

The work described in this report covers two aspects: the operations and nature of the Aerospace Materials Information Center (AMIC) and an experimental program to test the effect of the source(s) of index terms upon the retrieval effectiveness and input cost. A brief description of AMIC is given followed by considerations affecting the input to and retrieval from the system. The specific purpose of the experimental program was to provide quantitative data upon which a management decision could be made regarding the tradeoff between input cost and retrieval effectiveness.

1.1 DESCRIPTION OF THE AMIC SYSTEM

The Information Systems Section of the University of Dayton Research Institute (UDRI) has established and presently maintains and operates a document retrieval system in support of the Aerospace Materials Information Center (AMIC). The document retrieval system operated by the University of Dayton contains approximately 65,000 documents concerning materials research and development with new acquisitions being processed continually. The establishment, modification, and operation of the document retrieval system are described in references 1 through 9. The present report describes the work performed from 1 December 1971 through 30 November 1972.

The AMIC document retrieval system has been in operation with retrospective search capabilities since 1963. The purpose of the system is to provide scientific and technical information to qualified requesters in a timely and efficient manner. The information is supplied in the form of abstracts of documents pertinent to the search request; these abstract formats also contain complete bibliographic information, including AMIC access number, DDC AD number or NASA N number, generating agency, report number, title, author, contract number (if applicable), and date of issue of the document. The documents themselves are available from the Materials Documentation Center (MDC) maintained at the Air Force Materials Laboratory (AFML). Hardcopy documents are available on loan to AFML requesters. Microfiche documents are reproduced and the duplicate microfiche are provided to the requester if permanent retention is desired.

The AMIC document retrieval system is primarily concerned with the materials aspect of technical documents. Because of the concentration on materials, retrieval capabilities from a materials standpoint are very comprehensive. Retrieval can be quite specific. For example, a request for all information on the alloy Aluminum 2024 T6 can be readily satisfied; on the other hand, retrieval can be general in nature, e.g., high temperature fatigue of all metals and alloys. Similarly, a requester could ask for information on boron reinforced Epon epoxy composites or for aircraft structural applications of any composite material.

Retrospective searches encompassing the entire range of materials information are run regularly by UDRI in response to requests from the AFML. To ensure that the requester receives abstracts which are relevant to the request, all abstracts and index cards retrieved are screened for content by a UDRI information specialist to assess their relative pertinence to the originally-stated request.

AMIC also offers SDI services. SDI refers to Selective Dissemination of Information, which is the practice of providing timely, pertinent references to documents in particular areas of interest to a number of users, each user receiving only material of potential interest to him. The concept is also referred to as current awareness. The SDI program is based on the periodic input of document index data to the AMIC system.

1.2 INDEXING PHILOSOPHY

There are a number of viewpoints regarding the indexing of technical documents. For purposes of this discussion we will restrict our consideration primarily to coordinate indexing consisting of manual assignment of keywords or descriptors to serve as the set of retrieval access points for that document. We will not involve indexing by classification schemes. The topic we wish to address is the "depth" of indexing.

It is necessary at this point to define and differentiate "depth" of indexing from "specificity" of indexing, since these terms are often used interchangeably. By depth of indexing we mean the extent of the document which serves as a source of index terms; this can vary from only the title to the entire full text of the document. By specificity we mean the degree to which an index item is described in relation to the hierarchical possibilities for its description. To illustrate, let us consider the title "Fatigue Properties of Aluminum 2024-T6." Indexing from the title would be specific (including detailed specific nomenclature) but "shallow" (only the title was used as the indexing source). To be sure, there is a correlation between the depth of indexing and the specificity of indexing, but the two concepts are different.

The philosophical question regarding indexing is: How deep is deep enough? The Keyword in Context (KWIC) index which was developed by H. P. Luhn[10] depends on the display of significant words appearing in the title. Often only the title and abstract are used for manual indexing. A number of automatic indexing systems depend on the extraction of significant words from only the title and abstract; from this practice one could infer that these sources are considered adequate for indexing. Many people contend, however, that the full text of the document must serve as the source of indexing.

We believe that the answer to the question must be based on the scope of the information system and on the nature of the information needs of the user. All too often, in our view, the indexing function has governed the depth of indexing with too little attention being given to the actual user needs. Surely, the user needs should have some influence on how deeply and how specifically material is indexed in the first place. In the case of AMIC, which is specialized in the area of materials, there is a real need for highly specific retrieval regarding materials as evidenced by the types of requests presented by the user. Therefore this specificity in retrievability should be provided on input. In recognition of the need for specificity and realizing the correlation of depth and specificity of indexing, UDRI has used the full text of the document as a source of index terms.

However, with some years of indexing experience, UDRI has developed intuitively the notion that indexing time probably could be substantially reduced without serious loss of retrievability by restricting the source of index terms in documents to certain well-defined portions thereof, but not including the body. This idea had never been subjected to rigorous testing until the experimental program reported herein.

Miller[11] in studying the MEDLARS system concluded that a large percentage of index terms can be found in the document title. He cited the following evidence:

(a) 228 legal documents showed that 64% of the titles contained <u>all</u> the index terms

(b) 5 titles from the Physical Review contained 63% of the total number of index terms

(c) 4 titles from Chemical Abstracts Subject Index contained 57% of the total number of index terms

Depth of indexing in terms of titles and abstracts has been studied by Tell[12], in Sweden. He reported that titles and abstracts are good sources of index material based on relevance judgments of the users. Tell further suggests (perhaps wistfully) that the authors' knowledge that indexing will be accomplished from the title and abstract will cause them to write more informative titles and abstracts.

Lancaster[13] states that with regard to cost effectiveness, an indexing system can be improved either by: (a) altering the indexing/retrieval language in such a way that system costs are reduced while the present level of search effectiveness is maintained; or (b) making system changes that improve search effectiveness with no measurable increase in overall system cost. As stated earlier, a primary concern of UDRI was to improve the cost effectiveness of AMIC.

1.3 UDRI APPROACH TO THE PROBLEM

The approach taken by UDRI in its experimental study represents a variation of alternative (b) as proposed by Lancaster. An experiment was designed to test the effect of the depth of indexing on retrieval effectiveness of the AMIC system. According to Tell[14] studies have shown that exhaustive indexing will diminish retrieval effectiveness. There is a point when an increase of indexing depth will decrease retrieval quality. He goes on to state that indexers are more consistent when indexing from titles and/or abstracts than from full text. One of our studies[5] also suggested this phenomenon.

The hypothesis assumed by UDRI for the AMIC experiment is as follows:

> Indexing time, and therefore cost, will be substantially reduced without serious loss in the number of relevant documents retrieved if the source of index terms is limited to exclude the body of the document.

A model was designed to test this hypothesis. The data base for the model consisted of 984 documents indexed by three experienced indexers. Each indexer indexed about 330 documents. The indexers were instructed to select and designate terms derived from each of the five different sources within the documents. In designating the indexing source, only the additional terms obtained from each subsequent section were indicated; thus the categories were considered cumulative and not completely independent. These sources and combinations thereof represented varying depths of indexing. The five sources of index terms were categorized as follows:

Category 1	Title
Category 2	Abstract
Category 3	Table of Contents, and lists of Tables and Figures
Category 4	Author Keywords
Category 5	Body of document*

Index terms were coded as to the source (Category 1, Category 2, etc.) by using these code numbers on the index card. The code number for each term thus provided the key to determine the source of the term. The indexers also maintained records of the indexing time required for each category. Two sets of actual previously run searches (97 SDI searches, 15 retrospective searches) were run against the model data base. Each search was then

---

* Note that the body of the document is the narrative descriptive portion of the document exclusive of the title, abstract, table of contents and author keywords. The full text is represented by 1-5.

analyzed according to the number of documents retrieved as a result of the five indexing categories. From the data obtained, retrieval effectiveness was determined as a function of the depth of indexing and input cost was correlated with the depth of indexing.

SECTION II

EXPERIMENTAL PROGRAM

2.1 RETRIEVAL RESPONSE AS A FUNCTION OF THE SOURCE OF INDEX TERMS

The first part of this experiment addressed itself to the problem of determining the amount of raw retrieval or recall as a function of indexing depth. No effort was made in the analysis of these results to determine relevance. The documents retrieved on a given search contained not only the usual access number but also the category number, thus revealing the indexing source within the document by which the retrieval was made. By referring to Appendix A, it can be seen that some documents on a given search were retrieved in all five categories. For example, on SDI Search #99019, Document #69882 was retrieved by each of the five categories. In this particular example, any one of the five sources of index terms or any combination thereof would have been sufficient for retrieval. Other document were retrieved by only a single category. For example, on SDI Search #99071, document #200355 was retrieved only by Category 5. In later analysis, this document was judged to be relevant. In this example, evidently it was necessary to index from the body of the document to effect retrieval. The two examples cited above represent the extreme cases; most documents were retrieved with various combinations of Categories 1 through 5. A tally was made (see Appendix A) for each search showing the total number of documents retrieved and the percentage of the total number of documents first retrieved by each category. For example, if a document was retrieved on Category 2, then indexing from the title and abstract was sufficient to effect retrieval.

For the purpose of this experiment, all searches were assumed to retrieve all available documents, that is, retrieval obtained from full text indexing represented 100% recall. We here introduce the term "retrieval response." Retrieval response represents the ratio of the documents retrieved by a given category to the entire set of documents retrieved from full text. The retrieval response we will symbolize by S. Figure 1 shows the retrieval response for the various categories. The retrieval responses are presented for each category such that any given document is counted only for that category by which retrieval first occurred. The corresponding averaged data are presented in Table I. Raw data are provided in Table A-1, Appendix A.

TABLE I

RETRIEVAL RESPONSE AS A FUNCTION OF THE SOURCE OF INDEX TERMS (CATEGORY)

Category	Retrieval Responses S(%)	Cumulative Retrieval Response $\sum S(\%)$
1 - Title	19	19
2 - Abstract	43	62
3 - Table of Contents and List of Figures and Tables	9	71
4 - Author keywords	2	73
5 - Body	26	100

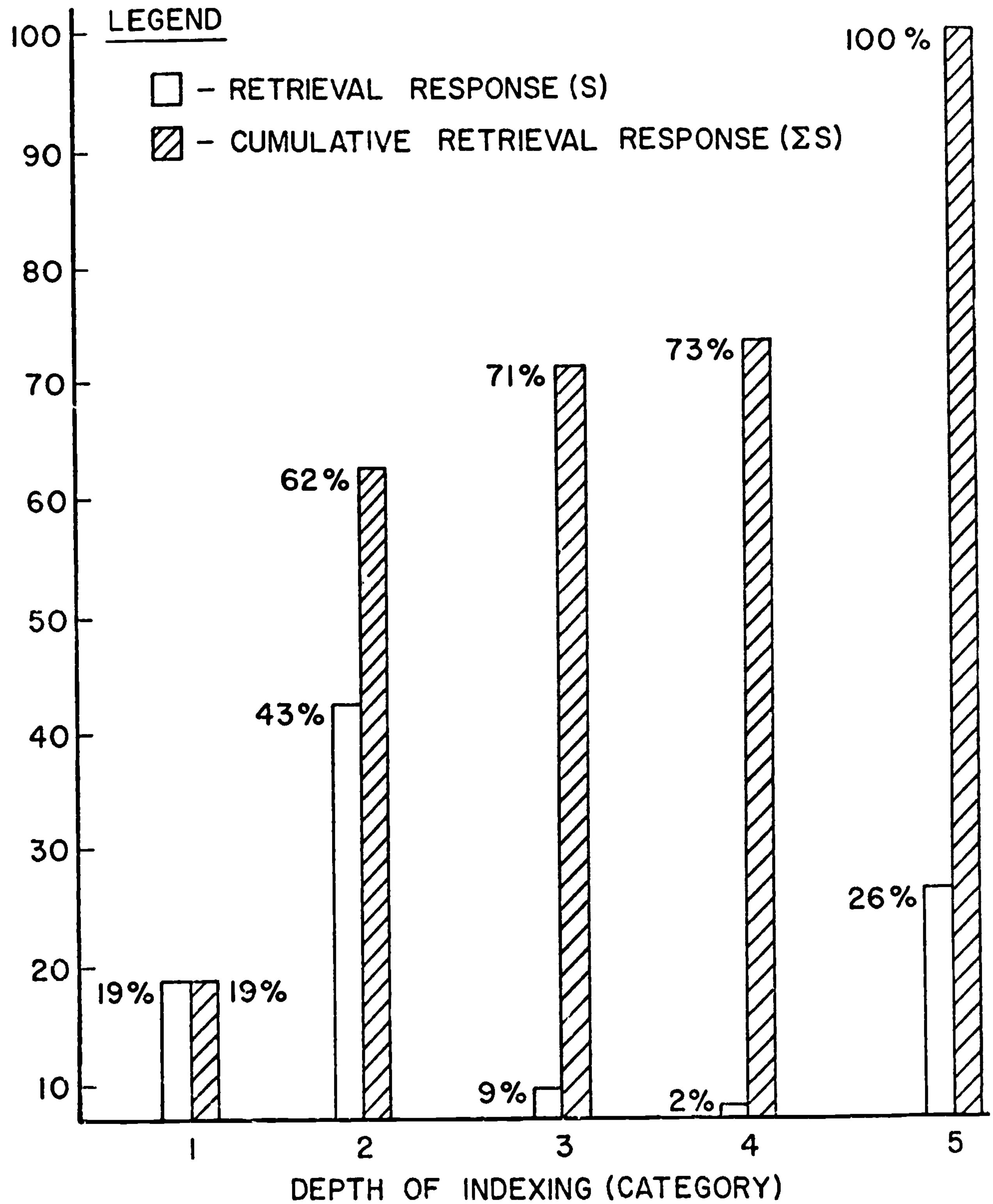

Figure 1. Retrieval Response as a Function of the Source of Index Terms.

An analysis of the retrieval response data reveals that the title alone does not yield a substantial portion of AMIC documents for the set of searches run against the data base. Addition of the abstract as a source of index terms provides an additional 43% of the documents retrieved from full text indexing; thus abstract and title together provide 62% of retrieval response. Interestingly, author keywords represent only a 2% improvement over title, abstract, and table of contents and list of figures and tables. If the body of the document is excluded, a retrieval response of 73% is obtained.

2.2 RELEVANT DOCUMENTS RESPONSE AND CATEGORICAL RELEVANCE AS A FUNCTION OF THE SOURCE OF INDEX TERMS

The second part of the experiment was concerned with the relevance of the retrievals. In particular, we were interested in the distribution of relevant documents among the five categories. From the retrieval response we learned that if the body is excluded as a source of index terms, 27% of the documents which would have been retrieved from full text indexing were, in fact, not retrieved. But of this 27%, how many were actually relevant? Obviously, if none of the 27% were relevant, we would be actually improving the retrieval effectiveness by not retrieving nonrelevant documents, while at the same time reducing input cost since the time for indexing the body of the document would no longer be required. Such an idealized situation would represent an increase in relevance with no corresponding reduction in recall. Actually we expected some loss in recall as the price for increased relevance.

In order to provide some answers, an analysis was performed to determine the relevance of the documents being retrieved. The searches run against the model data base were screened by judging the document to be either relevant or nonrelevant. Two experienced UDRI AMIC information specialists performed this task. The searches were screened independently, each information specialist screening a different set of searches. A few of the searches were screened by both specialists; no significant differences were observed. After screening, the relevant documents for each search were summed in order to determine both the distribution of relevant documents by category and the relevance factor for each category.

From the relevance data, we can consider relevance in various ways. One possibility is to consider the relevance as the ratio of number of relevant documents in a given category to the total number of documents in that category. This concept is defined as the "categorical relevance," symbolized by $\overline{R}$. Another way is to consider the ratio of the number of relevant documents retrieved in a given category to the total number of relevant documents available. We define this concept as the "relevant document response," symbolized by R. The relevant document response indicates the distribution of relevant documents by category. $\sum \overline{R}$ represents the usual concept of relevance, i.e., the number of relevant documents retrieved divided by the total number of documents retrieved. The relevant document response ($\bar{R}$) by category and the categorical relevance ($\overline{R}$) are shown in Figures 2 and 3. The corresponding averaged data are given in Table II. Raw data are provided in Table A-2 in Appendix A.

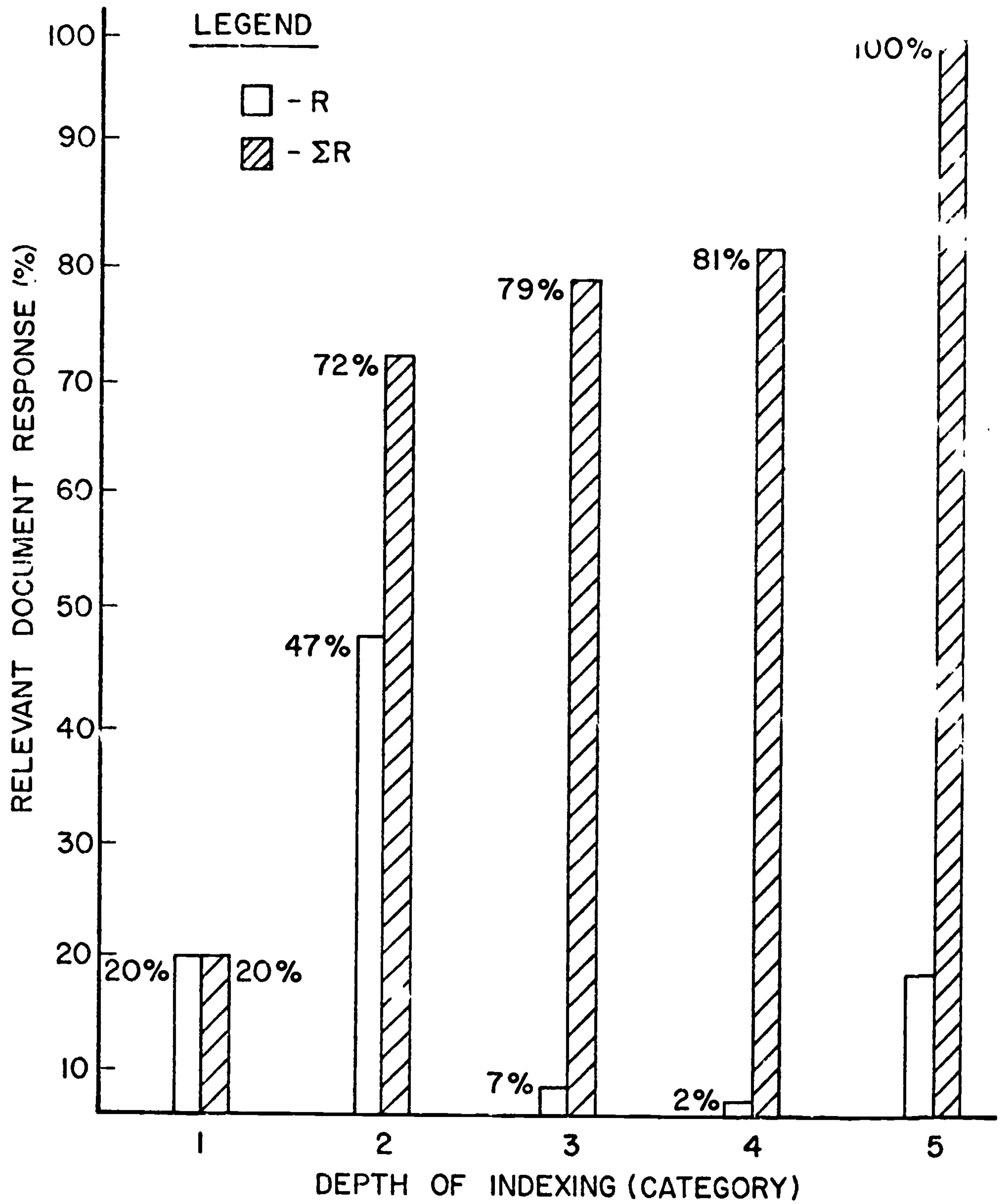

Figure 2. Relevant Document Response as a Function of the Source of Indexing Term

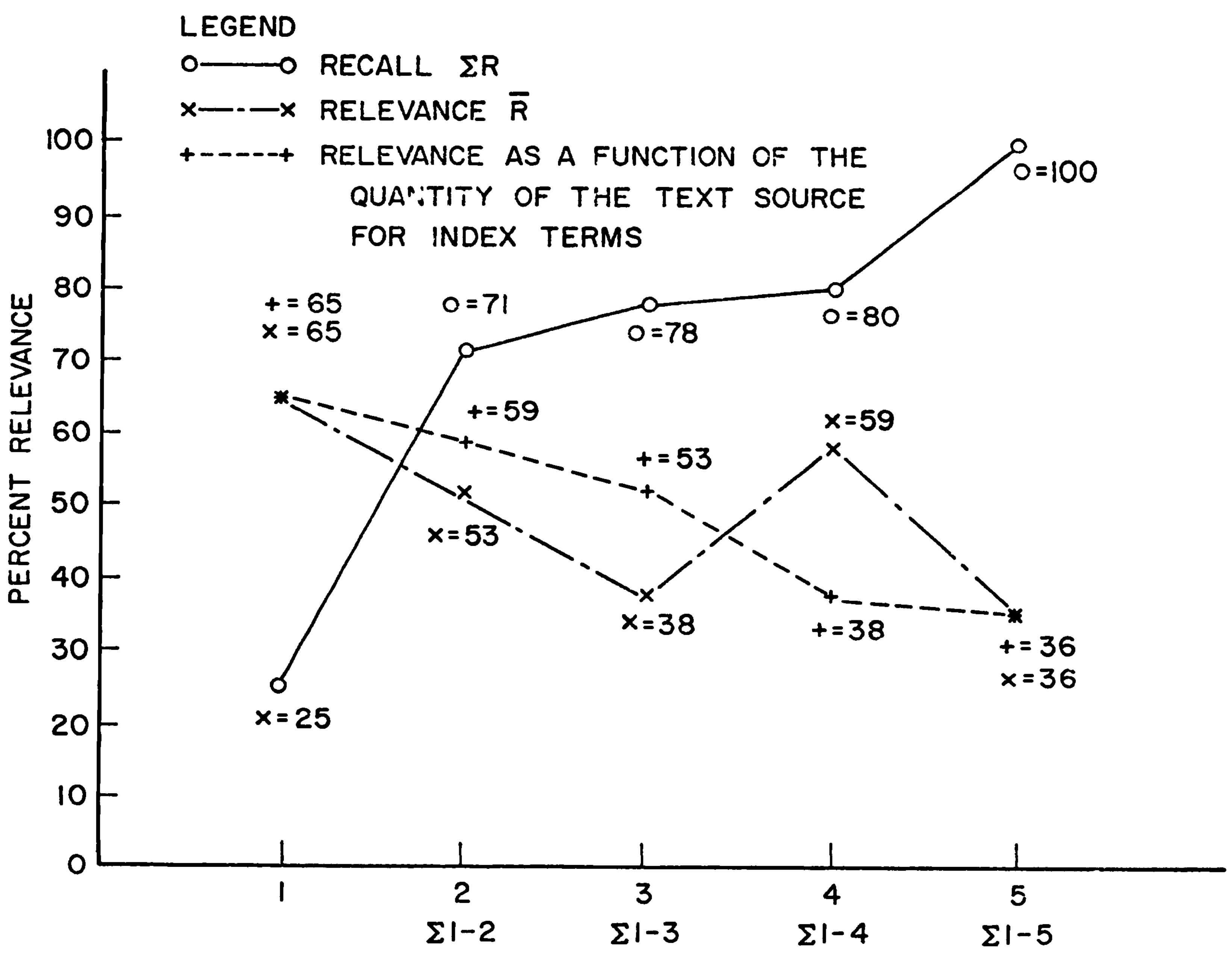

Figure 3. Relevance and Recall as a Function of Depth of Indexing

TABLE II

RELEVANT DOCUMENT RESPONSE AND CATEGORICAL RELEVANCE AS A FUNCTION OF THE SOURCE OF INDEX TERMS (CATEGORY)

Category	Relevant Document Response R	Categorical Relevance $\bar{R}$	Cumulative Values $\sum R$
1	25%	65%	25%
2	47%	53%	72%
3	7%	38%	79%
4	2%	54%	81%
5	19%	36%	100%
$\sum$1-5	100%	*	100%

* The standard definition of relevance is the ratio of the number of relevant documents retrieved to the total number of retrieved documents. For $\sum$1-5, i.e. $\sum \bar{R}$, the relevance is 50%

Analyzing the data, it can be seen that $\bar{R}$ is greatest for the title (65%), but this value varies considerably for the other categories. The significance of these data is that the index term sources are not particularly good discriminators as far as differentiating nonrelevant from relevant documents, although the general trend is toward lower $\bar{R}$ as the category increases. It should be pointed out that retrieval by the author keywords represents a high categorical relevance as would be expected. If one re-ordered the categories such that the categories correspond to the quantity of text therein, a relevance pattern would emerge such that as the quantity of text used as a source of index terms increased, the relevance would decrease. This is shown in Figure 3.

The implication of $\bar{R}$ is that one is about as likely to miss relevant documents by not indexing from the body of the document as by not indexing from the table of contents, and list of figures and tables. Indexing from the title, author keywords and abstracts, on the other hand should result in a good probability of retrieving relevant documents. The idealized situation described earlier in this section is not approximated in practice.

The relevant document response R, on the other hand, [illegible] showing the distribution of relevant documents [illegible]. One can see that 25% of the available relevant documents were retrieved based on indexing from the title alone, whereas fully 72% of the available relevant documents were retrieved from title and abstract indexing. Inclusion of indexing from the table of contents and list of figures and tables increased the retrieval of relevant documents from 72% to 79%; additional indexing from the body of the document resulted in further retrieval of 19% of the available relevant documents which would have been lost if the body had not been used as a source of indexing terms. If we assume that all available relevant documents were retrieved when the full text served as the source of index terms, it is interesting to plot the cumulative relevance and recall resulting from increasing depth of indexing, because this shows that increasing the depth of indexing increases recall to a lesser degree than might be intuitively anticipated, as is shown in Figure [illegible].

2.3 INDEXING TIME AS A FUNCTION OF THE SOURCE OF INDEX TERMS

The time required to index a document is directly related to the cost of indexing. It is important to consider the retrieval effectiveness as a function of index term source in terms of the time, and therefore cost, required for indexing from the various sources of terms. From the preceding discussions, we have seen that R and S are distributed as follows:

	S(%)	R(%)
Category 1	19	25
Category 2	43	47
Category 3	9	7
Category 4	2	2
Category 5	26	19

The body of the document (Category 5) represents the largest portion of text of the document. Therefore, at the outset, one could easily hypothesize that the major portion of the indexing time would be spent on the body. Yet the above results clearly indicate that index terms from the body of the document were responsible for only 26% of all the documents retrieved and only 19% of the relevant documents retrieved. The question yet to be answered is: what is the distribution of indexing time by category?

The answer to this question can be derived from an analysis of the data on indexing time for each category. These data were acquired by having each indexer record the stopwatch time required to index each document by the five categories. By knowing which documents and corresponding categories were retrieved on the searches, it was a relatively easy matter to determine the distribution of indexing time by category for precisely that subset of documents actually retrieved by the searches.

To facilitate handling the index time data and correlating the search output with index time, keypunch cards were prepared for each document in the data base. A print-out from the keypunched data was obtained listing documents in order by access number. From the search printout sheets, total indexing time per retrieved document was obtained for each search (see Table A-3). This process was repeated for each document retrieved for all searches. The result of the summation of indexing time by category for each search is shown in Table A-4. Table A-4 shows the distribution of total indexing time for the documents retrieved on a particular search among the various categories. Table III indicates the distribution of time spent in indexing those documents retrieved for the set of SDI searches run. The indexing time is symbolized by T. These data are shown graphically in Figure 4.

TABLE III

DISTRIBUTION OF TIME SPENT IN INDEXING THOSE DOCUMENTS RETRIEVED FOR THE SDI SEARCHES RUN ON THE EXPERIMENTAL DATA BASE

Category	Time Spent in Indexing T (%)	Cumulative Values
1	3	3
2	36	39
3	11	50
4	3	53
5	47	100

We now have data on retrieval response (S), relevant document response (R), and indexing time (T) by category. We can therefore determine the distribution of these values and make some inferences about the desirability of selective indexing, i.e., indexing by selected categories. Figure 5 shows the trend by cumulative category for all of these values.

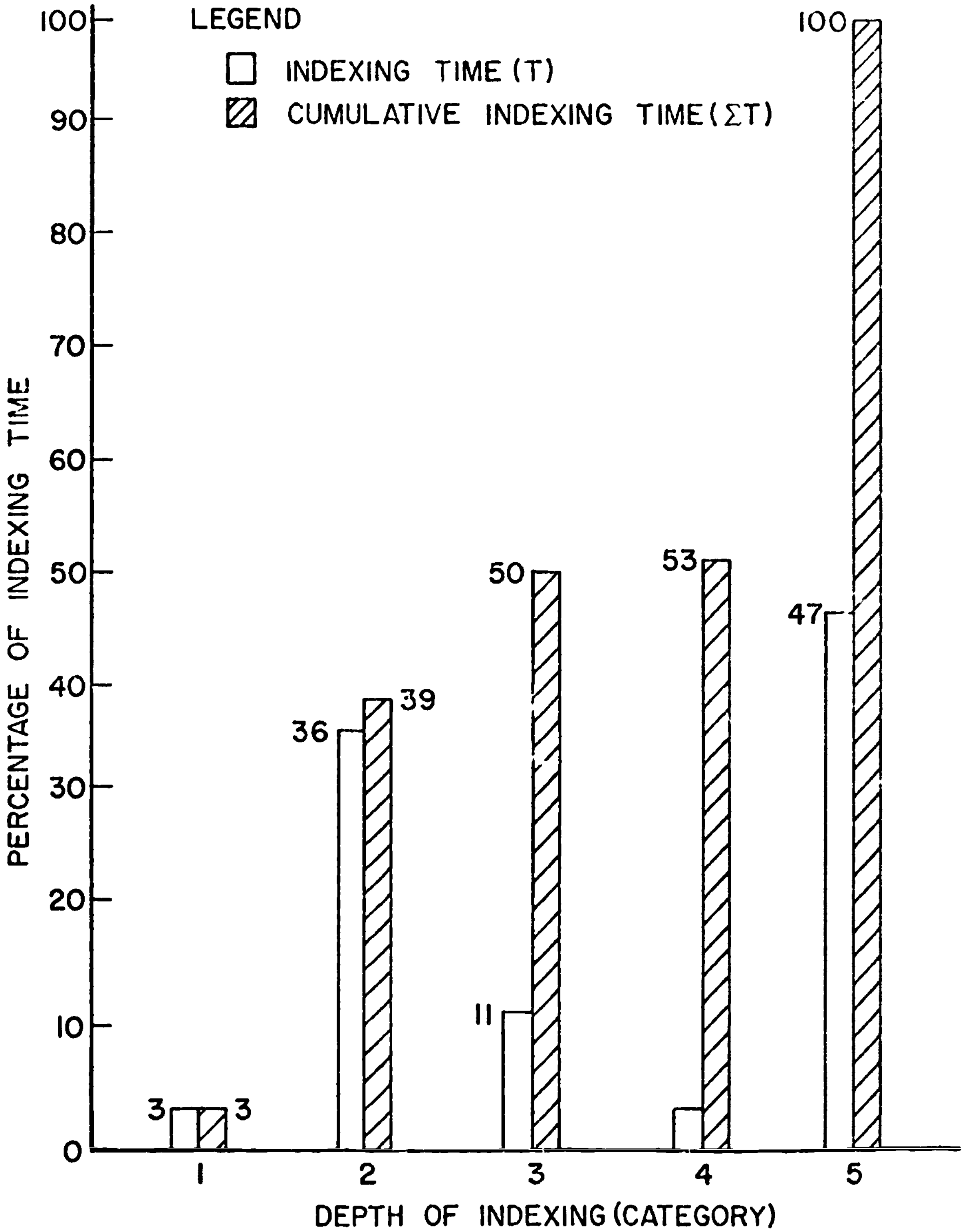

Figure 4. Distribution of Indexing Time by Category for the Documents Retrieved for the SDI Searches Run

It is interesting to consider the retrieval efficiency for various sources of index terms. The retrieval efficiency, E, is defined as the percentage of total documents retrieved divided by the percentage of the time required for indexing, i.e. $E = \frac{S}{T}$. The relevant document efficiency, $\overline{E}$, is defined as the percentage of relevant documents retrieved divided by the percentage of the time required for indexing, i.e., $\overline{E} = \frac{R}{T}$. These values are shown in Table IV.

TABLE IV

RETRIEVAL EFFICIENCY (E) AND RELEVANT DOCUMENT EFFICIENCY ($\overline{E}$) BY CATEGORY

Category	$E = \frac{S}{T}$	$\overline{E} = \frac{R}{T}$
1	6.34	8.33
2	1.19	1.30
3	0.82	0.64
4	0.67	0.67
5	0.55	0.40

From the above considerations we can state that indexing from the title is by far the most efficient, and that indexing from the body is rather inefficient. The reader is warned, however, that these values are interdependent rather than independent. In other words, the values for Categories 2 - 5 are dependent on the results from the preceding category. Our experiment really shows the added value achieved when additional portions of the documents are used as sources of index terms. The results would be somewhat different if the categories had been considered as mutually exclusive, completely independent entities serving as sources of indexing terms. However, our experimental design was formulated to indicate the additive efficacy of the sources of index terms. Our indexing philosophy and procedures follow a pattern such that one would always start with the title as a source of indexing terms and only then proceed to the abstract to determine additional index terms and thence to the table of contents and list of figures and tables for yet additional terms, etc.

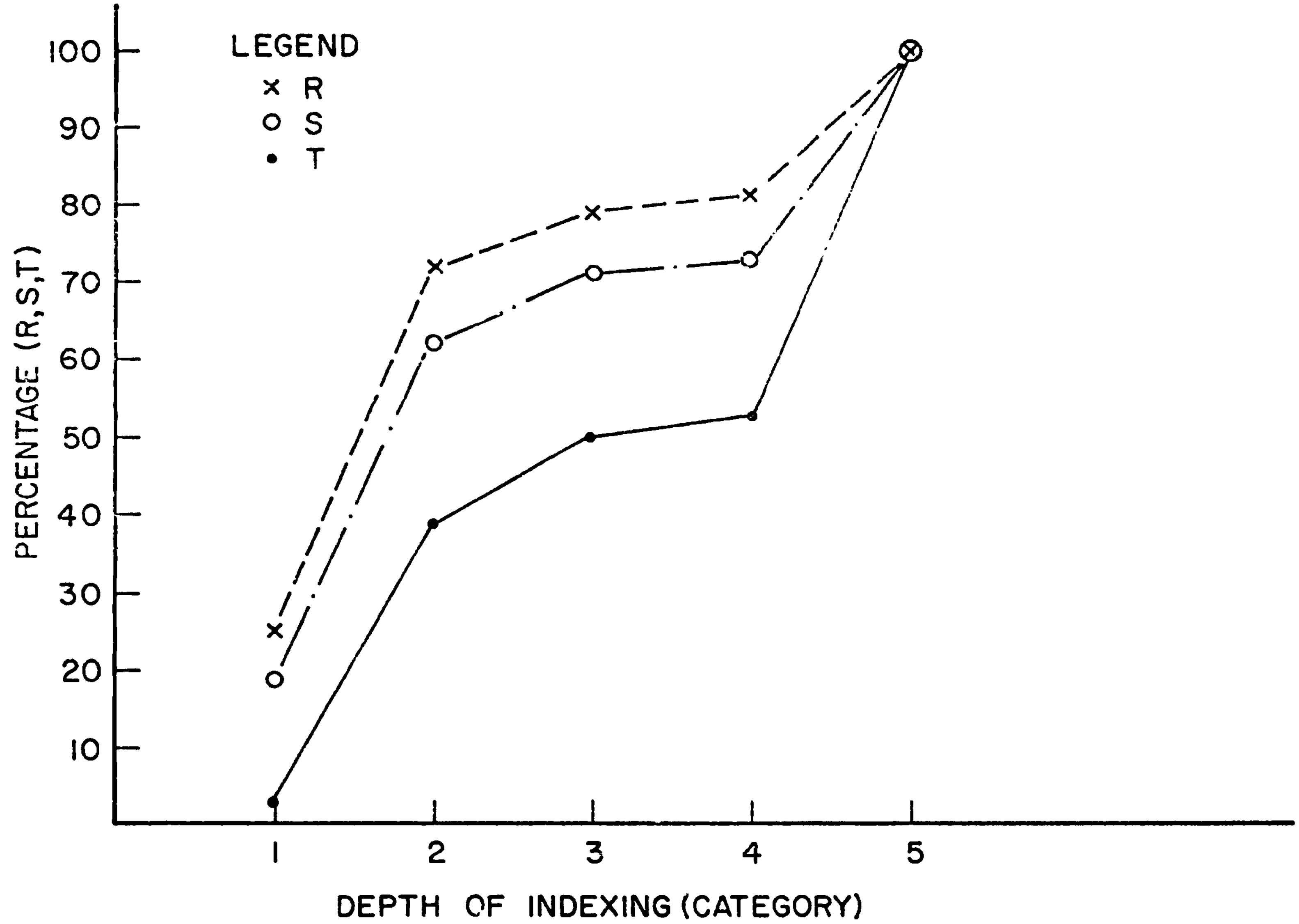

Figure 5. Cumulative R, S, and T by category

2.4 SUBSIDIARY INVESTIGATIONS

In addition to our primary study we wanted to investigate certain phenomena which would possibly influence the results. The data reported up to this point were with SDI searches. The nature of SDI searches is such that they tend to be more general in nature than retrospective searches. Therefore, since titles and abstracts, especially, tend to be more generally descriptive text than the body of the report itself, we expected that the results from SDI searches would tend to favor the less detailed indexing that necessarily would occur from using the titles and abstracts as index term sources. It was our original intent to test a mix of retrospective and SDI searches. However, the data base was so small (936 documents) that very few retrospective searches had any retrievals. We did obtain data for fifteen retrospective searches. These data are presented in Table V. The raw data are given in Tables A-5 to A-7 in Appendix A.

TABLE V

COMPARISON OF R, S, AND T FOR RETROSPECTIVE VS. SDI SEARCHES

Category	R		$\overline{R}$		S		T	
	SDI	Retro	SDI	Retro	SDI	Retro	SDI	Retro
1	25	7	65	100	19	5	3	1
2	47	50	53	64	43	54	36	42
3	7	2	38	33	9	5	11	8
4	2	2	59	50	2	3	3	2
5	19	39	36	80	26	33	47	48

From these data it can be seen that the times for indexing are virtually the same for both retrospective and SDI searches. The values for S, R, and $\overline{R}$ show significant differences. Interestingly, the document response, S, for the body of the document is about the same for both SDI and retrospective searches, but the category 1 relevance, $\overline{R}$, increases from 36% to 80%. Also, percentage of relevant documents retrieved by Category 5 indexing increases from 19% to 39%. It should be recognized that the data from retrospective searches is far less extensive than for the SDI searches, due both to the lower number of retrospective searches run and fewer retrievals per retrospective search. Nevertheless, the trends and differences are definitely significant. Table VI shows the efficiency factors for retrospective and SDI searches.

TABLE VI

COMPARISON OF E AND $\overline{E}$ FOR RETROSPECTIVE AND SDI SEARCHES

Category	$E = \frac{S}{T}$		$\overline{E} = \frac{R}{T}$	
	SDI	Retro	SDI	Retro
1	6.34	5.00	8.33	7.00
2	1.19	1.28	1.30	1.19
3	0.82	0.63	0.64	0.25
4	0.67	1.50	0.67	1.00
5	0.55	0.69	0.40	0.81

Another factor we wished to explore was the indexing time distribution for new, relatively inexperienced indexers vs. that for highly experienced indexers. We expected that the inexperienced indexer would tend to spend a particularly long time on indexing the body of the report. One new indexer at UDRI was selected and instructed to record time-category data for 50 documents just as the indexers in the primary experiment had done earlier. The results of the time distribution for experienced and inexperienced indexers is shown in Table VII.

TABLE VII

DISTRIBUTION OF INDEXING TIME BY CATEGORY FOR EXPERIENCED VS. INEXPERIENCED INDEXERS

Category	Experienced* (%)	Inexperienced (%)
1	13	11
2	38	20
3	10	11
4	6	6
5	32	53
Total time/Doc	24 min	47 min

*For all documents indexed; note difference between these values and T.

2.5 INPUT COSTS AS A FUNCTION OF THE SOURCE OF INDEX TERMS

An important consideration is the unit cost of input to the AMIC system. The indexing process represents a major portion of the costs of the operation of the AMIC system. Therefore any saving in the unit cost of input processing becomes significant for cost reduction when considered in terms of the number of documents added annually. A previous technical report[7] indicated that the cost for processing a report into the AMIC system is about $10 including clerical processing and indexing. The cost of indexing has been held down by using University students as paraprofessional indexers. These students are thoroughly trained in indexing by a student indexer training program developed[4] and validated[5] by the University of Dayton.

The cost of indexing (not including clerical processing) amounts to about $6.50 per document. From Table VII we can see that from 32% to 53% of stop-watch indexing time is expended in indexing using the body of the document as a source of indexing terms. The stopwatch time is not altogether realistic for determining actual indexing time, since there is a certain document handling time over and above the reported stopwatch time for indexing from the various portions of the document. Previous experience suggests that about one-third of the time required for indexing is document handling time. If we assume that approximately one third of the cost of indexing is taken up in the mechanics of document handling, then the cost corresponding to the actual stopwatch indexing time is about $4.33. From Table VII it is known that the indexing time per document can be reduced by 32% to 53% depending on the level of experience of the indexer. A reasonable approximation of the average is 40%. If we assume that the body of the document is eliminated as a source of index terms, then the cost of indexing (apart from document handling) can be reduced from $4.33 to $2.60, representing a saving of $1.73 per document. Considering an annual input of 5000 to 8000 documents per year it can be seen that an annual saving of $8,650 to $13,840 can result; this saving can be transformed into additional input into the system. Simply by eliminating the body as a source of index terms the unit cost of documents entered into the system would be reduced from $10 to $8.25.

2.6 CONCLUSIONS

An important purpose of this study was to assist us in a decision making situation. Specifically, we knew that for our system we could restrict the portions of the documents which would serve as sources of index terms, and thus reduce indexing time. As shown in the preceding section, this factor can be translated into reduced unit input cost. In order to make a decision concerning desirability of restricting the source of indexing terms, we needed experimental evidence to indicate the effect of restricting the index term sources. These effects, not only on indexing time, but also on retrieval effectiveness, needed to be determined. The data presented above provide us with the information we needed.

It should be pointed out that the data acquired in our experimental program apply to the AMIC system, and therefore the specific data and the analysis thereof may serve as guidelines for other situations and systems, but a precise transfer and application of our experience into another environment should not be expected. Also the concept of relevance must be approached with caution. Our previous experience[9] and the experiences of others have shown tendencies toward inconstancy and inconsistency of relevance judgments.

We found that the amount of time required for indexing from the first four categories amounted only to about 53%, whereas the time required for the body amounts to 47%. Yet for SDI searches, the required extra indexing effort results in an increase of relevant documents of 19%; for retrospective searches, an increase of 39% results, based on a limited number of retrospective searches. Thus there is a trade-off between the cost of input and retrieval effectiveness, i.e., the ability to retrieve the relevant documents. The tradeoff is more dramatic with SDI searches than with retrospective searches. It is interesting to note that the title and abstract alone achieves a relevant document response of 72% with an indexing effort of only 39%. By increasing the indexing effort also to include the table of contents and list of figures and tables, the relevant document response is increased to 81% with an additional indexing expenditure of 14%.

The figures listed above refer to SDI searches. [illegible] retro-spective searches, the tradeoff between retrieval effectiveness [illegible] indexing effort is not nearly as strong. However, one must [illegible] frequency of retrospective searches and SDI searches. SDI searches [illegible] periodically with each update. Retrospective search activity with [illegible] declining while the SDI program has increased in the number of participants, although there is not necessarily a cause-and-effect relationship [illegible].

In view of all the above factors, we have concluded that for [illegible] it is reasonable to use the first four categories as source of [illegible] and not to include the body of the report. By doing so we risk [illegible] of the available relevant documents, but we save 47% of the indexing [illegible], based on retrieved documents. Based on the indexing [illegible] (not just the subset retrieved), the indexing time savings [illegible] experienced indexers and 53% for inexperienced indexers. [illegible] are more appropriate, since in practice all documents are [illegible], obviously we could not predict the subset of documents which would be retrieved for all searches.

Figure 6 shows a composite of results considering a hypothetical subset of 100 retrieved documents, and comparing retrieval effectiveness for Categories $\sum 1-4$ with the additional retrieval provided by Category 5.

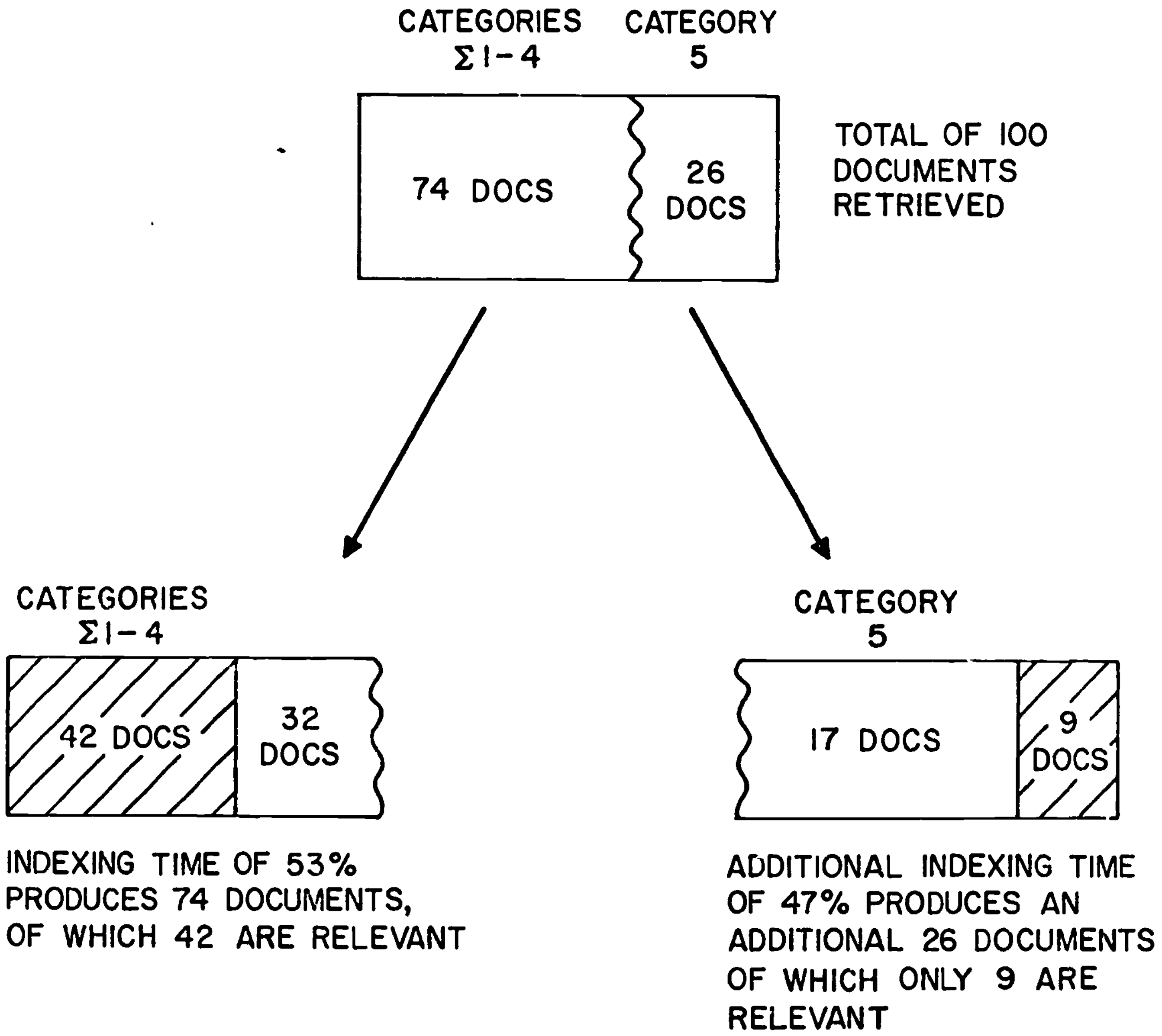

Figure 6. Retrieval and Distribution of a 100-Document Set, Comparing Categories $\sum$1-4 with the Additional Retrieval Effected by Category 5. Shaded areas represent relevant documents.

SECTION III

AMIC SYSTEM OPERATIONS

3.1 SUMMARY OF THE AMIC SYSTEM

In the normal course of document processing, AFML technical reports; documents on automatic distribution from AEC, DDC, NASA, and FTD; and other Government R&D publications of the Army, Navy, and Government contractors such as Westinghouse, The Boeing Co., the University of California, etc. are received at the UDRI's off-campus Materials Documentation Center, Bldg. 1', AFML. After checking for duplication and screening for relevance of the contents to the mission objectives of AFML, identifying accession numbers are assigned. The documents are then delivered to the University of Dayton campus.

At the University of Dayton the documents are indexed and abstracted by information scientists whose technical disciplines enable them to select pertinent keywords from a thesaurus displaying acceptable keywords. The indexed data from the documents is transferred to keypunch cards and then converted to magnetic tape for storage on the CDC 6600 computer in Bldg. 676 at WPAFB. The documents themselves are returned to the Materials Documentation Center, Bldg. 17, WPAFB for storage.

Retrospective searching for information requested by AFML personnel proceeds as follows: the request is made directly to an AMIC information specialist, either in person or over the phone. The Project Leader is present at WPAFB in Building 17 two days per week to take such requests. At other times, the contact can be made by phone. The search request is assigned to the appropriate information specialist for the formulation of an appropriate search strategy. This search data is provided to the CDC 6600 computer facility for a batch mode computer search of the data bank. Access numbers whose index data qualify them for retrieval according to the search strategy are retrieved. Access numbers, which identify documents, are printed out and returned to UDRI, where a file of abstracts is maintained for screening by the information specialist. Relevant abstracts are copied and sent on to the Materials Documentation Center for distribution to the AFML Requester.

The SDI searches are run periodically against the update data. In preparing an SDI profile, the AFML user discusses his subject request of continuing interest with an AMIC information specialist, through personal interviews. The statement of interest is processed into an SDI profile and the search data are prepared on magnetic tape for running on the CDC 6600 computer. The computer does a search of current update data only (information added to the data bank in the current and two previous years). Document access numbers corresponding to the SDI profile are used to select abstracts for copying. These abstracts are [illegible] Materials Documentation Center for distribution to the SDI users. [illegible] bank is updated approximately every six weeks. All of the processes [illegible] are presented in the form of charts and flow diagrams in Figures [illegible] 10.

3.2 REORGANIZATION OF THE AMIC

[illegible] improve the efficiency and effectiveness of the overall UDRI AMIC operation, [illegible] reorganization of the AMIC project was implemented during the report [illegible] to coordinate the activities of the off-campus Materials Documentation [illegible] maintained at the AFML more closely with those aspects of the AMIC operations performed on-campus. Specifically, the off-campus clerical operation was brought under the supervision of the UDRI Information Systems Section's Clerical Supervisor, thus bringing all clerical operations of the AMIC system together under one head.

Several advantages have accrued as a result of the reorganization. Clerical procedures and documentation methods have been made uniform for both on- and off-campus operations. Clerical personnel from each operation became more familiar with the activities of the other. The effect of the reorganization of the Aerospace Materials Information Center is to provide a more coordinated unit.

3.3 SDI PROGRAM

In an effort to enhance and expand the SDI program, a number of personal interviews were held with already active SDI participants as well as potential new clients. In interviews with the already active participants, feedback was obtained regarding the appropriateness of the abstracts distributed corresponding to the SDI search profiles. With the experience of having received abstracts over a period of time, it was often possible to pinpoint specific retrieval terms in the profile which were causing nonrelevant retrievals. In many other cases it was possible to ascertain terms which could be negated in order to suppress nonrelevant retrievals. A number of profiles were modified to incorporate these changes. In some cases profile modifications were made to reflect changes in the subject areas of interest of some individuals. These redirections in subject interest usually came about due to the phasing out of certain projects and the initiation of new projects, or because of personnel reassignments.

A number of referrals to other AFML persons who might be interested were made by active SDI users. These referrals were followed up with interviews and new SDI users were obtained by this method. By the end of the reporting period there were 166 active SDI participants. A list of SDI users is presented in Appendix B, and the SDI topics are given in Appendix C.

AMIC SERVICES AND FACILITIES

SERVICES

1. SUBJECT SEARCHES OF AMIC FILE OF INDEXED TECHNICAL REPORT LITERATURE
2. AUTOMATIC DISTRIBUTION OF CURRENT AWARENESS (SDI) ABSTRACTS FROM THE AMIC DATA BASE
3. ORDERING SPECIFIC DOCUMENTS NOT AVAILABLE IN AMIC
4. INITIATING SEARCHES OF OTHER LITERATURE SERVICES, e.g., NASA, DDC, DMIC, etc.
5. REPRODUCING MICROFICHE (MICROFICHE TO MICROFICHE)
6. PROVIDING PAPER COPIES OF CHARTS, GRAPHS, TITLE PAGE OR SELECTED PORTIONS OF MICROFICHE.

HOLDINGS

ABOUT 65,000 DOCUMENTS ON MATERIALS RESEARCH AND TECHNOLOGY ARE INDEXED AND ABSTRACTED. THESE DOCUMENTS ARE ON FILE IN BLDG. 17. IN ADDITION TO REGULAR R&D REPORTS, STATE-OF-THE-ART, SYMPOSIUM, BIBLIOGRAPHY AND HANDBOOK REPORTS ARE INCLUDED. SUBJECT SEARCHES ARE RUN ON THE CDC 6600 COMPUTER AT THE ASD COMPUTER FACILITY. DOCUMENTS CAN ALSO BE RETRIEVED THROUGH THE MATERIALS DOCUMENTATION CENTER BY:

1. AUTHOR
2. REPORT NUMBER
3. CONTRACT NUMBER
4. CONTRACTOR
5. SPONSORING AGENCY

Figure 7. Description of AMIC Services and Holdings.

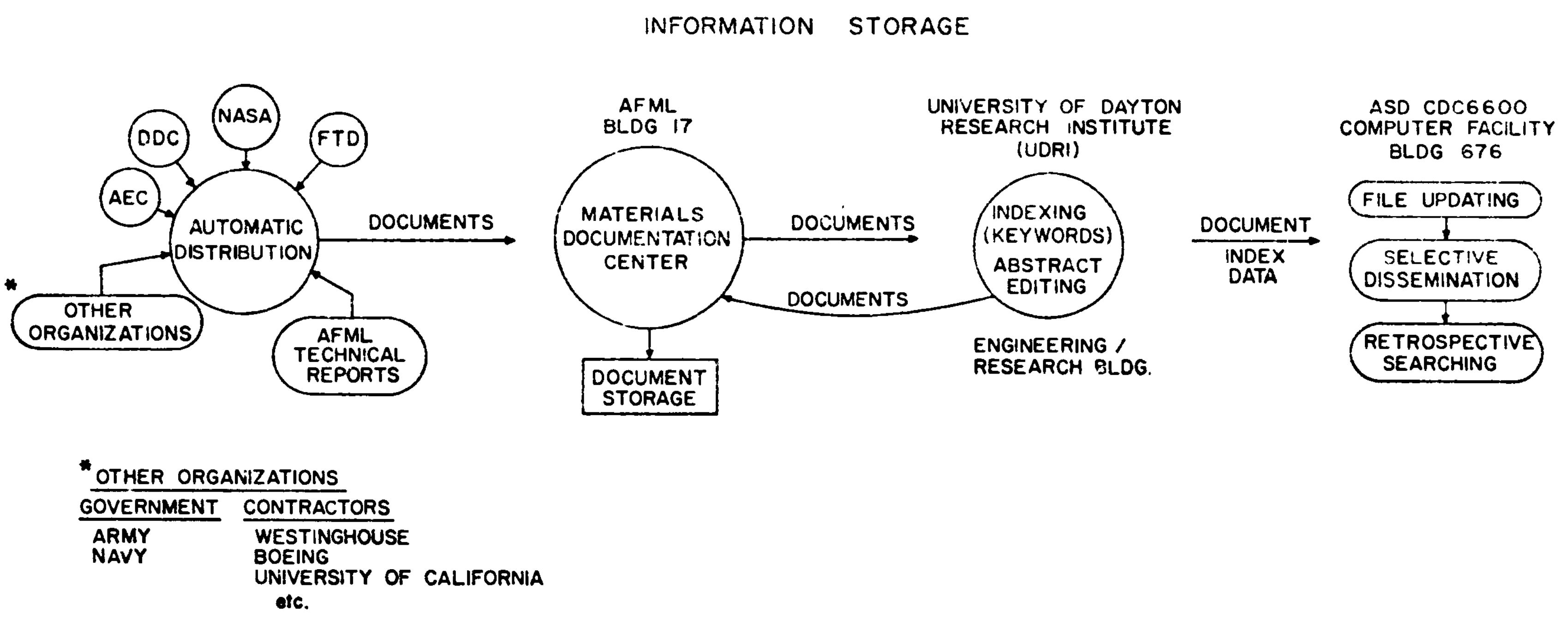

Figure 8. Processing of Documents into the AMIC System.

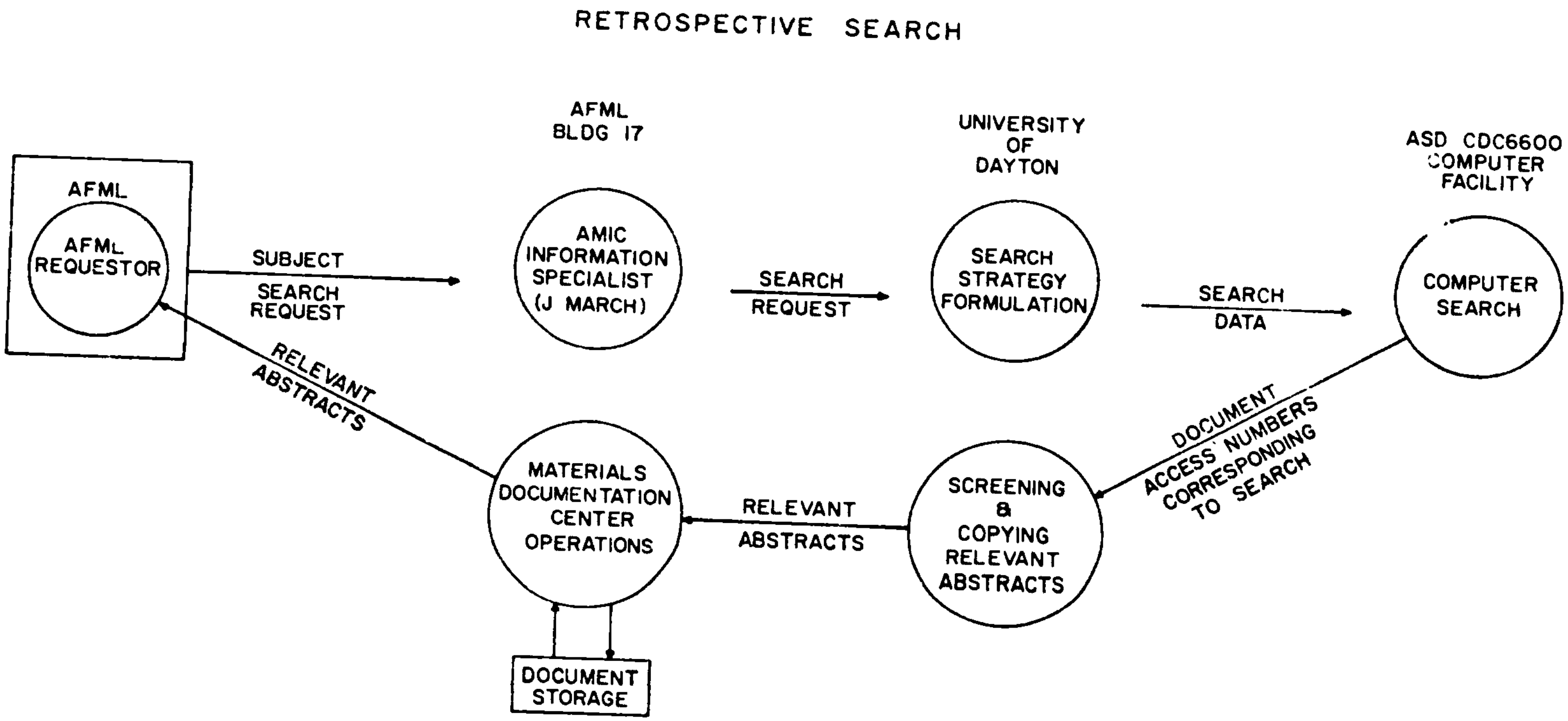

Figure 9. Processing of Retrospective Search Requests.

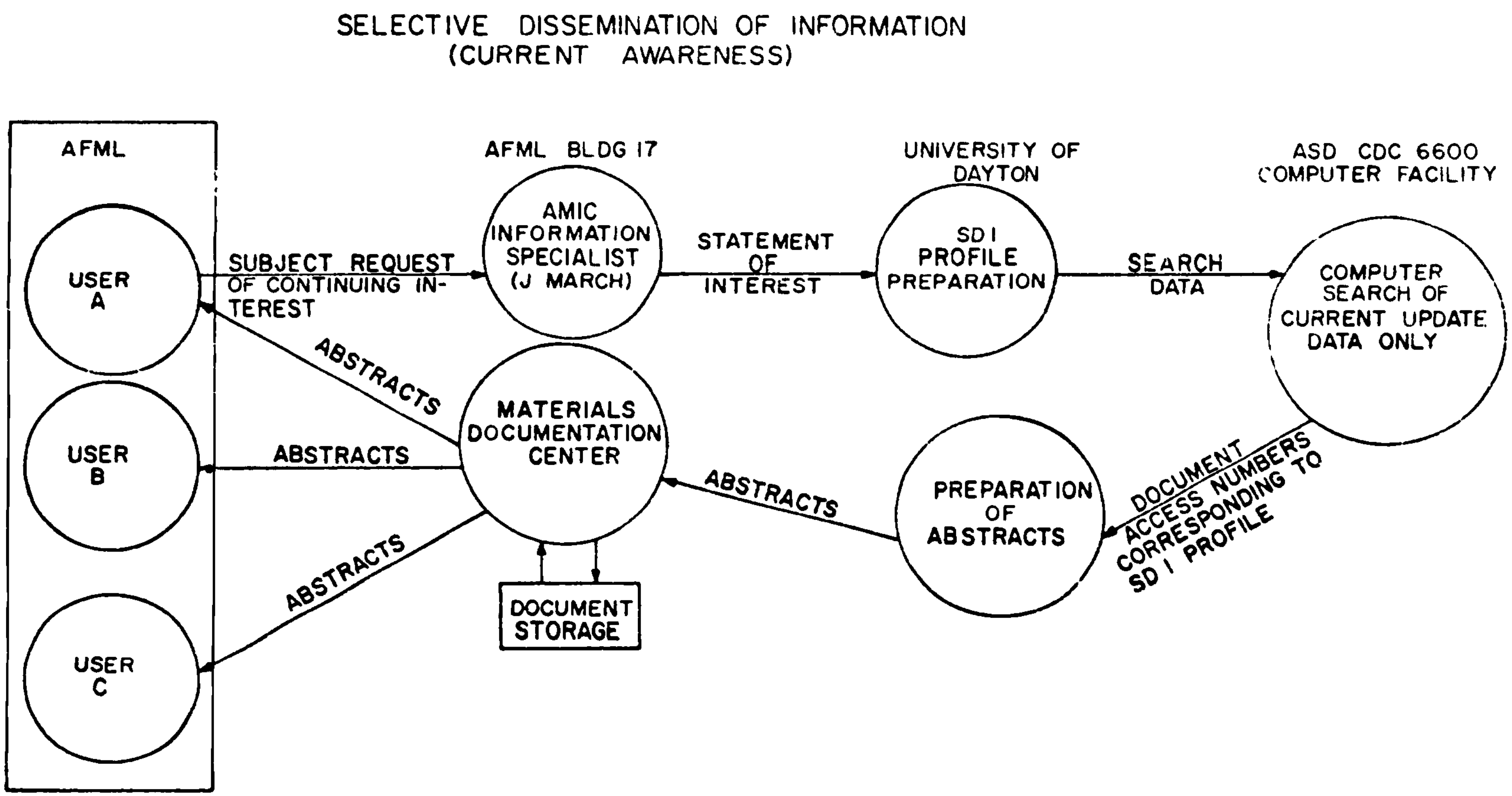

Figure 10. Processing of SDI Searches.

3.4 INPUT

During the period covered by this report, 1 Dec 71 through 30 Nov 72, 5704 documents were indexed and processed into the system. Of this number, 59 were handbooks, 139 were state-of-the-art reports, 58 were bibliographies, and 44 were symposium proceedings or papers. The documents were indexed with an average of 20.6 terms per document (exclusive of automatic generic postings) with an average indexing time of 32.3 minutes. There are now 65,467 documents in the AMIC document retrieval system. The distribution by subject category is given in Table E-1. The subject category definitions are shown in Appendix D.

3.5 SEARCHING

A total of 99 retrospective technical requests were processed by the Information Systems Section during the report period. An average of 24.3 abstracts was printed per search for forwarding to the search requesters. A list of retrospective search topics is given in Table E-2.

REFERENCES

1. E. A. Janning, Establishment of a Coordinate Indexing Retrieval System for the Air Force Materials Laboratory, RTD-TDR-63-4263, (AD 428 423), Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio. November 1963.

2. E. A. Janning, The Modification of an Information Retrieval System by Improving Vocabulary Control, Indexing Consistency, and Search Capabilities, AFML-TR-65-20, (AD 613 301), Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio. March 1965.

3. E. A. Janning, Operations of a Document Retrieval System Using a Controlled Vocabulary, AFML-TR-66-36, (AD 633 614), Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio. March 1966.

4. F. L. Scheffler, Student Indexer Training Program and the Improved Operations of a Document Retrieval System, AFML-TR-66-391, (AD 651 039), Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio. January 1967.

5. F. L. Scheffler, Indexer Performance Analysis and Operations of a Document Retrieval System, AFML-TR-67-379, (AD 666 462), Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio. February 1968.

6. F. L. Scheffler, and R. B. Smith, Document Retrieval System Operations Including the Use of Microfiche and the Formulation of a Computer Aided Indexing Concept, AFML-TR-68-367, (AD 686 804), Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio. February 1969.

7. F. L. Scheffler and J. F. March, User Appraisal and Cost Analysis of the Aerospace Materials Information Center, AFML-TR-70-27, (AD 670 597), Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio. March 1970.

8. F. L. Scheffler and J. F. March, Evaluation of the Selective Dissemination of Information (SDI) Program for the Aerospace Materials Information Center, AFML-TR-71-11, (AD 725 036), Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio. March 1971.

9. F. L. Schaffler and J. F. March, Determination of the Consistency of Relevance Judgments and the Reliability of Search Strategies Among Information Specialists for the Aerospace Materials Information Center, AFML-TR-72-51, (AD 751 977), Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio. April 1972.

10. H. P. Luhn, "Key-Word-In-Context Index for Technical Literature (KWIC Index)", American Documentation, Vol. 11, No. 4, pp 288-295, October 1960.

11. W. L. Miller, "The Efficiency of MEDLARS Titles for Retrieval", Journal of the American Society for Information Science, Vol. 22, No. 5, pp 318-321, Sept-Oct 1971.

12. B. V. Tell, "Retrieval Efficiency from Titles and the Cost of Indexing", Information Storage and Retrieval, Vol. 7, No. 5, pp 241-243, December 1971.

13. F. W. Lancaster, Vocabulary Control for Information Retrieval, Washington, D.C.: Information Resources Press 1972.

14. B. V. Tell, "Document Representation and Indexer Consistency - A Study of Indexing from Titles, Abstracts, and Full Text using UDC and Keywords", Proceedings of the American Society for Information Science, pp 285-292, 1969.

APPENDIX A

EXPERIMENTAL DATA

TABLE A-1
RETRIEVAL RESPONSE (S) BY INDEXING SOURCE CATEGORY

Search No.	Total Documents Retrieved	Retrieval Response S Category 1	2	3	4	5
99005	23	3	13	2	1	4
99006	23	3	7	3	0	10
99019	102	37	43	5	1	16
99022	18	2	12	0	0	4
99024	31	13	9	4	1	4
99045	19	6	7	3	1	2
99046	10	4	6	0	0	0
99047	4	1	2	0	0	1
99070	4	0	4	0	0	0
99071	19	4	8	4	0	3
99075	62	18	35	3	2	4
99077	34	1	13	2	1	17
99080	18	6	7	4	0	1
99082	5	0	3	1	0	1
99083	8	0	5	1	0	2
99085	5	2	2	0	0	1
99087	2	1	1	0	0	0
99088	1	0	1	0	0	0
99197	19	7	9	2	0	1
99199	7	3	4	0	0	0
99200	1	0	0	0	0	1
99201	9	1	2	2	0	4
99203	75	6	21	9	0	39
99204	4	0	2	1	0	1
99215	5	2	1	0	2	0
99218	6	1	2	3	0	0
99219	26	7	10	0	0	9
99220	8	1	4	1	0	2
99221	20	0	4	5	1	10

TABLE A-1 continued

Search No.	Total Documents Retrieved	Retrieval Response S				
		Category				
		1	2	3	4	5
99222	7	1	3	0	1	2
99223	7	0	6	0	0	1
99224	14	0	8	1	0	5
99225	74	11	40	6	1	16
99226	5	1	1	2	0	1
99227	19	1	7	5	0	6
99228	17	0	12	1	0	4
99229	9	0	0	0	0	9
99231	3	0	3	0	0	0
99232	32	10	11	4	0	7
99234	4	0	3	1	0	0
99235	6	0	3	1	0	2
99236	11	2	5	1	0	3
99237	34	6	10	7	0	11
99238	35	27	5	2	1	0
99239	19	5	9	3	1	1
99241	42	1	17	6	1	17
99278	63	12	27	5	3	16
99286	50	0	19	5	0	26
99288	23	8	12	1	0	2
99289	44	19	14	2	1	8
99290	5	0	4	0	0	1
99291	21	4	11	5	0	1
99292	13	3	4	2	2	2
99353	3	0	0	0	0	3
99354	3	0	1	1	0	1
99355	1	1	0	0	0	0
99357	1	0	0	0	0	1
99361	3	3	0	0	0	0

TABLE A-1 continued

Search No.	Total Documents Retrieved	Retrieval Response S				
		Category				
		1	2	3	4	5
99365	74	1	26	9	1	37
99366	21	0	9	2	0	10
99367	17	3	8	1	0	5
99368	51	7	24	9	0	11
99369	14	1	4	4	0	5
99374	19	4	12	0	0	3
99376	51	7	23	3	2	16
99377	12	1	4	2	0	5
99378	258	54	131	18	3	52
99379	1	0	0	0	1	0
99380	8	0	4	1	1	2
99385	12	2	6	1	0	3
99386	3	0	2	1	0	0
99387	3	0	2	0	0	1
99389	1	0	1	0	0	0
99390	1	0	1	0	0	0
99391	2	1	1	0	0	0
99392	5	0	2	0	0	3
99394	1	0	0	0	0	1
99395	1	0	1	0	0	0
99397	28	5	12	2	2	7
99399	2	0	1	0	0	1
99401	7	2	1	0	0	4
99402	90	10	40	4	1	35
99403	2	0	2	0	0	0
99405	4	0	0	2	1	1
994[illegible]	3	0	0	0	0	3
[illegible]	26	5	11	2	0	8
99413	1	0	1	0	0	0

TABLF A-1 continued

Search No.	Total Documents Retrieved	Retrieval Response S				
		Category				
		1	2	3	4	5
99415	13	1	7	0	0	5
99416	2	0	1	0	0	1
99421	11	1	6	0	0	4
99422	8	0	5	0	0	3
99424	6	2	2	0	0	2
99430	49	11	26	4	0	8
99432	8	0	5	2	0	1
99434	33	14	11	3	1	4
99436	12	5	4	0	0	3
$\sum$ Searches	1996	381	868	186	34	527
S(%)		19%	43%	9%	2%	26%
$\sum$S		19	62	71	73	99

TABLE A-2
RELEVANT DOCUMENT RESPONSE (R) AND CATEGORICAL RELEVANCE ($\bar{R}$) BY INDEXING SOURCE CATEGORY

Search No.	Total Documents Retrieved	Retrieval Response R				
		Category				
		1	2	3	4	5
99005	12	2	8	1	0	1
99006	16	3	3	3	0	7
99019	24	6	10	1	1	6
99022	6	1	3	0	0	2
99024	17	9	6	0	0	2
99045	9	5	4	0	0	0
99046	6	3	3	0	0	0
99047	1	1	0	0	0	0
99070	0	0	0	0	0	0
99071	17	3	7	3	0	4
99075	62	18	35	3	2	4
99077	27	1	9	3	1	13
99080	12	5	3	4	0	0
99082	3	0	2	1	0	0
99083	4	0	3	1	0	0
99085	2	1	1	0	0	0
99087	1	1	0	0	0	0
99088	1	0	1	0	0	0
99197	9	3	4	2	0	0
99199	3	2	1	0	0	0
99200	0	0	0	0	0	0
99201	3	1	1	0	0	1
99203	16	4	6	1	0	5
99204	2	0	2	0	0	0
99215	4	2	1	0	1	0
99218	4	1	2	1	0	0
99219	10	6	2	0	0	2
99220	3	0	2	1	0	0
99221	10	0	5	1	1	3

TABLE A-2 continued

Search No.	Total Documents Retrieved	Retrieval Response R				
		Category				
		1	2	3	4	5
99222	6	1	3	0	1	1
99223	5	0	4	0	0	1
99224	5	0	5	0	0	0
99225	29	6	14	1	1	7
99226	5	1	1	2	0	1
99227	8	0	3	2	0	3
99228	9	0	9	0	0	0
99229	1	0	0	0	0	1
99231	3	0	3	0	0	0
99232	12	3	6	1	0	2
99234	1	0	1	0	0	0
99235	3	0	2	0	0	1
99236	8	2	5	1	0	0
99237	12	5	5	0	0	2
99238	34	26	5	2	0	1
99239	4	3	1	0	0	0
99241	8	0	4	1	1	2
99278	32	8	16	3	1	4
99286	15	0	6	1	0	8
99288	19	8	10	1	0	0
99289	18	10	5	0	1	2
99290	1	0	1	0	0	0
99291	8	2	6	0	0	0
99292	1	0	0	1	0	0
99353	2	0	0	0	0	2
99354	1	0	1	0	0	0
99355	0	0	0	0	0	0
99357	0	0	0	0	0	0
99361	3	3	0	0	0	0

TABLE A-2 continued

Search No.	Total Documents Retrieved	Retrieval Response R				
		Category				
		1	2	3	4	5
99365	42	1	23	4	0	14
99366	10	0	5	0	0	5
99367	9	1	5	1	0	2
99368	10	7	3	0	0	0
99369	10	1	3	3	0	3
99374	13	3	8	0	0	2
99376	25	7	10	2	1	5
99377	3	0	1	1	0	1
99378	121	32	60	8	3	18
99379	0	0	0	0	0	0
99380	7	0	4	1	1	1
99385	5	0	4	1	0	0
99386	1	0	1	0	0	0
99387	3	0	2	0	0	1
99389	1	0	1	0	0	0
99390	1	0	1	0	0	0
99391	2	1	1	0	0	0
99392	3	0	2	0	0	1
99394	1	0	0	0	0	1
99395	1	0	1	0	0	0
99397	18	1	11	0	1	5
99399	0	0	0	0	0	0
99401	1	1	0	0	0	0
99402	58	8	27	1	2	18
99403	0	0	0	0	0	0
99405	1	0	0	0	1	0
99408	5	0	0	0	0	5
99410	15	4	7	1	0	3
99413	1	0	1	0	0	0
99415	10	0	5	0	0	5

TABLE A-2 continued

Search No.	Total Documents Retrieved	Retrieval Response R				
		Category				
		1	2	3	4	5
99416	1	0	1	0	0	0
99421	3	0	2	0	0	1
99422	3	0	2	0	0	1
99424	5	2	2	0	0	1
99430	40	11	18	4	0	7
99432	1	0	1	0	0	0
99434	17	8	6	1	0	2
99436	7	4	3	0	0	0
Rel. Docs, R	991	248	461	70	20	190
R(%)		25%	47%	7%	2%	19%
$\overline{R}$(%)	50%	65%	53%	38%	59%	36%

TABLE A-3

INDEXING TIME (IN MINUTES) CORRESPONDING TO SOURCE OF INDEX TERM
(Sample Format of Data)

Item	Document Access Number	Category 1 Title	Category 2 Abstract	Category 3 T of C , Figs Tables	Category 4 Author KW	Category 5 Body	Total
1	028093	3	4	6	5	6	24
2	033029	1	5	0	0	2	8
3	034001	1	2	0	1	1	5
4	034008	2	3	0	2	8	15
981	200396	1	6	4	2	1	14
982	200397	1	5	2	0	1	9
983	200398	1	5	3	2	1	12
984	208094	2	5	0	0	5	12

TABLE A-4
DISTRIBUTION OF TIME SPENT IN INDEXING OF THOSE DOCUMENTS RETRIEVED ON THE SEARCHES RUN

Search No.	Docs Retrieved	Summation of Indexing Time by Indexing Source Category					
		1	2	3	4	5	Total
99005	23	3	42	12	10	37	104
99006	23	6	47	26	1	86	166
99019	102	43	361	44	68	396	912
99022	18	2	55	3	0	66	126
99024	31	16	103	24	15	105	263
99045	19	9	50	35	12	36	142
99046	10	7	37	0	1	19	64
99047	4	2	14	3	2	6	27
99070		0	20	0	0	9	29
99071	19	7	70	21	4	87	189
99075	62	27	296	88	39	297	747
99077	34	2	108	32	8	194	344
99080	18	9	80	28	11	85	213
99082	5	0	15	6	0	18	39
99083	8	0	21	4	1	17	43
99085	5	3	17	0	1	15	36
99087	2	2	8	2	2	4	18
99088	1	0	6	0	0	5	11
99197	19	11	78	29	11	86	215
00100	7	5	24	5	1	31	66
99200	1	0	0	0	0	8	8
99201	9	2	18	19	3	51	93
99203	75	9	139	71	2	319	540
99204	4	0	9	2	0	5	16
99215	5	4	22	0	3	11	40
99218	6	2	26	18	0	14	60
99219	6	11	74	3	6	51	145
99220	8	4	13	7	1	13	38
99221	20	0	29	1	1	43	74

TABLE A-4 continued

Search No.	Docs Retrieved	Summation of Indexing Time by Indexing Source Category					
		1	2	3	4	5	Total
99222	7	1	18	0	1	11	31
99223	7	0	27	5	0	17	49
99224	14	0	53	0	0	68	121
99225	74	16	230	66	2	251	565
99226	5	2	9	22	3	28	64
99227	19	2	37	34	2	83	158
99228	17	0	55	5	0	58	118
99229	9	0	0	0	0	81	81
99231	3	0	18	0	0	16	34
99232	32	16	115	41	3	112	287
99234	4	0	15	15	0	17	47
99235	6	0	15	3	0	28	46
99236	11	2	43	8	0	45	98
99237	34	6	[illegible]	7	3	101	131
99238	35	31	137	48	24	91	331
99239	19	8	70	33	9	67	187
99241	42	2	134	64	9	190	399
99278	63	17	138	42	15	191	403
99286	50	0	88	55	3	244	390
99288	23	15	119	32	24	99	289
99289	44	34	181	42	12	117	386
99290	5	0	26	0	0	13	39
99291	21	5	90	26	1	71	193
00202	13	4	20	7	3	51	85
99353	3	0	0	0	0	13	13
99354	3	0	7	3	0	7	17
99359	1	0	0	0	0	0	10
99357	1	0	0	0	0	1	1
99361	3	4	19	0	3	11	37

TABLE A-4 continued

Search No.	Docs Retrieved	Summation of Indexing Time by Indexing Source Category					
		1	2	3	4	5	Total
99365	74	2	207	106	5	429	749
99366	21	0	59	17	0	116	192
99367	17	5	59	26	17	101	218
99368	51	7	178	57	22	266	530
99369	14	2	18	9	3	45	77
99374	19	10	104	32	24	117	287
99376	51	12	183	49	13	298	555
99377	12	2	28	24	10	72	136
99378	258	95	966	200	73	1218	2552
99379	1	0	0	0	1	5	6
99380	8	0	18	7	1	27	53
99385	12	5	47	14	4	59	129
99386	3	0	9	2	0	9	20
99387	3	0	9	0	0	22	31
99389	1	0	4	0	0	2	6
99390	1	0	5	0	0	4	9
99391	2	4	4	4	0	8	20
99392	5	0	11	5	0	10	26
99394	1	0	0	0	0	5	5
99395	1	0	4	0	0	2	6
99397	28	10	103	30	12	107	262
99399	2	0	4	10	0	17	31
99401	7	3	13	0	1	53	70
99402	90	16	257	61	14	434	782
99403	2	0	5	0	0	2	7
99405	4	0	0	11	2	23	36
99408	3	0	0	0	0	6	6
99410	25	6	69	12	4	103	194
99413	1	0	4	0	0	5	9

TABLE A-4 continued

Search No.	Docs Retrieved	Summation of Indexing Time by Indexing Source Category					
		1	2	3	4	5	Total
99415	13	0	51	9	0	63	123
99416	2	0	10	0	0	8	18
99421	11	2	44	0	0	63	109
99422	8	0	40	7	0	53	100
99424	6	4	30	0	3	29	66
99430	49	22	218	81	31	239	591
99432	8	0	28	4	1	21	54
99434	33	27	137	34	11	104	313
99436	12	5	60	8	2	49	124
Time	1996	590	6447	1860	574	8400	17,871
Distribution by Category (%)		3%	36%	11%	3%	47%	100%
Time/Doc = 10 min.							

TABLE A-5
DOCUMENTS RETRIEVED FROM EXPERIMENTAL DATA BASE FOR RETROSPECTIVE SEARCHES

Search No.	Total Documents Retrieved	Retrieval Responses				
		Category				
		1	2	3	4	5
12603	14	1	8	1	0	4
12629	1	0	1	0	0	0
12630	2	0	1	0	0	1
12636	1	0	1	0	0	0
22634	7	0	3	0	0	4
22635	14	2	7	0	1	4
32428	3	0	1	0	1	1
32604	4	0	2	0	0	2
32605	3	0	2	0	0	1
42543	5	0	3	1	0	1
42637	2	0	0	0	0	2
62504	1	0	0	1	0	0
92594	1	0	1	0	0	0
92601	1	0	1	0	0	0
92624	2	0	2	0	0	0
	61	3	33	3	2	20

TABLE A-6
RELEVANT DOCUMENTS RETRIEVED FROM THE EXPERIMENTAL DATA BASE FOR RETROSPECTIVE SEARCHES

Search No.	Total Documents Retrieved	Retrieval Response				
		Category				
		1	2	3	4	5
12603	9	1	5	0	0	3
12629	0	0	0	0	0	0
12630	2	0	1	0	0	1
12636	0	0	0	0	0	0
22634	5	0	2	0	0	3
22535	12	2	6	0	1	3
32528	2	0	1	0	0	1
32604	1	0	0	0	0	1
32605	3	0	2	0	0	1
42543	5	0	3	1	0	1
42637	2	0	0	0	0	2
62504	0	0	0	0	0	0
92594	1	0	1	0	0	0
92601	0	0	0	0	0	0
926	0	0	0	0	0	0
	42	3	21	1	1	16

TABLE A-7
DISTRIBUTION OF TIME SPENT IN INDEXING OF THOSE DOCUMENTS RETRIEVED FOR THE RETROSPECTIVE SEARCHES RUN

Search No.	Docs Retrieved	Summation of Indexing Time by Indexing Source Category					
		1	2	3	4	5	Total
12603	14	2	55	6	0	50	113
12624	1	0	6	4	0	6	16
12630	2	0	5	0	0	12	17
12636	1	0	5	0	0	0	5
22634	7	0	20	3	0	43	66
22635	14	3	45	0	4	55	107
32528	3	0	7	0	5	18	30
32604	4	0	16	20	0	8	44
32605	3	0	12	0	0	8	20
42543	5	0	17	8	0	32	57
42637	2	0	0	0	0	4	4
62504	1	0	0	3	0	0	3
92594	1	0	4	0	0	3	7
92601	1	0	10	0	0	5	15
92624	2	0	14	0	0	5	14
	61	5	216	44	9	249	518

APPENDIX B

SDI REQUESTERS

SDI

REQUESTER INDEX

REQUESTER	ORGANIZATION
Adair, A M	AFML/LL
Allinikov S D	AFML/MXE
Anderson, C S	AFML/LTF
Anspach, W F	AFML/LNE
Arnold, F E	AFML/MBP
Arnson H L	AFML/LLS
Askins, D. R	UDRI
Auman, G W	AFML/LTE
Bentley, F F	AFML/LP
Benz, R S	AFML/LTF
Bertke, R S	UDRI
Bialrt. M.	AFML/LTE
Blakeslee, H W	Franklin Institute Research Lab.
Baynton T A	AFML/LTE
Browning, C E	AFML/MBC
Buckley, M J.	AFML/LL
Campbell, G L	AFML/LTM
Champa, R A	AFML/LPH
Clark, L	AFML/LTM
Cohen B.	AFML/MXA
Corbly, D M.	AFML/LLN
Crane, R L	AFML/LL
Crawford, W J	AFML/LPA
Crosby, J J	AFML/LL
Cunningham, A.	Lockheed-Georgia Co.
Davidson, J E	UDRI
Davis, K. A	AFML/LN
Davis, S O	AFML/LL
Denman, G L	AFML/MXS
Denson, D D.	AFML/MBP

REQUESTER	ORGANIZATION
DePierre, V	AFML/LL
Dimiduk P W	AFML/LP
Donlan, V L	AFML/LPE
Drzal L T	AFML/LNX
Dueweke, P W	UDRI
Duvall, D	UDRI
Ekman, W J.	AFML/MBP
Emrich B R.	ASD/YHEF
Engle, A G.	UDRI
Evers, R C.	AFML/MBP
Ezekiel, H M	AFML/LNF
Farmer, R W	AFML/MBC
Fiscus, I	UDRI
Frederick, W G D	AFML/LPE
Fujishiro, S.	AFML/LL
Garrett, H J	AFML/LTE
Gehatia, M T	AFML/MBP
Geisendorfer, R F.	AFML/LL
Glenn, G. M.	AFML/LTM
Gloor, W. H.	AFML/LN
Goldberg, W	AFML/LP
Goldfarb, I. J	AFML/MBP
Grandt, A F	AFML/LL
Grant, R.	UD
Graves, R	UDRI
Haggard, D K	AFML/LL
Hall, J A	AFML/LL
Harmer, R S	UDRI
Haury, G L.	AFML/LPH
Headrick, R. E.	AFML/MBE
Hecht, N.	UDRI

REQUESTER	ORGANIZATION
Heinrich, J. P.	AFML/LPF
Helminiak, T. E	AFML/MBP
Hemenger, P. M	AFML/LPE
Henderson, J. P	AFML/LL
Hickmott, J. P.	AFML/LPE
Hollenberg, G	AFML/LI
Hopkins, A. K	AFML/LPH
House, P. L.	AFML/MXE
Hutchens,	AFML/LPE
Iller, W. J	AFML/MXA
Iglauer, N	AFML/LP
Jerina, K. L.	AFML/MBE
Johnson, W. P	AFML/LNE
Jumper, G	AFML/MXS
Kennard, R	AFML/LTM
Kirkpatrick, N. B	AFML/LL
Klarquist, N. E	AFML/LLM
Knight, M	AFML/MXE
Koenig, J. R	AFML/MXS
Kopell, L	AFML/LTM
Kuhl, G. E.	AFML/LP
Lee, T.	AFML/LPH
Lehn, W. L	AFML/MBE
Leinberger, K.	UDRI
Lituak, S	AFML/LTN
Lopez, A.	AFML/LTN
Loughran, G. A	AFML/MBP
Lyon, S. P	AFML/LL
McDevitt, N. T	AFML/LP
Marcus, H.	AFML/LPT
Material Science Corp.	Material Science Corp.

REQUESTER	ORGANIZATION
May, D. R	AFML/LN
Metzger, G. E.	AFML/LL
Meulamans, J I.	AFML/LTE
Meyer, F H.	AFML/MXA
Mildrum, H.	UDRI
Morris, G J	AFML/LN
Morrissey, E J.	AFML/MXE
Neff, R. M.	AFML/LC
O'Hara, W.	AFML/LTP
Olson, J C	AFML/LPE
Opt, P C.	AFML/LN
Parrish, P.	AFML/LL
Peters, L. J	AFML/LTM
Pierce, B. J.	AFML/LPE
Pierce, C. M.	AFML/LL
Powell, W R.	AFML/LP
Poynter, J. W.	AFML/LLS
Pratt, C. A.	AFML/MXS
Ramke, W. G.	AFML/MB
Ray, J. D	AFML/MBC
Reimann, W.	AFML/LL
Reinhart, T.	AFML/MBC
Rhodehamel	AFML/MXE
Rice, D. A.	AFML/LL
Rolinski, E. J.	AFML/LP
Rondeau, R. E.	AFML/LPH
R [illegible]do, [illegible]	AFML/LPH
Ro[illegible]g, Harold	AFML/MBP
Rosenberg, Herbert	AFML/LPH
Ross, J. H.	AFML/LN
Rowand, P. R.	AFML/LL

REQUESTER	ORGANIZATION
Rubey, W	UDRI
Ruh, R	AFML/LL
Russo, T. J	AFML/LC
Rutner, E	AFML/LP
Ryan, M. T	AFML/MBP
Sajadai, R. J	AFML/DOP
Schmidt, D	AFML/MBC
Schmitt, G. F	AFML/MBE
Schulman, S	AFML/LN
Schwartz, H. S.	AFML/MB
Schwenker, H	AFML/LN
Shillito, K. P	AFML/LLP
Shimmin, K. D	AFML/LL
Shinn, D. A	AFML/MXA
Simpson, R. P	AFML/LLP
Smith, C. F	AFML/MBP
Smyth, R. R.	AFML/LLD
Snyder, C. E.	AFML/LNL
Srp, C. O	AFML/LL
Standage, A	UDRI
Stanton, R. M.	AFML/LN
Starks, D	AFML/LTN
Stevison, D. F.	AFML/LP
Strang, J. R.	AFML/LN
Sullivan, J. J.	AFML/MXE
Tamborski, C	AFML/MBP
Tanis, C	AFML/LTN
Tanner, H. A.	AFML/LPE
Tarrants, E. H	AFML/LTE
Tesson, J. T.	AFML/MXS
Tolley, L. G	AFML/LPH

REQUESTER	ORGANIZATION
Pringle, H K	AFML/LTE
Tsai, S W	AFML/CA
Vahldiek, F	AFML/MXS
Voss, D. K.	AFML/MBC
Voss, D P	AFML/LL
Wheeler, E.	AFML/LTN
Wheeler, W. H.	AFML/MXS
Williamson, J R.	AFML/LTN
Winn, R. A	AFML/LP
Wittebort, J I.	AFML/LTE
Wittman, R. E.	AFML/MXE
Zimmerman, P.	AFML/MBE

APPENDIX C

SDI PROFILE TOPICS

SDI SEARCH REQUESTS PROCESSED
1 DECEMBER 1971 - 30 NOVEMBER 1972

SERIAL NO.	SEARCH TITLE
99007	Organic Fluorine Compound
99008	Ferrocene, Compounds, Metallocene Polymers
99009	Spiropolymers, Spirocompounds
99022	Damping, Flotation Fluids
9903[illegible]	Properties of High Temperature Polymer Composites
99033	Testing of Polymer Composites
99034	Process of Polymer Composites
99037	Transparent Films for Windows
99041	Cleaning of Aircraft
99046	Transparent Materials
99048	Radar Absorbing Materials
99070	Aircraft Armor Materials Impact
99071	Carbon Fiber Research/Technology
99072	Three Dimensional Fibers
99082	Environmental Effort on Fibrous Materials
99083	Fabric Properties
99084	Flammability of Materials Fabrics
99085	Recovery, Safety of Personel
99086	Parachute System-Loading
99087	Expandable Structures
99088	Coated Fabrics
99089	Properties, High Strain Rate-Fibers
99094	Fiber Optics
99095	Electrically Conductive Fibers
99099	Compressor Blades for Aircraft Engines
99112	Forming of Metals
99117	Powder Metallurgy Techniques

SEARCH NO.	SEARCH TITLE
99128	Paints, Primers, Surface Finish
99134	Polymer Composite Tankage
99135	Batteries - Materials
99151	Electrical Powder Devices, Electro - Chemical
99152	Radomes High Temperature Dielectrics
99167	Fluoro Organic Compounds
99168	Fluorinated Polymers
99169	Melting of Metals & Alloys
99175	Mathematical Analysis of Metal Working
99177	Temperature Measuring Instrumentation
99179	Polyacrylonitrile - Decomposition Production
99180	Rain Dust Erosion Phenomena
99182	High Temperature Ceramics
99183	Ceramic, Metal Composites
99184	Orthopedic Implant Materials
99197	Crystalline Carbon Fibers, Thermal Analysis
99198	Rare Earth Alloys Crystal Structure
99199	Rare Earth Co Magnetic Materials
99200	Holography Crystal Deformation
99201	Ceramic Coatings, Flame Spraying
99202	Mechanical Properties of MgO Glasses
99203	Design of Instrumentation
99218	Gas Chromatography Decomposition of Polymers
99233	Ceramic Substrates Packaging for Magnetic Devices
99237	Energy Conservation Materials
99238	Masers and Lasers
99239	Luminesence, Optical Property Special Materials
99243	Metal Processing
99254	Molecular Vibration Spectra of Materials
99255	Instrumentation for IR Spectra

SEARCH NO.	SEARCH TITLE
99257	Nonmetallic Radomes Fabrication
99259	Decomposition of Polymers
99287	Joining, Welding of Metals
99288	Metal Composites
99289	Carbides, Cermets Phase Diagram
99299	Fibrous Materials for Clothing
99301	Processing of Ablative Composites
99302	Ablation, Phenomena Mechanism
99303	Adhesives - Properties and Interfacial Phenomena
99305	Effect of Electrical Field on Interfaces
00306	Composites
99317	Properties of Aerospace Materials
99318	Synthesis of Perfluoro Compounds
99319	Ozone Chemistry
99320	SeO_2 Oxidation of Perfluoro Materials
99321	Cvclic Organic Peroxides
99324	Coating Wear and Erosion
99329	Superconductivity
99342	Metallic Composites
99345	Solid State Electronic Materials
99346	Physical Chemistry
99347	Ceramics
99350	Laser Radiation on Materials
99351	Viscoelasticity and Fracture
99354	IR Scanning Devices
99355	Semiconducting Glasses
99357	Differential Thermal Analysis
99360	Electron Microscopy
9936[illegible]	Laser and IR Windows
99363	Cadmium Telluride and Zinc Selenide
99365	Microstructure, Mechanical Properties, Working

SEARCH NO.	SEARCH TITLE
99366	Heat Treatment of Titanium
99367	Powder Metallurgy
99368	Qualitative Microscopy
99374	Mechanics of Metal Composites
99379	Synthesis of Hydraulic Fluids
99382	Functional Laser Trimming
99383	UHF Broadband Amplifiers
99393	Polymeric Protective Coatings
99394	Erosion
99395	Directionally Solidified Eutectics or Composites
99398	Aluminum - Chromium Binary
99401	Liquid Lubricants
99402	Fatigue, Mechanical Properties of Aluminum and Steels
99405	Solid Lubricants Compacts
99406	Rhenium Ductilizing of Tungsten
99407	Solid Solution Softening BCC Metals
99408	Ablation and Ablative Material
99420	Laser Window Materials
99424	Direct Solidified Eutectics
99425	Transformations in Ti Ni Co Nb
99427	Corrosion Data
99428	Electrodeposition
99429	Storage Material Capabilities
99432	Analysis of Polymers
99435	Magnetic Materials and Properties
99440	Powder Metallurgy Technology
99441	Glass Fabrication
99442	Metal Polymer Interfaces
99443	Polymer Composite Interfaces
99444	Polyphenolquinoxaline Resins
99445	Polymer Degradation

SEARCH NO.	SEARCH TITLE
99446	Working, Alloy Development of Ti, Al Alloys
99447	Structure of Perfluoro Organization of Fluoro Metallic Compounds
99448	Corrosion and Embrittlement of 4340 and D6aC
99449	State of the Art of Epoxy Polymers
99450	Chemical Types, Curing of Epoxies
99452	Magnetic Resonance
99453	Acoustical and Optical Radiation
99454	Compatibility of Metals
99455	Structural Application of Metals, Composite
99456	Fabrication of Metals Composites
99457	In-Service Corrosion Failure
99458	Lubricants for Aerospace Systems
99459	Thermoplastics
99461	Advanced Composite Application
99462	Transparent Materials
99463	Lasers - Materials and Effects
99464	CO_2 Lasers
99465	RAE Magnetic Materials
99466	Elastomers Sealants, Polymers
99467	II-VI Semiconductors
99468	Laser Damage on Materials
99483	Tooling for Composites
99484	Organic Compounds-Nuclear Magnetic Resonance
99485	Liquid Crystals
99486	Laser Effect on Materials
99487	Laser Effect on Materials
99506	Al Composites with Boron Fibers
9950[illegible]	Fracture Theory of Metals
99519	Properties of Rigid Polymers
99520	Metal Matrix Composites

SEARCH NO.	SEARCH TITLE
99521	Phase Transformation of Defects
99522	Joining Oxides of Metals Alloy Development
99523	Hydraulic Fluids and Lubricants
99524	Polymeric Protective Coating
99525	Rain and Dust Cloud Simulation
99526	Thermal Protection Systems
99527	Inorganic Nonmetallic Reinforced Fibers
99528	Polymeric High Strength Fibers
99529	Stress Corrosion, Cracking
99530	Shock Phenomena
99531	Cutting Tools Ti and Superalloys
99533	Fluids and Lubricants
99535	Oxidation and Coating of Metals
99536	Mechanical Properties, Testing
99537	Shells, Panels - Structural
99538	Acoustical Effect on Materials
99539	Temperature Effects on Microstructure
99540	Dynamic Loading Behavior of Materials
99541	Bearing Systems for Space
99542	Testing Lubricant - Bearing System
99543	Physical Metallurgy
99544	Ladder, Spiro, Thermal Stability Polymer
99545	Nonflammable Fibrous Materials
99546	High Strength Polymer Fibers
99547	Composite Reinforcements
99548	Heat Flow in Fibrous Materials
99549	Liquid Fuel Fires
99550	Laser Hardened Materials
99551	Math, Statistic, Prediction of Behavior
99552	High Temperature Plastic Coatings

SEARCH NO.	SEARCH TITLE
99553	Composites Data
99554	Thermal Protection Systems for Rockets
99555	Laser Window Materials
99556	Crack Initiation at Notch
99557	Joining Welding, Brazing
99558	IR Laser Window Materials
99559	Metal Composites Height and Weight
99560	Emission From Material Under Stress
99561	Elastomers and Applications
99562	Reinforced Polymer Composites
99563	Stress Corrosion Kinetics
99564	Mass Spectrometry
99565	Radar IR, UV Absorption Materials
99566	Optical Properties of Inorganic Materials
99567	Structural Adhesives
99568	Surface Analysis
99569	Aerothermodynamics
99570	Reaction Kinetics
99571	Thermodynamics
99572	Chemical Physical Behavior in Ablative Wakes
99573	Wear Properties of Titanium
99574	Fracture Mechanics
99575	Mechanical Fasteners
99576	Titanium Alloy Properties
99577	Vacuum Deposition Techniques
99578	Polymeric Dielectric Coating
99579	Measurement of Optical Properties
99580	Energy Effect on Materials
99581	High Temperature Corrosion Protective Coating
99582	Paint, Coating Formulation Camouflage

SEARCH NO.	SEARCH TITLE
99583	Properties of Ni Superalloys
99584	Infrared Detectors, Photoconductivity
99585	Amplification of Surface Acoustics
99586	Elastomeric Fluid Seals
99587	Rolling Technology of Metals
99588	Manufacturing Technology Steel Ti Al Mg Be
99589	Nondestructive Testing, Quality Control
99590	Design Behavior of New Composites
99591	Effect of Laser Radiation on Materials
99592	Welding of Titanium Alloys
99593	Oxidation of Ni Superalloys
99594	IR detectors - amplifiers
99595	Ceramics - Properties and Application
99596	Lubricant Composites with Titanium
99597	Thin Metal Foils - Preparation
99598	Powder Alloys Ti Al Co Ni
99599	Temperature Control Coatings
99600	Rain Resistance
99601	Carbon Fibers - Pyrolysis of Organic Fibers
99602	Weldability of Titanium Alloys
99603	E M Windows IR Laser Radiation
99604	Fuel Tank Sealants
99605	Ferroelectric Materials
99606	High Temperature Application of Materials
99607	Coating Processes
99608	Theory of Metal Plastic Deformation
99609	Microwave Ferrites
99610	Properties of Textiles
99611	Optical Contamination of Spacecraft Surfaces
99612	Ultrasonic Testing
99613	Fracture Mechanics

SEARCH NO.	SEARCH TITLE
99614	Semiconductor Materials; Properties
99615	Semiconductor Compounds
99616	Garnets Ferrites and Computers
99617	Fabrication Process - Electronic
99618	Electro-Optical Materials
99619	Thermionic Tubes - Materials and Processes
99620	Dielectronic for Electronic Devices
99621	Epoxy Graphite Composites
99622	Coating Vs Erosion
99623	Forming Techniques
99624	Properties of Composites for Missiles

APPENDIX D

DEFINITION OF SUBJECT CATEGORIES

APPENDIX D

DEFINITION OF SUBJECT CATEGORIES

AMIC	COSATI	CATEGORY
01	01	Aeronautics
		Aerodynamics
		Aeronautics
		Aircraft
		Aircraft flight control and instrumentation
		Jet engines
02	03+04	Astronomy, Astrophysics, Atmospheric Sciences
		Astronomy
		Astrophysics
		Atmospheric physics
		Meteorology
03	06+07	Chemistry, Biology, Medical Sciences
		Biochemistry
		Bioengineering
		Biology
		Chemical analysis
		Chemical engineering
		Inorganic chemistry
		Life support systems
		Organic chemistry
		Physical chemistry
		Radiochemistry
		Toxicology

AMIC	COSATI	CATEGORY
04	09	Electronics and Electrical Engineering
		Components
		Electronic and electrical engineering
		Telemetry
05	11A	Adhesives
		Ceramic cements
		Organic resin adhesives
		Potting compounds
06	11A	Seals, Sealants
		Ceramic-metal bonds
		Mechanical seals
		O-rings
07	11B	Ceramics, Refractories, Glasses, Minerals
		Borides
		Carbides
		Carbon, graphites
		Mixed oxides
		Nitrides
		Single oxides
08	11C	Coating, Paints, Oxide Films
09	11D	Composites Materials, Laminates, Sandwich Structures, Honeycomb
10	11E	Fibers, Textiles, Cloth
11	11F	Metallurgy, Metallography
		Alloys
		Metals
12	11H	Oils, Lubricants, Heat Transfer Fluids, Greases, Hydraulic Fluids
13	11I	Polymers, Plastics
14	11J	Elastomers
15	11K	Cleaning Compounds, Surface Active Agents

AMIC	COSATI	CATEGORY
16	11L	Wood and Paper Products
17	21	Fuels, Propellants, Propulsion Systems, Explosives
18	13	Mechanical, Industrial, Civil and Marine Engineering
		Civil engineering
		Construction equipment, materials, supplies
		Containers and packaging
		Couplings, fittings, fasteners, joints
		Industrial processes
		Machining, tools, machine elements such as bearings, gas lubrication systems
		Marine engineering
		Pumps, filters, pipes, fittings, tubing, and valves
		Safety engineering
		Structural engineering
19	14	Methods and Equipment
		Apparatus
		Detectors
		Laboratories, test facilities, and test equipment
		Recording devices
20	18	Nuclear Science and Technology
		Fuel elements; fuel, nuclear
		Nuclear explosions
		Nuclear power plants
		Nuclear reactors
		Radiation shielding
		Radioactive wastes

AMIC	COSATI	CATEGORY
21	20	Physics
		Acoustics
		Crystallography
		Electricity and magnetism
		Fluid mechanics
		Masers and lasers
		Optics
		Particle accelerators
		Particle physics
		Plasma physics
		Quantum theory
		Solid mechanics
		Solid-state physics
		Spectrometry, spectroscopy
		Thermodynamics
		Wave propagation
22	10, 16, 22	Space Technology and missiles
		Astronautics
		Energy conversion, solar cells
		Launch vehicles
		Missile technology
		Re-entry vehicles
		Rockets
		Satellites, artificial
		Spacecraft
		Trajectories and re-entry

APPENDIX E

RETROSPECTIVE SEARCH REQUESTS

TABLE E-1

DISTRIBUTION OF INPUT DOCUMENTS BY SUBJECT CATEGORY

AMIC Category	Documents	
	No.	%
01	112	1.7
02	71	1.1
03	1341	20.1
04	6	1.0
05	30	0.4
06	40	0.6
07	163	2.4
08	125	1.9
09	220	3.3
10	42	0.6
11	1089	16.3
12	192	2.9
13	136	2.0
14	46	0.7
15	10	0.1
16	20	0.3
17	126	1.9
18	324	4.9
19	330	4.9
20	384	5.8
21	1581	23.7
22	229	3.4

RETROSPECTIVE SEARCH REQUESTS PROCESSED
1 DECEMBER 1971 - 30 NOVEMBER 1972

SEARCH NO.	SEARCH TITLE
2655	Convection Transfer
2656	Critical Strain Grain Growth
2657	Rare Earth Ni-Co-Fe Alloys
2658	Wear Fretting of Titanium
2659	Flexural Testing of Laminates
2660	Rare Earth Co-Ni-Fe Alloys
2661	Rare Earth Co-Fe-Ni Alloys
2662	Thickners-Liquid Hydrocarbon Fuels
2663	Cost of Titanium
2664	Documents on Titanium Ni-Alloys Eutectoid
2665	High Temperature Nickel Superalloys
2666	Iodoform Heat of Vaporization
2667	Inorganic Polysulfides
2668	Polysulfide Preparation
2669	Alloy 713-C Mechanical Properties
2670	Inconel 71 ·C High Temperature Fatigue
2671	Thermochromic Compounds (25-50°C)
2672	Testing Adhesives Bonding Joints
2673	Silver Bearing Corrosion Inhibition
2674	Surface Finish of Aluminum
2675	H_2 Effect on Titanium-Al-Sn-V
2676	Hydrogen Pickup of Titanium Alloys
2677	Weldbonding Aerospace Structures
2678	Mechanical Fasteners for Aircraft
2679	Fretting Corrosion
2680	Fatigue of Incoloy 901

SEARCH NO.	SEARCH TITLE
2681	Hot Deformation of Alloys
2682	Ballistic Impact Testing
2683	Impact Testing of Composites
2684	Composites Research and Development, AFML
2685	Graphite Fiber Reinforced Composites
2686	Mesophase From Pyrolysis
2687	Stress Corrosion Cracking of Titanium Alloys
2688	Powder Metallurgy of Titanium and Titanium Alloys
2689	Light Laser Beam Choppers
2690	Graphite Fiber Processing
2691	Hot Deformation on Alloys
2692	Textured Titanium
2693	Phosphine Oxide Polymers
2694	Phosphorus Containing Polymers
2695	Ultrasonic/Defect Interactions
2696	Adhesive Bonding Beta III
2697	UV Stabilizers for Polymers
2598	Low Outgassing Polymers
2699	Hydroforming
2700	Ultrasound Attenuation
2701	Tooling for Composites
2702	Metal-Polymer Interfaces
2703	Polymer-Composite Interfaces
2704	Polyphenolquinoxaline Resins
2705	Polymer Degradation
2706	Corrosion and Embrittlement of 4340 and D6
2707	Structure of Perfluoro Organofluoro Metallic Compounds
2708	State of the Art of Epoxies
2709	Chemical Types - Curing of Epoxies

SEARCH NO.	SEARCH TITLE
2710	Working, Alloy Development of Al-Ti Alloys
2711	Perfluoro Aldehydes or Ketenes
2712	Reflective Coatings for Air Materials
2713	CO_2 Laser Photometry
2714	Titanium Fires in Aircraft
2715	Rare Earth Magnetic Materials
2716	Prestress Effect on Fatigue Life
2717	Elastomers, Sealants, Polymers
2718	II - VI Semiconductor
2719	Organic Fluoride Compounds
2720	Organic Compound NMR
2721	Liquid Crystal Display Devices
2722	Zinc Selenide - Heat Conductivity
2723	Al Composites with Boron Fibers
2724	Fracture Theory of Metals
2725	Instability Fractures
2726	Nitroso Elastomers
2727	Fasteners for Composites
2728	Thermal Shock Behavior
2729	Oxidation, Sulfidation of Metals
2730	Polyphenylquinoxalines
2731	Corrosion Inhibitors
2732	Laser Hardened Materials
2733	Electric Field Controlled Heat Transfer
2734	Oxidation of Niobium
2735	Reinforced Thermoplastics
2736	Service Life of Ni Superalloys
2738	Stainless Steel 13-8
2739	Properties Behavior Ni Superalloys

SEARCH NO.	SEARCH TITLE
2740	Hydrogen Embrittlement, Ferrous Alloys
2741	Deformation Mechanics of Ni Superalloys
2742	Microstructure of Ni Superalloys
2743	Titanium-Aluminum Alloys
2744	Historical Analysis Materials Development
2745	Decomposition of Polyurethanes
2746	II - IV Compounds Properties
2747	E-M Generation Acoustic Waves
2748	Structural Adhesives
2749	Metal Surface Preparation
2750	Epoxy and Polymide Resins
2751	Trans- -rent Materials
2752	Carbon Foams
2753	Chemical Analysis of Fuel Tank Residue
2754	Rigid Mullite

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) University of Dayton Research Institute 300 College Park Ave. Dayton, Ohio 45469	2a REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b GROUP
3 REPORT TITLE The Use of Selected Portions of Technical Documents as Sources of Index Terms and Effect on Input Costs and Retrieval Effectiveness	
4 DESCRIPTIVE NOTES (Type of report and inclusive dates)	
5 AUTHOR(S) (First name, middle initial, last name) H. H. Schumacher, J. F. March, F. L. Scheffler	
6 REPORT DATE April 1973	7a TOTAL NO. OF PAGES 78 / 7b NO. OF REFS 14
8a CONTRACT OR GRANT NO F33615-71-C-1069 b PROJECT NO 7381 c Task No. 738103 d	9a ORIGINATOR'S REPORT NUMBER(S) UDRI-TR-73-10 9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFML-TR-73-53
10 DISTRIBUTION STATEMENT Approved for public release; distribution is unlimited	
11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY Air Force Materials Laboratory Wright Patterson AFB, Ohio 45433

13 ABSTRACT Recall (the retrieval of all available relevant documents) should decrease with the quantity of text serving as a source of indexing. However, the time for indexing and therefore the input cost should be less, establishing a tradeoff between input cost and retrieval effectiveness. To quantify the effect of restricting the source text on both retrieval effectiveness and input cost, an experiment was designed in which the full technical document text was divided into five categories: 1, title; 2, abstract; 3, table of contents and lists of figures and tables; 4, author-assigned keywords; and 5, the body. An experimental data base was prepared whereby the index term source category and the indexing time were recorded. Sets of SDI and retrospective searches were run against the data base, and retrievals were analyzed by category in terms of retrieval response, S; relevant document response, R, categorical relevance, $\overline{R}$; indexing time, T; and retrieval efficiency, E and $\overline{E}$. For the subset of documents retrieved, 81% of the available relevant documents were retrieved from Categories 1-4; the indexing time required for these four categories was only 53% of the total indexing time. For the entire set of documents input into the experimental data base, the portion of indexing time for the first four categories was 60%. Based on these results, it was decided that the body of the document could be excluded as a source of index terms. This decision was translated into a reduction of unit cost from $10 to $8.25.

DD FORM 1 NOV 65 1473

UNCLASSIFIED

Security Classification

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Aerospace Materials Information C nter						
Indexing						
Cost Effectiveness						
Document Retrieval						
Information retrieval						
Subject indexing						
Information retrieval effectiveness						
Recall						
Sources of index terms						
Depth of indexing						
Specificity of indexing						
Information Systems						

☆U.S.Government Printing Office: 1973 — 759-493/522

Security Classification