

TION/AEROD YNAMIC/HEA TING/HYPER SONIC/FLOW /GAS/FLOW/
 UM/FLOW/RE SULTS ARE PRESENTED FROM CALCU LATIONS ON
 F AN IDEAL DISSOCIAT ING GAS TH ROUGH A HY PERSONIC N
 SHOCK TUBE WITH THE DATA FROM EXPERIM ENTAL RESEARCH PUBLISHE
 SIMILAR CONDIT IONS. QUALITATI VE AGREEME NT IS FOUN D BETWEEN
 55 UNCLASS * UNIV OF SOUTHAMPTO N/OXYGEN /RECOMBINA
 ZLE--COMPARISON WITH DATA FROM SHOCK TUBE WIND TUNN EL EXPERIM ENTS ON AI R PUBLISHE D BY THE G
 SEP 1960. * /BRAMAKIN /1960/YEAR /74555/ACC /U/IDEAL/G AS/DISSOCI ATION/HYPE RSONIC/NOZ ZLE/FLOW/A
 N. ENGLAN D. R-143. ARL-TN-6 0-178.* RE COMBINATIO
 TAL RESULT S FOR AIR FROM GENER AL ELECTRI C RESEARCH
 /74555/ACC /U/IDEAL/G AS/DISSOCI ATION/HYPE RSONIC/NOZ

INFORMATION RETRIEVAL WITH A SMALL COMPUTER

Librarians have staffed library information desks since the earliest days of printed books, their task being to furnish titles to inquiring readers. As collections grew to the extent that librarians could no longer remember all the titles on their shelves, they kept lists in notebooks; still later, the card catalog became the key to a library's wealth of knowledge. Now, with quantities of publications of all kinds increasing exponentially, and information requirements becoming increasingly exacting, the librarians' reference problems have again become critical.

In the world of modern technology, librarians have begun to look to the electronic computer for a solution to their reference problems. It is with this subject—the use of machines for storing and supplying APL Document Library reference data—that this paper is concerned.

Investigations at APL into the use of computers for *information retrieval* (as this relatively new library system is known) began with large computers like the Univac 1103 and the IBM 7090. The cost of owning or leasing and operating such large equipment, however, makes them impractical for most libraries. Since only scant attention had been paid the use of small, hence inexpensive, computers for information retrieval, APL undertook to develop a storage and retrieval system compatible with the IBM 1401. Hourly charges in-

curred by this equipment are approximately 10% of those on the "7090." The system now approaching operational status at the Laboratory was established through the joint effort of F. L. Kennedy, R. P. Rich, and Jane Olmer.*

Information Retrieval System

The retrieval system at APL has one major distinguishing characteristic, other than that of low cost, that makes it especially attractive, namely direct storage. Direct storage is defined simply as that in which all information relating to a single document is stored on one continuous segment of tape. Indirect storage, by contrast, is the method by which information common to a group of documents is stored on tape by subject headings. With the direct method, the computer searches the master tape linearly on request—a matter of simple selection rather than requiring the interim sorting step that is usual for larger and faster computers.

The information retrieval system can be divided into two major phases. The first, or *storage*, phase comprises the several steps necessary to get all information pertaining to a document accurately

* Mr. Kennedy is Supervisor of the APL Document Library; Dr. Rich is Supervisor of the University Computing Center, of which the APL Computing Center is a part; and Mrs. Olmer is a member of the staff of the APL Computing Center.

	1	2	3	4	5	6	7	8	9	10	
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	
0	CC668	745	55 UNCLASS	* UNIV OF	SOUTHAMPTO	N. ENGLAN	D. R-143.	ARL-TN-6	0-178.*	RE COMBINATIO	N OF OXYGE
1	N IN A	NOZ	ZLE--COMPA	RISON WITH	EXPERIMEN	TAL RESULT	S FOR AIR	FROM GENER	AL ELECTRI	C RESEARCH	LABORATOR
2	Y. 27P.	SEP 1960.*	/BRAMAKIN	/1960/YEAR	/74555/ACC	/U/IDEAL/G	AS/DISSOCI	ATION/HYPE	RSONIC/NOZ	ZLE/FLOW/A	
3	IR/ATOMI	C/RECOMBINAT	ION/OXYGEN	/RECOMBINA	TION/AEROD	YNAMIC/HEA	TING/HYPER	SONIC/FLOW	/GAS/FLOW/	FLUID/FLOW	
4	/GAS/DYNA	MIC/FLUID/M	ECHANIC/NO	NEQUILIBRI	UM/FLOW/RE	SULTS ARE	PRESENTED	FROM CALCU	LATIONS ON	THE EFFEC	
5	TS OF ATOM	IC RECOMBI	NATION ON	THE FLOW O	F AN IDEAL	DISSOCIAT	ING GAS TH	ROUGH A HY	PERSONIC N	OZZLE. TH	
6	ESE RESULT	S ARE COMP	ARED WITH	DATA FROM	SHOCK TUBE	WIND TUNN	EL EXPERIM	ENTS ON AI	R PUBLISHE	D BY THE G	
7	ENERAL ELE	CTRIC RESE	ARCH LABOR	ATORY FOR	SIMILAR CO	NDITIONS.	QUALITATI	VE AGREEME	NT IS FOUN	D BETWEEN	
8	THE THEORE	TICAL AND	EXPERIMENT	AL RESULTS	.S						

Sample of an editing printout, showing numbered spaces and lines for the convenience of an editor.

stored on a master tape. The second phase is that in which a request for information is put into the computer and the information requested is retrieved (extracted). This we call *retrieval*.

Storage

The APL Document Library's activity in the storage and retrieval system begins with determining the information required to describe completely a given document and its contents, and then to put this onto a Master Library Tape. Such material is of two kinds: identification, and content description. After this material is collected, it is translated accurately into a language that the computer will accept and from which accurate listings can later be obtained.

The librarian first extracts the identification information (APL accessions number, source, title, author, report number, date, etc.) and the author's abstract. A series of words and phrases is then compiled to describe the contents of the report in more detail. Such words and phrases are known as *descriptors*.

After studying various methods of selecting descriptors, it was decided that they should be taken from the documents during the cataloging process rather than beforehand. Selected in this way, they comprise a listing, or thesaurus, with built-in flexibility. A pre-established thesaurus, on the other hand, has the distinct disadvantage of requiring the cataloger to fit the contents of a document to it. A thesaurus compiled directly from documents has the further value of being more specific and more in accord with the organization using it. While a thesaurus compiled in this way will ultimately become "pre-established," it will retain a utility and flexibility peculiarly adapted to the APL Document Library.

The meanings of descriptors selected in this way will be as restricted as the situation requires, some being limited to a single meaning when the possibility of ambiguities exists, and others having a multiplicity of meanings. They may be single words or combinations of words. They may be condensed words or abbreviations (HI for HIGH, TEMP for TEMPERATURE) in order to shorten both cataloging and machine search time. Synonyms are often used, but in such cases the thesaurus must define carefully the limits of their meanings; these definitions are called scope notes.

As pointed out, descriptors may be combinations of words, e.g. FIRE CONTROL, when they occur frequently and naturally in technical literature. The use of such combined forms makes the cataloger's work easier, especially at the time of search, since he need only accept them as one

Request received from an APL staff member:

Recombination of oxygen or other gases, except hydrogen, in a hypersonic nozzle.

Request translated into search language:

SASASOSE/OXYGEN/RECOMBINATIONSE/GAS/RECOMBINATIONS
N/HYDROGENSE/HYPERSONIC/NOZZLE

A sample search request, including the original request for information.

descriptor rather than having to combine them by means of special search language symbols. These combined forms are also of value in preventing printout of quantities of unwanted document records, such as can occur when the words are used separately.

As the first phase in the information retrieval system, then, the cataloger reviews each document, selects the pertinent identification data, descriptors, and abstract, and punches this information on cards which are then merged with the Library Master Tape by the Computing Center.

Whenever the master tape is updated with new information, an *editing printout* is returned to the library for editing. This printout is designed for the convenience of the editor. The one hundred character positions for each line of data are divided and numbered by tens; the lines for each document (each *section* of the printout) are numbered consecutively; and each section is numbered. This permits an editor to locate an error without tediously counting spaces and lines, editing time is greatly reduced, and the possibility of compounding original errors is prevented. Corrections are sent to the Computing Center for revision of the master tape.

Retrieval

The Library Master Tape is the heart of the information retrieval system just as the card catalog serves a public library collection. Either, properly used, will give a careful researcher the knowledge he needs about pertinent books and documents in the library stacks.

In the machine system, the staff member submits his request to the library, which, in turn, translates his question into search language that can be given to the computer. The information requested will then come to the requester in the form of a computer printout that contains three blocks of information for each document pertaining to the subject of the inquiry. The first section is the identification data, the second is a list of descriptors that embraces the contents of the document, and the third section is the abstract. Taken

CF-2998 UNCLAS* UNIV OF TEXAS, DEFENSE RES LAB, AUSTIN, TEX AS. DRL-495, CF-2998* THE DESIGN AND TESTING OF A UHF PHASE SHIFT RESOLVER. 1P. AND FIGURES. SEPT 1962. NORD 16498.*

/+FISHER/+DAVIS/+WRIGHT/1962/YEAR/CF-2998/ACC/U/UHF/PHASE/SHIFT/RESOLVER/DESIGN/UHF/PHASE/SHIFT/RESOLVER/TEST/UHF/RESOLVER/MECHANICAL/ROTATION/VS/ELECTRICAL/PHASE/SHIFT/UHF/PHASE/SHIFTER/FEASIBILITY/

THE DESIGN AND TESTING OF A UHF RESOLVER WITH AN ACCURACY (MECHANICAL ROTATION VERSUS ELECTRICAL PHASE SHIFT) OF PLUS OR MINUS TWO DEGREES AND AN INSERTION LOSS OF 26.8 PLUS OR MINUS ONE-HALF DECIBEL IS DESCRIBED.

Sample of a final printout in sectionalized form.

together, this information helps the requester to select from the library only those documents that are of particular interest to him.

Before we describe the search method, a characteristic of the system that pertains to both storage and retrieval must be made clear. As will be described, there are inserted in a search request certain symbols, called *operators*, that signal the computer to search for certain ordered combinations of descriptors. While these operators are not used in the storage of information, the library staff must know their use and take them into account during cataloging.

THE SEARCH REQUEST—The search request contains several symbols that are keys to successful searching. They are the virgule (/), magnitude operators \$E, \$G, and \$L, and the logical operators \$A, \$O, and \$N. Used in a search request, these symbols instruct the computer to search the given descriptors in certain clearly established ways. The magnitude equal operator \$E, being the most important of these operators, will serve to illustrate the way in which they are used.

If the descriptor

/SUPERSONIC AIRCRAFT/

appeared in a search request, the computer would search the master tape for all documents having to do with *supersonic* aircraft, but no other class of aircraft. If the requester wished, then, to see documents having to do with FIRE CONTROL of supersonic aircraft, the search request would be made to read

\$E/SUPERSONIC AIRCRAFT/FIRE CONTROL/.

In this instance, the \$E would signal the computer to search the tape for those documents containing the descriptor FIRE CONTROL. When such a document was found, the computer would then look to see if SUPERSONIC AIRCRAFT was also listed as a descriptor for this same document. If this proved to be the case, the document record would be printed out; if not, the computer would ignore this document and continue the search.

Virgules are used to set descriptors apart from each other, each descriptor in a search request being an exact duplicate of one that has already been listed in the thesaurus. The computer will

search only for the descriptors enclosed by virgules, i.e. /SUPERSONIC AIRCRAFT/ in the example above, but not AIRCRAFT/FIRE.

THE PRINTOUT—The printout as it is finally received by the requester is taken from an interim tape rather than directly from the master tape. When a search request on punched cards is first fed into the system, the Library Master Tape is searched by the “1401,” and an identification listing is printed out. The search time for one master tape containing 18,000 records is estimated as approximately 50 minutes. The operator checks the identification listing for two items of importance: the tape label, to ascertain that the correct master tape has been searched, and the number of *hits* or individual document records found in answer to the request. Assuming the operator finds all items to be in order, an interim tape is made up, run through the computer, and printed out in final form.

If it happens that the number of hits appears to be excessive, the librarian is asked to furnish an additional descriptor (in a sense, another limiting factor). With this additional descriptor added into the search request, the *interim* tape is then searched and a final printout is made. The value of this tool is, of course, that the technical staff member will ultimately receive a concise listing of only those documents that are pertinent to his request.

In a scientific organization where so much depends on keeping current with technological change, scores of documents received daily must be made available to staff researchers. These documents should be in *live* storage, subject to immediate use. Primarily, they should be available without involving staff scientists in hours of tedious searching through card files. The electronic computer storage and retrieval system for documented information has the unmistakable advantages of speed and accuracy. To this is now coupled the low operating cost associated with small computers.

Corollary to this discussion, organizations of all types that now use the “1401” or equivalent computers often have many hours of computer free time. This highly uneconomical free time can be put to excellent use in serving the needs of libraries that do not have their own computer facilities. This is the case with the APL Document Library, where the bulk of “1401” time is in support of technological research throughout the Laboratory. Information retrieval tasks can be scheduled into the Computing Center in such a way that the “1401” will operate with a high level of efficiency and so that the library can adequately support the information requirements of the APL staff.