(4) Costello, J. C., Jr., "Storage and Retrieval of Chemical Research and Patent Information by Links and Roles in du Pont," *American Doc.*, **12**, 111 (1961).
(5) Costello, J. C., Jr., "Some Solutions to Operational Problems in Concept Coordination," *American Doc.*, **12**, 191 (1961).
(6) Morse, R., "Information Retrieval," *Chem. Eng. Prog.*, **57**, 55 (1961).
(7) Holm, B. E., "Information Retrieval—A Solution," *Chem. Eng. Prog.*, **57**, 73 (1961).
(8) Vickery, B. C., "Thesaurus—A New Word in Documentation," *J. Doc.*, **16**, 181 (1960).
(9) Wall, E., "A Practical System for Documenting Building Research," in "Documentation of Building Science Literature," publn. 791 (1960), p. 15; U. S. Senate Report, 86th Congress, 2nd Session, Document No. 113, 1960.
(10) Taube, M., private communication. Also "Final Report: Feasibility Study for Coordinate Indexing and Machine Retrieval of Chemical Warfare Laboratory Library Documents." Prepared under Contract Number EA18-108-405-CML-676 for the Chemical Research and Development Laboratory, U. S. Chemical Corps, Ded. 1960, and "Handbook of Procedures and Comprehensive Report" prepared in conjunction with above feasibility study.
(11) Leibowitz, J., Frome, J., and Andrews, D. D., "Variable Scope Patent Searching by an inverted File Technique," U. S. Patent Office, Research and Development Report Number 14, Nov. 17, 1958.
(12) Levery, F., Reibell, B., Cohendet, R., de Picciotto, S., Biet, F., Picot, G., Deribere-Desgardes, M. L., Slove, D., Escaravage, R., and Fournier, M., "Une Experience de Selection Automatique de Documentation," published by IBM France and the Compagnie de Saint-Gobain, Dec., 1960.
(13) Montague, Barbara A., "Patent Indexing in du Pont by Concept Coordination Using Links and Roles," *American Doc.*, **13**, 104 (1962).
(14) Yngve, V. H., "The Feasibility of Machine Searching of English Texts," in "Proceedings of the International Conference on Scientific Information," Washington, D. C., Nov. 16–21, 1958, Vol. 2, p. 975.

Five Operational Years of Inverted Index Manipulation and Abstract Retrieval by an Electronic Computer*

B. K. DENNIS, J. J. BRADY, and J. A. DOVEL, Jr.
Technical Information Center, Flight Propulsion Division,
General Electric Company, Cincinnati 15, Ohio
Received May 9, 1962

In this presentation we have attempted to record the salient facts about the conception, development, operation and continued improvement of our Automatic Information Retrieval System (AIRS). We have had a great many experiences in our information retrieval work, both good and bad. The path we have followed over the past five years has been rocky. We have stumbled and fallen flat on our faces more times than we care to recall, but we are still moving forward. We are happy to note that our progress has accelerated during the past year, showing specific pay-off in terms of a growing body of customers satisfied with value received from our services.

As our discussion proceeds, you should bear in mind that the General Electric Automatic Information Retrieval System has not been developed and operated as an experiment under research laboratory or "hothouse" conditions. We have not had a staff of information system theoreticians and specialists along the way to tend it, nurture it and keep it progressing toward carefully established and well defined experimental objectives with adequate financing to reach those objectives. To the contrary, our information retrieval system was developed in response to a felt but not clearly defined need. It has been considered a production tool from the outset, available to its users on a specific charge basis. Its design, development and operation has been solely in the hands of information service and computer programming practitioners. Its rate of development has been dictated by a widely fluctuating level of financial support.

Since the unveiling of our retrieval system in 1958, many people have visited or contacted our Center to learn more about our automation work, with many visitors providing much valuable advice. There have been those who have taken exception to our approach. Early in our work with the IBM 704, one outspoken critic accused us of "using a sledge hammer to drive a carpet tack." That comment always came to mind every time we were required to break up a machine search and re-run it because the memory of our 704 could not handle it. There have been times when our searchers felt that they were attempting to drive a railroad spike with a tack hammer and would have been pleased to inform the "sledge hammer" critic accordingly.

During this discussion, we will attempt to provide answers for many of the questions which have been asked by visitors during the past few years. Of necessity, much of what we have to say will be of an intuitive or "educated guess" nature. One of the unfortunate aspects of attempting to advance the state-of-the-art in a service organization is the fact that the more sophisticated scientific or experimental objectives may often be compromised by the immediate necessity to provide a service. Therefore, much experience data which would be

*Presented before the Division of Chemical Literature, ACS National Meeting, Washington, D. C., March 23, 1962.

interesting and useful in this discussion is not currently available.

BACKGROUND

Early in the life of the Technical Information Center, an information retrieval method was sought which economically would provide a significantly greater depth of penetration into the documents described and with a higher degree of resolution than could be afforded by conventional means at comparable costs of input, index maintenance, and search. After investigating several available systems, the Center decided to develop a uniterm coordinate index for access into its documents other than books. Books would still be catalogued until such time that it would prove desirable to index them also. To launch its uniterm system on a sound basis, the Center retained the services of Dr. Mortimer Taube and associates in October, 1953, under whose direction a uniterm coordinate index was established and Center personnel were briefed in its care and feeding. Four years later, this index served as the base from which the Center developed its Automatic Information Retrieval System. In those early days, it was believed by the Center that the uniterm approach to information retrieval permitted lower cost document analysis through the use of non-technical, untrained personnel. Indexing is now performed by technical people. It is our opinion that careful selection of really significant terms in the analysis process is prerequisite to a reasonable degree of retrieval selectivity and completeness.

During the first fews years of uniterm indexing, the Center depended upon title, abstract, and table of contents to provide all terms used for that document. An estimated 8–10 terms per document were assigned on this basis. (Figure 1 shows trend in indexing depth. Data in this chart are derived from a very limited sampling and shows only trend.)

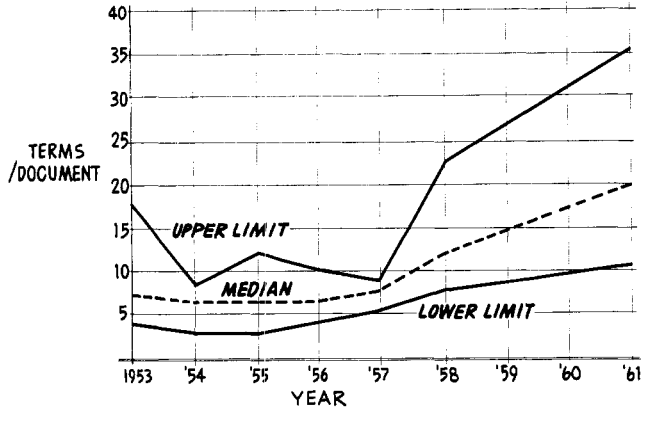


Fig. 1.—Indexing depth trend.

By the end of the first full year of operation approximately 5000 items had been indexed into the Center's uniterm system. Rate of input for subsequent years is shown in Figure 2. Input included technical reports and memoranda generated internally and also those obtained from external sources (particularly Defense Agencies and their contractors in the aerospace industry). Technical society papers, journal and trade press articles, U. S. and foreign patents, translations, and many miscellaneous media of scientific and technical information were also included in the file. Cumulative growth of the file is seen in Figure 3.

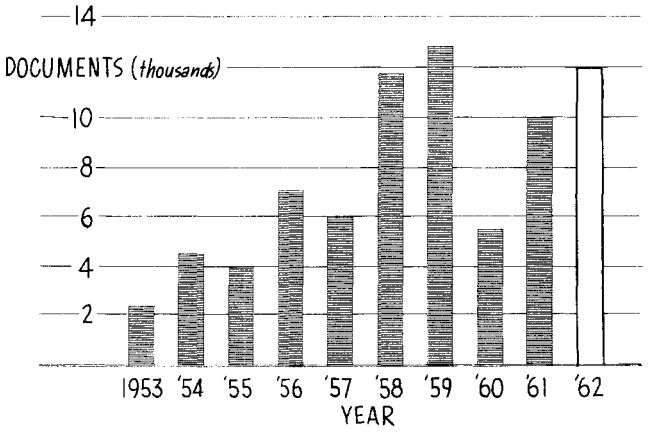


Fig. 2.—Rate of indexing.

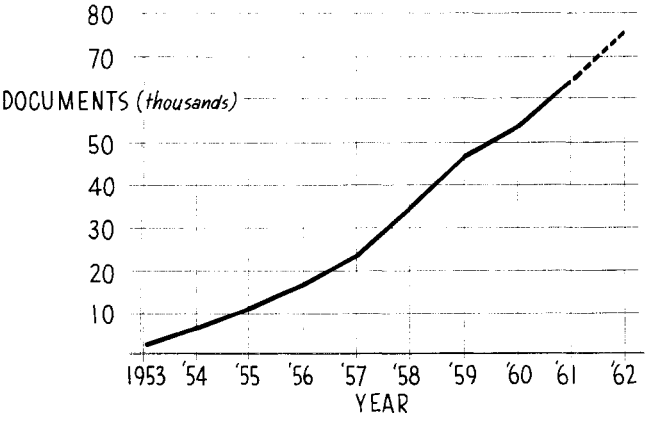


Fig. 3.—Growth of AIRS file.

Utility of the Center's manual system had diminished by 1957 until it was being used only as a matter of last resort. Some terms such as COMPRESSOR, ALLOY, ENGINE and DESIGN each had been posted over 2,000 times, filling several cards with their accession numbers. And, unfortunately, it was too often necessary to coordinate one high density with others nearly as large so that even the limiting effect of lightly posted terms was of little help. It became increasingly apparent that the Center's users, who were required to conduct their own searches, were making so little use of the manual index (less than one search per week) that the Center had lost a great deal of its effectiveness. Mainly requests for specific documents were handled and little bibliographic assistance was provided.

We became convinced that the more than 1,000 technical people who made frequent use of our Center's services and facilities and the additional 1,500 to 2,000 technical and management personnel who were either infrequent

or potential users could benefit significantly by a fast reacting, technically qualified information searching service and that an adequate service could be provided only by people skilled in the subject matter of their clients. We recognized that our manual uniterm coordinate index was a dull tool for an information specialist to use, regardless of how technically competent he might be. At the same time, however, we believed the coordinate index method to be a sound approach to information retrieval. Therefore, we decided to investigate the feasibility of mechanizing our existing manual uniterm coordinate index.

II. EQUIPMENT SELECTION

Many people have asked why we used a huge computer like the IBM 704 for information retrieval. Our decision to utilize an IBM 704 (and later an IBM 7090) was based on expediency more than any other single factor. There was an IBM 704 computer system in the same building with us and we had reasonably assured access to it. Besides that, on our Computations staff were two imaginative, young systems analysts who were also good salesmen. These two analysts worked closely with us on our manual system problems and in 1957 we accepted their proposal to mechanize our system on the 704. It should be mentioned that we gave serious consideration to using EAM equipment which was also available in the Plant before we decided in favor of the 704. Operating costs, simultaneous search capability, biblioabstract output, magnetic tape *vs.* card storage and handling, ease of file maintenance and the ease of incorporating new system philosophy were the key factors in making this decision.

Now that a GE 225 computer is available to the Center, we will eventually consider the feasibility of transferring our system to it for operation. We do not feel committed to any particular equipment but will seek to make use of available machines which we believe best fit our requirements.

III. GETTING AUTOMATED*

In 1958, we programmed the computer, performed a minor amount of vocabulary clean-up in our inverted index, unbound terms and fragmented others, keypunched our terms with their accession numbers and abstracts with source and security data, began checking out the system in July and initiated production searching in September.

As originally conceived, the Automatic Information Retrieval System was programmed for the logical product operation with the capability to perform 99 multiterm products simultaneously. No provision was made for either the logical sum or negation. In retrospect, we were overly cautious and tended to restrict our use of the machine. Probably two factors accounted for this. First, we didn't like to stray to far from that which had been proven, that is, our manual index, and secondly, we had visions of creating a system too expensive to operate—a trap easy to step into, by the way, when one is charged $350–400 per hour for the use of a large computer. There was considerable discussion as to whether we should retain our inverted file or convert to a conventional index. Due largely to the increase in computer operating costs which would be incurred by a conventional index because of its longer search file (possibly three times as much tape required as for an inverted index in our case) and considering other factors having to do with index manipulation we elected to continue with an inverted file.

From the searcher's viewpoint the Automatic Information Retrieval System performs two operations. Part I operation is illustrated in Fig. 4. Search questions in the form of index terms arranged in logical products, new terms and their accession numbers, and program instructions, all on punched cards, were stored on an input tape

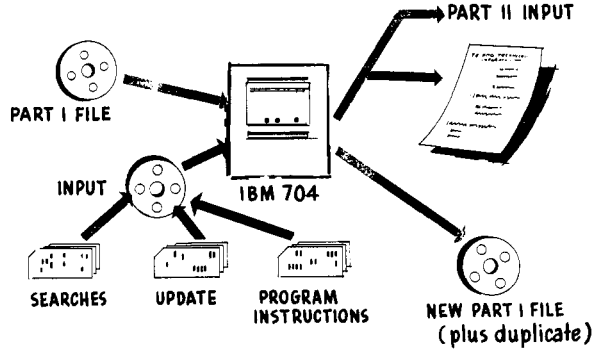
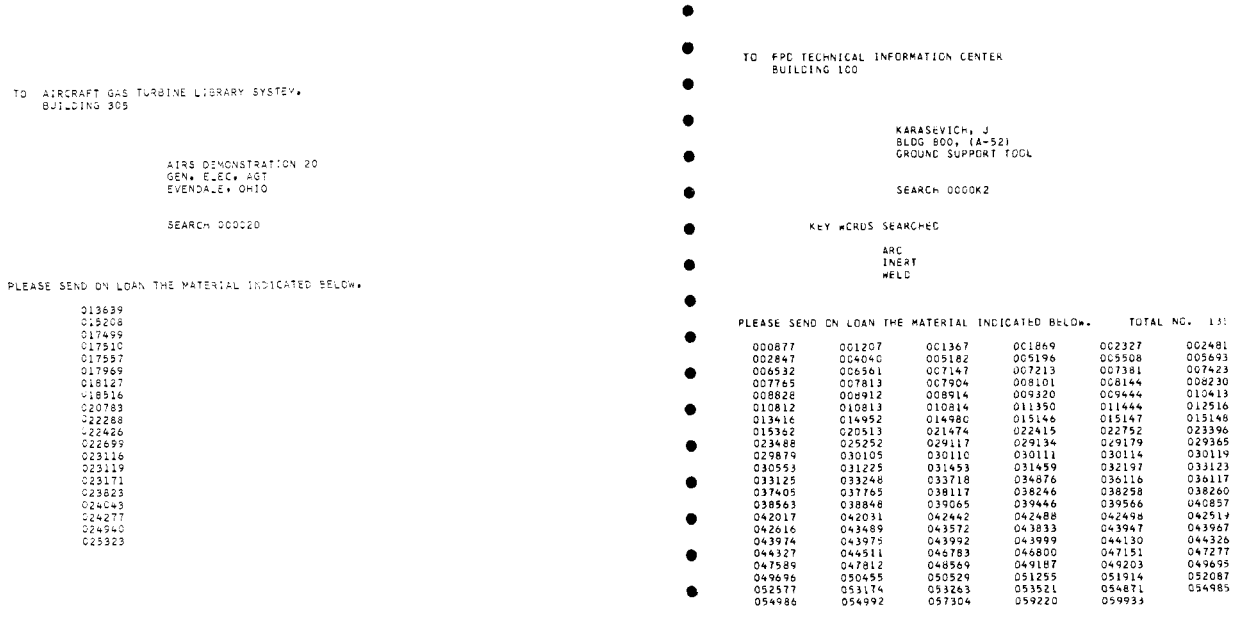


Fig. 4.—How AIRS works: Part I. . . . . Comparison.

on peripheral equipment. The input tape was then run on the 704 along with the Part I file (unifile). Up to 99 multiterm products could be handled on a single search run. Updating the file while searching allowed us to search our total processed information and distribute file updating costs across search customers. This idea was good in theory but was difficult to implement consistently at first. On more than one occasion, several boxes of Unifile input on punched cards would be discovered stacked in the keypunch room, or perhaps beside a programmer's desk or even in the peripheral equipment room long after we had assumed them to be on tape. We finally developed a check-off procedure for machine room operators which cleared up this and several other operational problems of a similar nature.

Output from Part I searching was access numbers and customer identification data and was stored on an output tape. At the searcher's option the search run could be stopped at this point and the output tape printed on off-line equipment. The initial printing format did not include the terms that were coordinated. It was necessary to return to the searcher's work sheet to match question to output. The program soon was modified to print terms on Part I output as seen in Fig. 5.

*Barton, A. R., "Information Retrieval on a High-Speed Computer," G. E. Report No. R59AGT253, paper presented at Western Joint Computer Conference, San Francisco, California, March 3-5, 1959. Dennis, B. K., "High Speed Literature Searching on an IBM 704," paper presented at IBM Information Retrieval Systems Conference, Poughkeepsie, New York, September 21-23, 1959.



TO AIRCRAFT GAS TURBINE LIBRARY SYSTEM
BUILDING 305

AIRS DEMONSTRATION 20
GEN. ELEC. AGT
EVENDALE, OHIO

SEARCH 000020

PLEASE SEND ON LOAN THE MATERIAL INDICATED BELOW.

013639
015208
017499
017510
017557
017969
018127
018516
020783
022288
022426
022699
023116
023119
023171
023823
024043
024277
024940
025323

before

TO FPD TECHNICAL INFORMATION CENTER
BUILDING 100

KARASEVICH, J
BLDG 800, (A-52)
GROUND SUPPORT TOOL

SEARCH 0000K2

KEY WORDS SEARCHED

ARC
INERT
WELD

PLEASE SEND ON LOAN THE MATERIAL INDICATED BELOW. TOTAL NO. 131

000877	001207	001367	001869	002327	002481
002847	004040	005182	005196	005508	005693
006532	006561	007147	007213	007381	007423
007765	007813	007904	008101	008144	008230
008828	008912	008914	009320	009444	010413
010812	010813	010814	011350	011444	012516
013416	014952	014980	015146	015147	015148
015362	020513	021474	022415	022752	023396
023488	025252	029117	029134	029179	029365
029879	030105	030110	030111	030114	030119
030553	031225	031453	031459	032197	033123
033125	033248	033718	034876	036116	036117
037405	037765	038117	038246	038258	038260
038563	038848	039065	039446	039566	040857
042017	042031	042442	042488	042498	042514
042616	043689	043572	043833	043947	043967
043974	043975	043992	043999	044130	044326
044327	044511	046783	046800	047151	047277
047589	047812	048569	049187	049203	049695
049696	050455	050529	051255	051914	052087
052577	053174	053263	053521	054871	054985
054986	054992	057304	059220	059933	

after

Fig. 5.—AIRS Part I Output.

Unless otherwise instructed the search program proceeded automatically into Part II (Fig. 6) and looked up abstracts and bibliographic and security data. Initially in the 704 system, Part II tape updating was accomplished along with searching. However, this practice was discontinued as a matter of economy and abstract tape

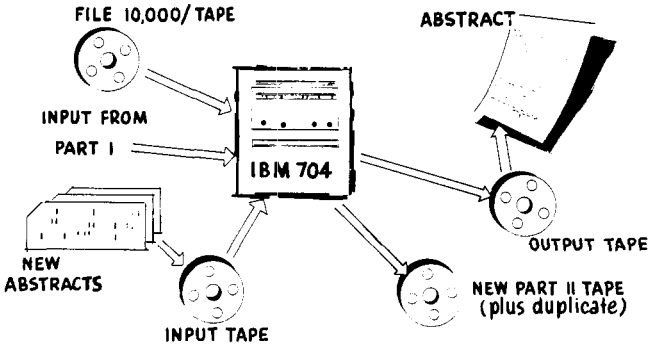


Fig. 6.—How AIRS works: Part II. . . . . Abstract Lookup.

updating was performed monthly just ahead of Current Awareness service runs. An abstract is shown in Fig. 7. The length of our abstracts has been kept at 40–50 words. It may be interesting to consider that this abstract length bears no relationship to the computer nor does it have a basis in communication or statistical theory. It just so happens that an abstract of this length will fit on a 3 × 5 standard library card.

IV. EARLY AIRS EXPERIENCE

Soon after getting into operation with our 704 search system in 1958 it became apparent that the "logical product" did not provide enough flexibility in our search strategy. Because of the relatively pristine nature of our index and due also to much "chaff" introduced into our vocabulary through inadequate quality control when it was transferred from our manual system to magnetic tape, the searchers often found it necessary to use eight or more parallel machine questions to cover adequately a customer's problem. This usually resulted in some abstracts being duplicated several times in a sisngle search. This

KARASEVICH, J
BLDG 800, (A-52)
GROUND SUPPORT TOOL

GROUP 0000K1

CLASSIFICATION
OF REPORT-

ACCESS NUMBER. 053521
REPORT NUMBER. GE DM 61-2
DATE. 01/10/61
AUTHORS. + JONES, E S

TITLE AND ABSTRACT

INVESTIGATION OF THE PHYSICAL METALLURGY OF JOINING TUNGSTEN AND COLUMBIUM. INERT ARC WELDING OF TUNGSTEN SHEET, TRANSVERSE BEND TESTS, METALLOGRAPHIC AND FRACTROGRAPHIC EXAMINATION, TANTALUM FILLER, TUNGSTEN ALLOY WELD FILLER DEVELOPMENT. BRAZING OF COLUMBIUM AND TUNGSTEN, BRAZE ALLOY EVALUATION.

PREPARED BY FPD TECHNICAL INFORMATION CENTER

Fig. 7.—AIRS Part II Output.

wasted machine time to look-up and print-out duplicate abstracts and wasted clerical time to weed them out. Therefore the machine program was modified to perform a "sum" operation, thus eliminating all duplication and providing the searcher more freedom. This "sum" or "composite" feature is illustrated in Fig. 8.

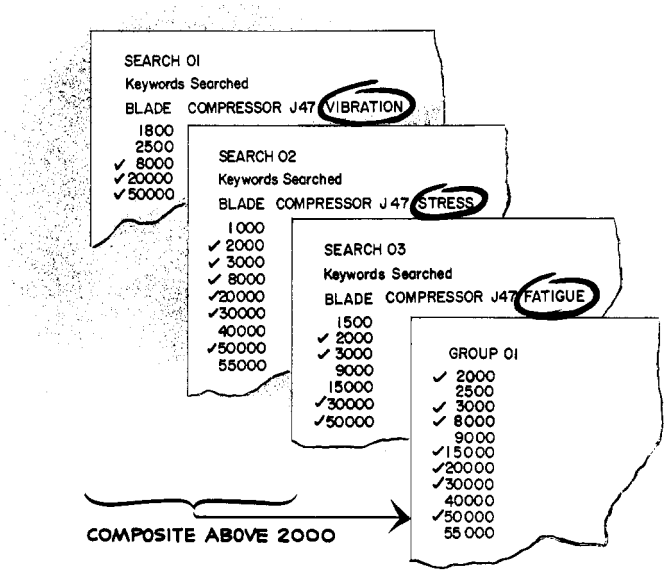


Fig. 8.—AIRS "Sum" or Composite.

After mechanization, vocabulary growth was very slow, because of our attempts to accept only one form of a word in our vocabulary without regard to concept. For example, the keyword included the concepts weld, weldment, weld joints, welder, weldability, and welding. It was our intent to restrict the growth of the machine unifile, thereby controlling running costs, and increasing retrieval power. It seemed to us that even though concepts appeared to be distorted or lost by so doing, they would be restored during the coordination process. For example in the search question, INERT—GAS—WELD—EQUIPMENT, WELD seems reasonably well defined.

Our practice of using single terms and mixing concepts along with our lack of a full logical sum capability led to a serious memory problem that plagued us until we transferred to the 7090. In spite of our 32K memory on the 704, by the time our file had grown to 35,000 items we found ourselves exceeding (or as we referred to it "tilting") the memory area allocated for search term storage with increasing frequency even though we had programmed the machine to make two passes on the Unifile. This meant that the search had to be broken into two smaller runs to get by the "tilt" condition and still search on all questions. Since we were on an overnight schedule, this sometimes meant a one day delay in delivering search results to the customer. Our senior searcher devised a trick to outguess the machine. By knowing the posting density of the terms in his search, the searcher would artifically induce a second pass on the Unifile at an advantageous point in the search run by introducing a "false" question. In so doing he would sequence the coordination of terms to insure a high probability of the run's successful completion.

This is but one of several tricks devised by the searcher to get around the machine limitations in our 704 system. He also developed many short cuts and evasive tactics to compensate for language problems.

The "tilt" problem encountered on the 704 has long since been solved on the 7090 with plenty of room for expansion. To do so we expanded our search question storage area from 9,000 to 12,000 words, changed the search logic to use less density posted terms as filters to the more dense terms in any one product, and used a secondary tape for temporary storage of rejected unifile terms. The secondary tape can be read and re-read as many as five times to complete the coordination of all search questions. Since these modifications to the uniterm search program logic, we have had no "tilt" condition.

Our language problems are now under concentrated attack.

Several features were programmed into the system as we gained operating experience. As a result of these system refinements, the searcher could run Part I only and stop; could run Part I, print it out for a review of his results and then, if satisfied, go on to Part II; could go automatically into Part II without review; could exercise a Part II print-out limit per machine question to avoid excessive machine costs; could exercise an accession number range control for a search run so that he could designate portion of a file to be searched (this was primarily for Current Awareness service); and could elect to avoid "blank sorts" by instructing the machine to ignore terms causing zero output on any machine question.

Several supporting products are generated by the AIRS programs. Those programmed for the 704 included a machine printed alphabetical listing of the system vocabulary showing frequency of use for each term, a machine printed dictionary containing each index term with all of its associated accession numbers and its frequency of use count, and an index term association analysis report which shows for any given term all other terms that have appeared as index terms in documents with it and the number of times the association has occurred. Additional features have been programmed for the 7090. These will be mentioned later in the discussion.

V. OUR DARKEST HOUR.

During the first year or so that we operated our machine system, its initial programmer, who was a member of the Evendale Computations Operation, kept in close contact with it, maintaining and modifying our programs and keeping the system functioning smoothly. Unfortunately for us, however, he was promoted and transferred to another city. After his departure there was no one left in the Computations Group familiar with our programs. Also, in our efforts to operate on a production basis, we had not found time to document them adequately. Consequently, it was virtually impossible for anyone unfamiliar with our system's programs to be of much immediate help as difficulties arose. Our system's reliability began to slip and problems began to appear more frequently.

During 1960 a local business recession resulted in cutting the activities of the Technical Information Center in half. Attempts to improve our retrieval system came to a halt. Input into the system dropped to half its previous rate. To complicate matters, the Plant's Computations Operation became deeply immersed in problems associated with several new complex computer applications that taxed available manpower and made the 704 nearly inaccessible. And then for the coup de grace, several problems arising from the turnover in personnel responsible for the maintenance and operation of our system prevented us from making successful search runs for days at a time. We more than half seriously contemplated how we might "unmechanize." But even though we felt as if we were hanging on the edge of a cliff with someone stepping on our fingers, we survived the onslaught, a little sadder but wiser for all of it.

At the close of 1960, we won adequate financial support to continue our work on a sounder basis. By adding an information systems specialist to our staff we introduced a degree of continuity into our work not available before. And with sufficient funds to purchase full time system analysis and programming from our Computations Operation we were able to move forward again.

VI. INFORMATION RETRIEVAL ON THE IBM 7090

At the beginning of 1961 the Evendale 704 was replaced by an IBM 7090–1401 computer system. To get into operation with our Automatic Information Retrieval System on the new computer a straight transfer was made at once. However, the difficulties that had been accumulating in our 704 programs were carried along to the 7090. One of these, the "tilt" problem and its solution, was mentioned earlier. Because of other difficulties, parts of our abstracts on magnetic tape became scrambled or lost while another problem caused temporary "loss" of terms and access numbers on our Part I tape.

In addition, major modifications were made to our retrieval system programs to exploit the greater capabilities of the 7090. The tabulation summarizes the improvements that have been made to our AIRS programs to solve problems, increase search flexibility and reliability, reduced operating costs, and accommodate the expanding parameters of our growing file:

AIRS PROGRAM DEVELOPMENT

	Logic	704	7090
1.	Output options	Part I separate; Part I and Part II combined	Part I separate; Part II separate; Part I and Part II combined
2.	Re-starting	Beginning Part I	Beginning Part I or beginning Part II
		Part I Improvements	
3.	Searches/run	99	1300 (approx.)
4.	Result buffer (BUCK)	9000	12000
5.	Search term storage order	Same as on Unifile	Smaller terms first
6.	Possible file passes/run	2	5 (on secondary buffer tape)
7.	Composite (SUM)	12 products per composite maximum	Unlimited
8.	Limit controls	(a) Limit access number output per product.	(a) Same (b) Specify access number for search run. (c) Specify access number band width for each composite group.
9.	Error controls	(a) Search stopped if results exceed buffer capacity (9000).	(a) Search stopped if results exceed buffer capacity (12000). (b) Misspelled search term halts search after Part I.
		Part II Improvements	
10.	Limit controls	(a) Eliminates access numbers for tapes not mounted.	(a) Same (b) Can eliminate any selected group of access numbers.
11.	Search sequence	Read abstract tapes serially.	Searches two abstract tapes at a time. Cycles to next tape when a tape is completed.
12.	Abstracts printing	Abstracts written on output tape as they were retrieved.	Prints through a scheduler—i.e., retrieved abstracts printed as output channel becomes free.
13.	Input-Output logic	Serial non-buffered	Reads and writes are buffered. Output is packed.

Today's Automatic Information Retrieval System is significantly improved over its progenitor. What its greater simultaneous search capacity has meant to our information searchers and their customers is indicated in Fig. 9. The ability to develop more search questions per

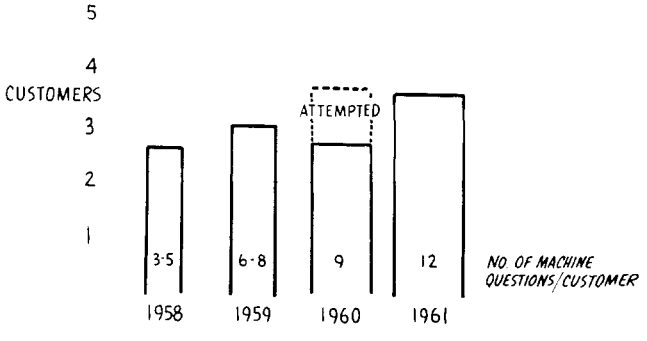


Fig. 9.—AIRS Customers/Run and Questions/Customer.

customer and to composite them into machine screened output has led our searchers to increase the thoroughness of their searches. This, of course, has meant more complete coverage of the file for the customer.

VII. AUTOMATION COSTS

Using a computer for information retrieval has introduced a new and significant item of expense in the operating budget of the Technical Information Center. It also has focused attention on the costs of information searching, both machine and manual. Contrary to the popular misconception that information services are "free," only a cursory examination of any information service or library will reveal that literature or information searching involves considerable costs, many of which are hidden beneath a benevolent cloak called "overhead." As an organization introduces a higher order of selectivity and quality control into its information handling work it must expect to pay the price.

Our Technical Information Center has incurred costs to develop, maintain and operate its Automatic Information Retrieval System. Our automation expenses for 1961 and those forecast for 1962 are shown in Fig. 10. The 16% decrease in total automation costs forecast for 1962 is largely attributable to reduced programming requirements

ITEM	1961 $ ACTUAL	1961 % TOTAL	1962 $ FORECAST	1962 % TOTAL
RETROSPECTIVE SEARCHES	7020	23	7000	28
CURRENT AWARENESS	820	3	2000	8
AIRS PROGRAM DEVELOPMENT	16,160	54	8500	34
PROGRAM MAINTENANCE	1000	3	1000	4
SPECIAL PROJECTS & PRINTOUTS	960	3	3000	12
AIRS KEYPUNCHING	1945	6	500	2
AIRS UPDATING	2195	7	3000	12
	$30,100		$25,000	

Fig. 10.—AIRS automatic costs.

and to improved operating economies. These figures reflect not only those costs due to our Automatic Information Retrieval System but also include several supporting services. It should be pointed out that automation expense forecast for 1962 represents less than 10% of the Center's total operating costs.

Cost of performing retrospective searching is shown in Fig. 11. This operating cost is based on the present file of over 65,000 documents. It is interesting to note that the search price range of $115 maximum to $22.50 minimum for 10–12 machine questions for a customer (the price depending upon batching) compares with a range of $130 maximum to $20 minimum for eight machine questions per customer and a file of fewer than half as many documents in the early days of our 704 system.

VIII. UTILIZATION

The extent to which our retrieval system has been used for retrospective searching is shown in Fig. 12. Obviously, at the rate of less than one machine search per day we have not been deluged with requests. Or, could this be a reasonable volume of retrospective searches in our kind of business? Could it be that the information service needs of our scientists and engineers include something other than retrospective searching? Also, does changing the customer not tend to restrict his use of the service? We believe the answer to all three questions may be yes. Based partly on the slow growth in utilization of our system and partly on the degree of success attained thus far with a broad scale Current Awareness Service, which we have conducted for the past year for over 100 aerospace materials scientists and engineers, we are becoming increasingly convinced that the major application for our Automatic Information Retrieval System is selective dissemination of information with retrospective retrieval a by-product.

It is true too that hanging a price tag on our information searching service has separated the curious from those who really require information. That this practice is bad with respect to the objectives of traditional "free" library service probably is true. However, what better measure of the value of a service than a body of paying customers coming back for more? Also, the "free" service objectives tend to become somewhat academic when the service organization cannot meet its payroll. And in our situation, direct sales of services account for approximately one-third of our support.

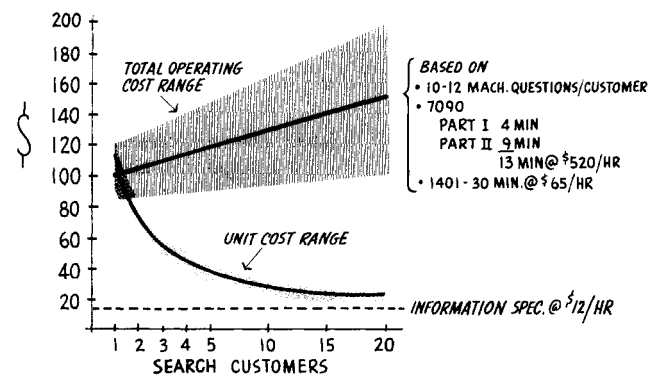


Fig. 11.—Automatic information retrieval system, retrospective search cost.

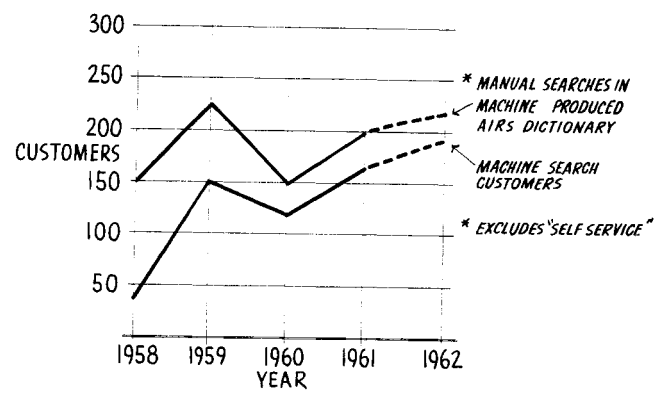


Fig. 12.—AIRS utilization: Retrospective searches.

IX. CONCLUSION AND FORECAST

This presentation has been a sketchy review of five years of experience with the Automatic Information Retrieval System. Perhaps we have over-emphasized our difficulties. We hope this is not the case. From the machine standpoint our system has been running efficiently over the past several months with a high degree of reliability. We can now provide more information at far less cost than would be possible without it. In fact, we are now pricing our Current Awareness Service at $25.00 per month. This includes labor, machine and abstracts and is based on semi-monthly runs. If we are successful in building up our volume, this service could become a significant source of income for we can now handle 20 or more customers with complex interest profiles for less than $80.00 total machine cost per run. As volume increases we expect to lower the price to the customer accordingly.

We are planning no major extensions to the Automatic Information Retrieval System's search logic during 1962. However, we are carrying on several projects related to the system.

We will soon complete 7090 programing effort required to machine print FIND-X, a manual version of our information retrieval system, in a terminal digit listing format and to update both FIND-X dictionary and abstracts on a quarterly basis. This manual retrieval system is used both in the Evendale Plant and at remote locations.

Programming necessary to "clean up" our existing vocabulary and to print a thesaurus now in preparation will begin in the near future. We have not yet decided what available equipment (7090, GE 225, 1401, 407) will be used to print the Thesaurus.

During 1962, we expect to convert our weekly bulletin TIPS from a subject category format to a permuted title index. Or possibly, we may incorporate both methods thereby providing browsability and selective dissemination in the same bulletin. We have tested this idea out on an experimental basis and it looks good.

The only major development we plan for our Automatic Information Retrieval System this year is a new unifile update program for Part I. This new 7090 program will incorporate the mechanical editing of our input vocabulary now done on EAM equipment and to improve economy will update our unifile separate from search runs.

Longer range system development work is in progress or being contemplated. For example we have performed preliminary investigation of an approach to vocabulary control that we refer to an "concept mapping." Work completed using the concept TITANIUM looks promising. When automated, this approach will provide full machine control at the searcher's option, of specific-generic and other term relationships. This will be accomplished without conventional generic posting with its attendant costs.

We have considered a general statement search program. This approach would use the machine to permute a search question into a composite under conditions specified by the searcher. Printed search results would pyramid from most general at the base to most specific at the apex. In the computer field, this would be analogous to the information searcher's changing from using an assembler language to describe his search to a compiler language. This approach should reduce the searcher's effort in creating his search questions. It may lead to direct questioning of the Automatic Information Retrieval System by engineers and scientists.

OBSERVATIONS

Perhaps surviving nearly five years of automation entitles us to offer a few words of advice to those who may be contemplating the use of machines in information work. Therefore, we make the following suggestions drawn from our own attempts thus far to blend an inverted coordinate index, abstracts, machines and people into an efficient and reliable information handling system.

A. Fundamental

1. Define your proposed information service users' needs. (At least make a reasonable attempt).
2. Develop both short and long range action and organizational plans to achieve your established service objectives.
3. Make a best effort attempt to estimate costs of implementing your plans (and then double it).
4. Review your plans and costs and adjust them to the needs of the business of which you are a part. (This will take some real doing).
5. Sell your proposed system and plans to the appropriate level of management (those who really control the pursestrings).

B. Assuming the above Steps Have Been Successful

6. Hire an imaginative, machine oriented information systems specialist. Get your brakes relined so you can hold him down to reality.
7. Hire an expert documentalist with considerable technical depth in those subject areas of major interest to your users (a Ph. D. in at least three areas of specialization is about the right level of competence).

C. If Your Money is Still Holding Out

8. Organize your information systems work into manageable phases.
9. Plan each phase to result in a useful, measurable service at the earliest feasible date.
10. Periodically review the system's services to insure that they are directed to users' needs. Build in systematic feedback.

D. And Here Are a Few Specific System Suggestions

11. Build an audit function into the system. Keep lots of data.
12. Be sure to document each phase of the work. Don't leave an obscure trail.
13. Design maximum flexibility into the system. Your knowledge and users' needs will change.
14. Strive for users' early understanding, utilization and acceptance.
15. Standardize—be consistent (being consistent may be more important than being right). Strive to handle the "typical" not the exception.

If you give full consideration to these suggestions and incorporate them in your work, a successful effort is not necessarily assured. However, your system development work will surely start off on a sounder basis and you can concentrate on making other mistakes.

Library Information Retrieval Program

By GRETCHEN W. KORIAGIN
Missiles and Space Systems Engineering,
Douglas Aircraft Company, Inc., Santa Monica, California
Received May 17, 1962

The Missile and Space Systems Engineering Library was first established as a small department library in 1956. At first all technical reports were indexed according to originating company and author only and were filed by originating company. There were no title or subject files. In early 1958 the expansion of the library holdings made improved report indexing a necessity. Various indexing systems were studied, nearby technical and university libraries were visited, and librarians consulted. The manual Uniterm System of Coordinate Indexing[1] was adopted, and these files and procedures were established late in 1958:

1. The cataloger completed a work sheet indicating the titles, authors, originating agencies, date and subjects[2] of the document.
2. The subjects were uniterms, or unit concepts, selected from the document.
3. The subjects, with appropriate cross references, were compiled in a card file authority list for use by the catalogers.
4. Accession numbers (prefix "ML" for Missile Library) were assigned to each document.
5. The documents were filed by this accession number to conserve filing space (previous files of documents by originating agency required leaving spaces for expansion in each file), to eliminate misfiling, and to aid in retrieval of documents.
6. Card files were established by originating company, title, author, and accession-number so that a report could be located by many types of reference.
7. The index card files were made by typing masters from the work sheets, printing on card stock, and cutting to 3″ × 5″ size. Many copies of each card were made to be placed in the various files. The filing points were indicated by red lines.
8. The subject index for the documents consisted of the 5″ × 8″ Uniterm Subject card for each subject used. The accession numbers were manually posted on the Uniterm Subject cards from the work sheets.

The Library performed manual literature searches upon request. The manual literature search of the documents was, and still is, conducted as follows:

1. One or more uniterms which best describe the subject are selected.
2. The Uniterm Subject cards for each subject are compared for common accession numbers. Those accession numbers which are common to all of the cards represent the technical reports on that subject in the Library.
3. The numerical accession number file is then consulted to further identify the selected documents.

During investigation of indexing systems used in southern California, we examined closely the Uniterm System that was being used at the Douglas Aircraft Company, Inc., Long Beach Location Library. They were pleased with the system but were also finding that it was becoming cumbersome to match cards for heavily posted subjects. At that time the number of reports indexed exceeded 10,000. While we felt that the advantage of the Uniterm System outweighed this disadvantage, we were aware that growth would soon cause a similar problem in the Missile and Space Systems Library. Our cataloging volume was almost 5,000 documents a year. Therefore, the Library was extremely interested when representatives of the Computing Engineering Section approached it in January 1959 with questions and ideas regarding mechanization of information retrieval systems. We had the problems of an indexing—retrieval system that would soon become difficult to operate manually. The Computing Engineering Section had high speed computers and printing equipment, and an interest in this information retrieval problem. We began a period of close cooperation and study of operating retrieval systems with the Computing Section. We found the work done by Mr. B. K. Dennis at General Electric especially interesting.

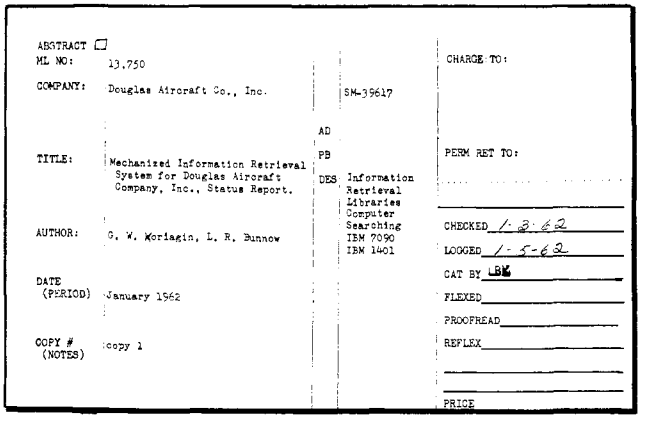

ABSTRACT ☐
ML NO: 13,750
COMPANY: Douglas Aircraft Co., Inc.
SM-39617
AD
PB
TITLE: Mechanized Information Retrieval System for Douglas Aircraft Company, Inc., Status Report.
DES: Information Retrieval Libraries Computer Searching IBM 7090 IBM 1401
AUTHOR: G. W. Koriagin, L. R. Bunnow
DATE (PERIOD) January 1962
COPY # (NOTES) copy 1
CHARGE TO:
PERM RET TO:
CHECKED 1-3-62
LOGGED 1-5-62
CAT BY LBK
FLEXED
PROOFREAD
REFLEX
PRICE

Fig. 1.—Worksheet

[1] Presented before the Division of Chemical Literature, ACS National Meeting, Washington, D. C., March 23, 1962.