

SWIFT: Computerized Storage and Retrieval of Technical Information*

H. J. ACKERMANN, J. B. HAGLIND, H. G. LINDWALL†, and R. E. MAIZELL
Olin Mathieson Chemical Corp., Technical Information Services Department,
275 Winchester Avenue, New Haven, Conn. 06504

Received December 4, 1967

A unique method of computerized storage and retrieval of technical information is applied in the SWIFT (Significant Word In Full Title) program. SWIFT chooses potential keywords from the titles, compares the keywords with an exclusion word glossary to remove insignificant words and an internal glossary to prevent duplication of terms. The keywords may be either a full or fragmented term. An option is available to index also by author. Indexes, containing full citations, are printed periodically and cumulatively. The magnetic tape file is available for computer search through a sort and print program. Responses to inquiries conducted through the computer program are listed in full citation format.

The Olin SWIFT program is a unique method for computerized storage and retrieval of technical information. It is intended primarily for internal company Technical Reports but is not limited to these. The SWIFT program is so named for its ability to select Significant Words In the Full Title; it also obtains keywords, the terms chosen by the indexer, from the text of the document. It was originally written in symbolic programming systems language for the IBM 1620. Parts of the program have since been modified to take advantage of sort routines for the IBM 1410.

The SWIFT index differs from conventional coordinate indexes and from KWIC Indexes. SWIFT resembles KWOC, or Keyword-Out-of-Content Indexes. However, it is more extensive than WADEX, the word, author KWOC Index of *Applied Mechanics Reviews*, in that the SWIFT Index contains complete citations—i.e., retrieval number, title, author and description (in which appears the source of the document, report number, date written, period covered, and other pertinent information). Sample printouts are described under "Output," and appear as figures below. The citations are assembled in alphabetic sequence, on the first part of each title, beneath each significant keyword. The keywords are listed in alphabetical order.

FRAGMENTING AND FILE USE

The fragmenting of words is one of the basic ground rules of the system and is one of the most significant aspects of the entire system. Chemical names are broken into fragments indicative of the moieties in the molecules. Each meaningful fragment becomes a keyword in the SWIFT processing of the title.

The fragments may be brought together to a Bound Term. Thus, SWIFT takes on an added feature which permits *generic* searching by fragments, and *specific* searching by Bound Terms.

For example, BENZOTHAZOLE is fragmented into BENZO, THI, and AZOLE in the Title section, and is also entered as BENZOTHAZOLE in the Bound Term section of the input. Through this feature of the program, the *specific* name of the compound is one index point. The more *generic* portions of the name are also searchable through conjunctive logic used in the search program.

Certain terms are not fragmented, such as ACET-ALDEHYDE, ANTIFREEZE, HYDROCHLORIDE, and POLYETHER, because they occur so frequently as single terms. The terms ACID and ALCOHOL are tied to their respective modifiers with hyphens; thus, ACETIC-ACID and ETHYL-ALCOHOL are specific terms. Collectively, these terms are designated as nonfragmented terms.

The computer-printed SWIFT Index is used as the primary point of reference to the file of Technical Reports. The periodic printouts are searched for terms pertinent to the document. Through the keywords, generic or specific searches are made of the indexes.

The updated master subject files are also computer searchable. More complex logic may be used in these interrogations than in manual searches of the printed indexes by single terms. Output from the computer search contains the inquiry numbers followed by complete citations (retrieval number, title, author, and description) for the pertinent documents, listed in alphabetical sequence on the first part of the title.

INPUT

The research chemists and engineers, who write the Technical Reports, help index their own documents by providing a list of suggested retrieval words, and by following a standard format on the title page of each report. They consult with the Technical Editor (1) in preparing meaningful titles for their reports and in the selection of retrieval words.

The standard format of the title page permits the Information Retrieval Analyst to mark on this page proper fragmentation of terms, to expand titles, and to add pertinent bound terms or other subject tracings. Keypunchers

*Presented at the Division of Chemical Literature, 151st Meeting, ACS, Pittsburgh, Pa., March 1966.

†Present address: 111 Dessa Drive, Hamden, Conn.

SWIFT: COMPUTERIZED STORAGE AND RETRIEVAL OF TECHNICAL INFORMATION

can prepare punched card input directly from the original document which has been so marked.

The Input Form, Figure 1, is used in the training of personnel connected with the information retrieval system. It can also be used to transmit data to be keypunched. Input may extend for 99 punched cards and even beyond. Cards 1 through 10 are used as Title cards, cards 11 through 13 as Author cards (there are specific fields for

up to nine author names), and cards 14 through 22 as Description cards (for such information as date written, period covered, original report number, and a brief abstract). The additional Description cards, 23 through 36, have specific fields for journal citations, and other information pertinent to input of external literature. Bound Terms or added index words are entered in cards 37 through 99.

1	2	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
CARD NO.		LOCATION (COUNTRY)				YEAR		DOCUMENT, SERIAL, or PATENT NO										LINK		TYPE DOC	
1	2	23 TITLE: EXPANDED TITLE or ABSTRACT.																			80
1	3																				
1	4																				
1	5																				
1	6																				
1	7																				
1	8																				
1	9																				
1	10																				
1	11	23 AUTHOR or INVENTOR:																			80
1	12																				
1	13																				
1	14	23 DESCRIPTION:																			80
1	15																				
1	16																				
1	17																				
1	18																				
1	19																				
1	20																				
1	21																				
1	22																				
1	23																				
1	24																				
1	25																				
1	26																				
1	27																				
1	28																				
1	29																				
1	30																				
1	31																				
1	32																				
1	33																				
1	34																				
1	35																				
1	36																				
1	37																				
1	38																				
1	39																				
1	40																				
1	41																				
1	42																				
1	43																				
1	44																				
1	45																				
1	46																				
1	47																				
1	48																				
1	49																				
1	50																				
1	51																				
1	52																				
1	53																				
1	54																				
1	55																				
1	56																				
1	57																				
1	58																				
1	59																				
1	60																				
1	61																				
1	62																				
1	63																				
1	64																				
1	65																				
1	66																				
1	67																				
1	68																				
1	69																				
1	70																				
1	71																				
1	72																				
1	73																				
1	74																				
1	75																				
1	76																				
1	77																				
1	78																				
1	79																				
1	80																				
1	81																				
1	82																				
1	83																				
1	84																				
1	85																				
1	86																				
1	87																				
1	88																				
1	89																				
1	90																				
1	91																				
1	92																				
1	93																				
1	94																				
1	95																				
1	96																				
1	97																				
1	98																				
1	99																				
1	100																				

(Additional Bound Terms on Reverse Side)

*Hold Term

Figure 1. Information input form

The identification number or retrieval number extends through columns 4 through 22 on all cards. Card columns 1 and 2 are reserved for card numbers, and card column 3 is left blank. In card columns 4 through 8, an alphabetic code is used for the Divisions of the company and the location from which the report originated. Card columns 9 and 10 are used for the last two digits of the year in which the document was written. Card columns 11 through 18 are used for a document number, serial number, or patent number. This number may resemble the original report number assigned to the hard copy report. The

entry may have alphabetic characters left-justified extending from columns 11 to 16 and numeric characters right-justified which may extend the full length of the field, columns 11 to 18.

Link numbers are entered in card columns 19 and 20. Thus, information may be extended beyond the normal 99 cards. Link numbers are also used to give separate document status to distinct parts of a report, especially those reports which cover a variety of topics. The link numbers are indicative of various parts or sections of the original report, identifiable by the retrieval number

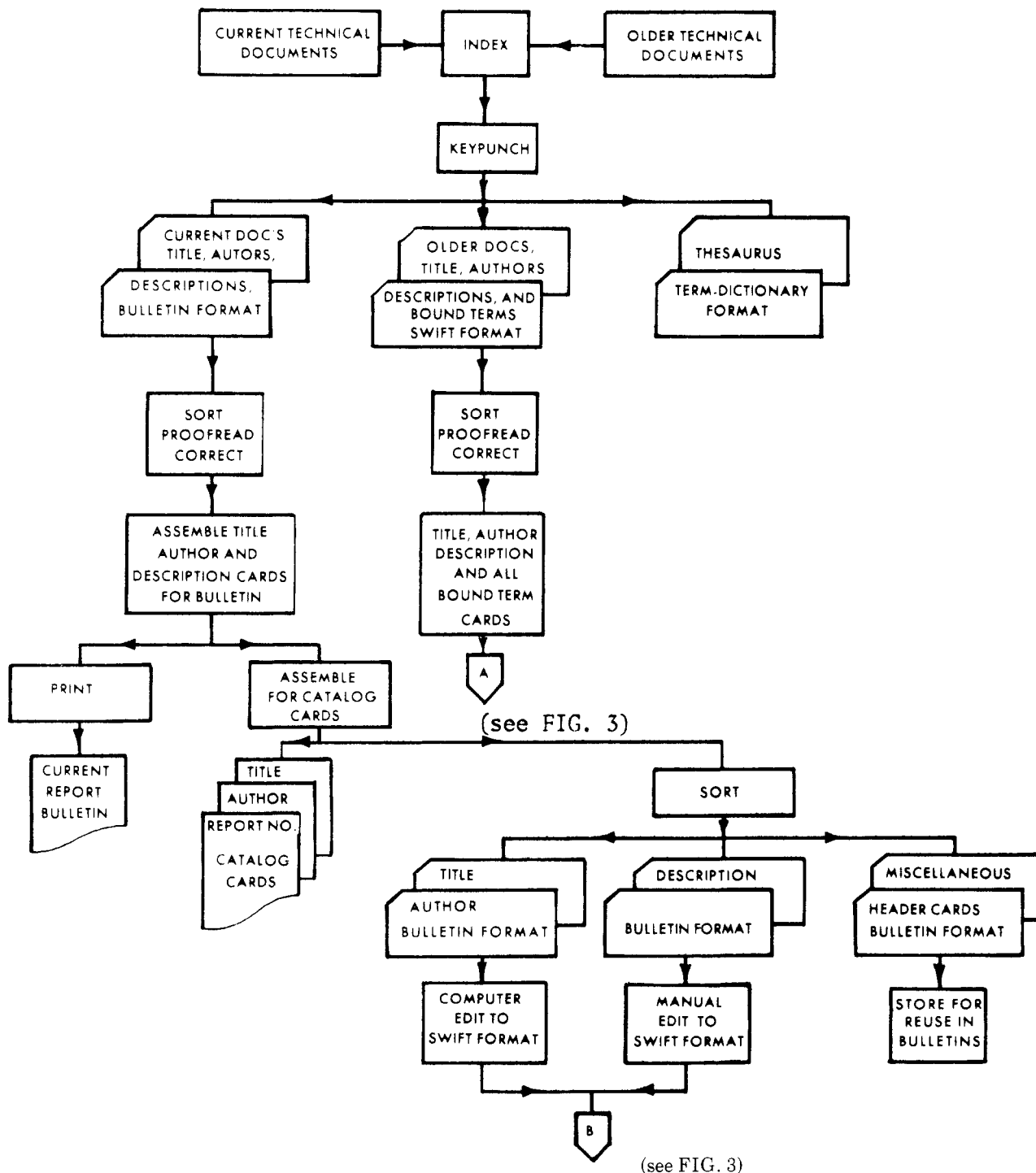


Figure 2. Olin information system

SWIFT: COMPUTERIZED STORAGE AND RETRIEVAL OF TECHNICAL INFORMATION

appearing in columns 4 to 22 of the input form.

In card columns 21 and 22, an alphabetic code is used to identify the document type—e.g., RR is used for research reports, OR for outside reports, RS for research summaries, and SR for survey reports.

Card columns 23 through 80 are assigned for information such as titles, authors' names, description, and the Bound Terms—i.e., additional index points.

The flow chart (Figure 2) illustrates the various components of the Olin information retrieval system. In the indexing step, the documents are marked for the keypuncher. Titles are expanded to make them more meaningful. Words are fragmented in the titles or bound to conform with the nonfragmented term list. Formulas are changed to chemical names. Bound Terms are selected by joining together terms which were fragmented in the title, or taking them from the text of the report. Only one Bound Term may be keypunched per card.

Cards are keypunched in conformance with the input format for a listing of current reports which is printed on an IBM 870 Document Writing system. Signals are inserted for upper case and lower case characters. Key punching is terminated at the end of a word before the end of a card is reached. Margins in the title and description sections are indented.

The input format for older documents (which have not been entered through the current report listing) varies slightly from that described; this is referred to as SWIFT input format. The input is keypunched in upper case characters, word by word to the end of a card, and if necessary, continuing the word without hyphenation in the card column 23 of the next card, for title and description input. The authors' names are keypunched within the specified fields. Only one Bound Term is keypunched per card.

While the document is still available to the indexer and

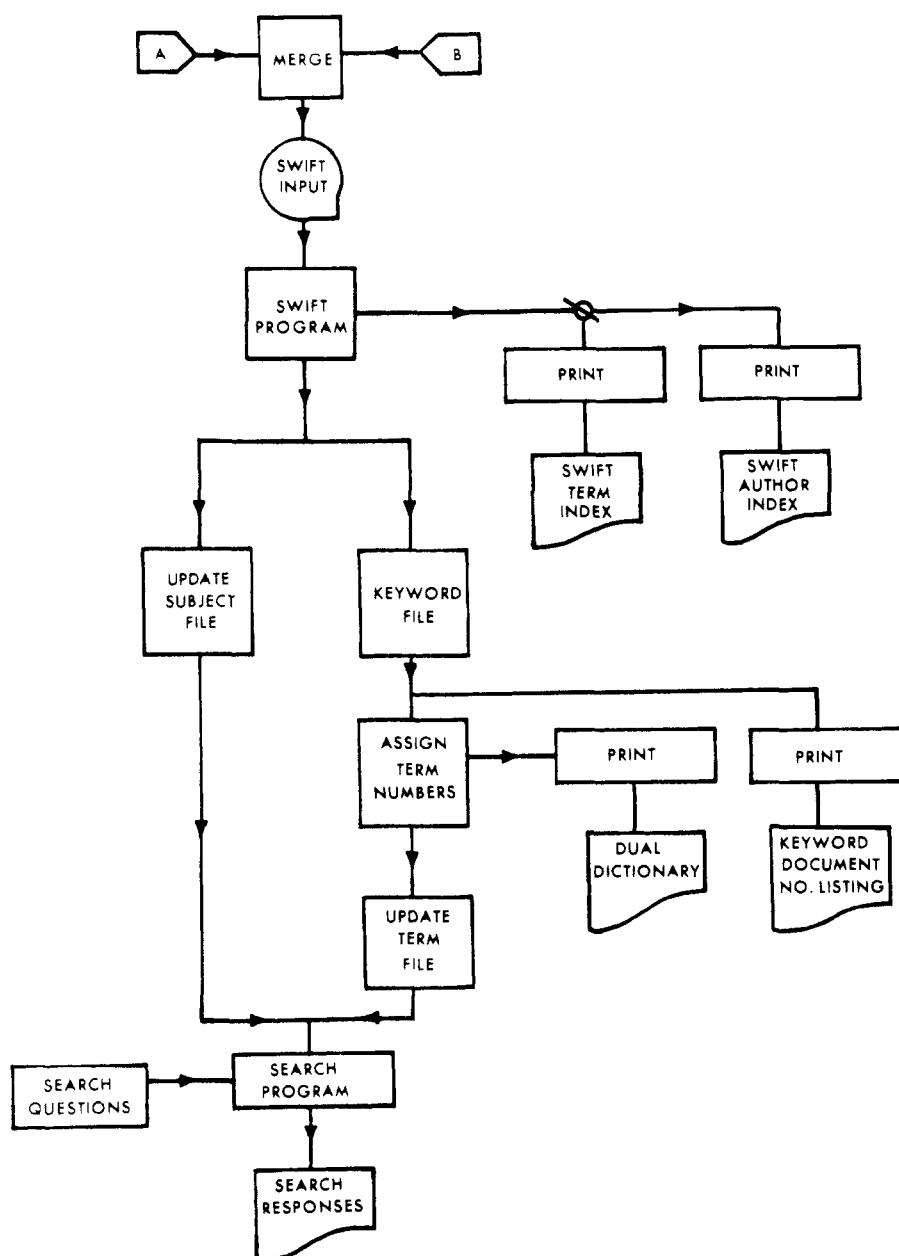


Figure 3. Olin information system

the keypuncher, cards are keypunched for new terms to be added to the thesaurus-term Dictionary. Established thesauri are used in assigning hierarchical order to the terms with the "use" and "used for," "see also," and "scope note" references. The proofreading of all keypunched data is performed from 8%, or direct print, lists of the contents of the cards against the original material. The title, author, and description cards of the older documents and the Bound Term cards for both types of reports are set aside. (See A, Figure 2). The title, description, and author cards of the current technical reports are sorted in card number sequence for the listing of current reports on the IBM 870. These same keypunched cards are sorted again for printing 3 by 5 cards for a conventional catalog as a backup or alternate approach to the files of Technical Reports.

The title and author cards are removed and, via a computer edit program, the data are converted from the current report listing input format to the SWIFT input format. The description information is re-keypunched to the SWIFT input format. The edited cards are merged with the computer edited title and author cards (See B, Figure 2).

Header and divider cards are set aside for production of the next listing of current reports.

As shown in Figure 3, A (the Bound Terms, and the title, author, and description cards for older reports) is merged with B (the cards for the more current reports, edited to SWIFT input format) into sequence by card number, within document number, within year, within location, within division. An 8% list, or direct print of these cards is printed and proofread. The cards are duplicated and interpreted. One set is retained at the Technical Information Services Department and the other set is forwarded to the Computer Center for added protection of input against loss. Prior to SWIFT processing, the input is transferred from card to magnetic tape. A tape sort is executed on the records of the documents to arrange them in alphabetic sequence on the first 30 positions of their titles. The records are then processed by the SWIFT program.

In the SWIFT program the title section is examined character by character and the characters are stored until a blank space is sensed indicating the end of a term. The term is then compared with the external exclusion term glossary, a list of words considered to be insignificant. If it is not matched with that file, the term is "keyworded" and transferred with the complete citation following it, to the work tape. The term is also placed in an internal glossary of exception terms (a temporary list of significant terms for each document) while the citation is "SWIFTed".

The program continues in the title section, sensing again character by character. The next term is picked up and compared with the external glossary. If unmatched, the term will be compared with the internal glossary to prevent duplication of the keywording for this citation. If the term has *not* been entered previously into the internal glossary for this document, it is keyworded and the citation is brought over behind it, to the work tape. This continues until more than one space is unoccupied by a character in the title section indicating the end of this section of the record.

The description section is not processed via the SWIFT program, and therefore it is not searchable. The test proceeds to the Bound Term section. The entire field, which on the cards extended through card columns 23 to 80, will be transferred as a keyword to the work tape. The program tests on the first position, equivalent to card column 23. If a character is sensed, the entire block is picked up as a term. It is not compared with the external or the internal glossaries, but simply transferred to the work tape with the citation behind it. Thus, it is the responsibility of the Information Retrieval Analyst to avoid duplication of terms already appearing in the title. But it also is his option to force a term past the external exclusion term glossary if he finds it to be a term likely to be used in the preparation of an inquiry.

Upon selection of proper switches on the computer control panel, the option is available to include the Author

```

NITROGEN
AIR POLLUTION FROM NITROGEN DI OXIDE,
POOL, LM                                ACO=58-    139-    -LS
HETEROCYCLIC NITROGEN COMPOUNDS PREPARATION,
LOO, CH                                JNY=34-    017-    -LS
PREPARATION, PROPERTIES, REACTIONS AND USES OF NITROGEN TRI
CHLORIDE, SMITH, AB                     ALE, JA     CHEMICAL ABSTRACTS
1907-1952,                                BAL=52-    1150-    -RX
NONIONIC
NONIONIC SURFACTANTS, MARKET-SURVEY
LOO, CH                                JNY=53-    102-    -MS
PHOSPHORUS
REPORT ON PHOSPHORUS- SULFUR SYMPOSIUM
CHARLES, FL                           DX=080160, NORDIO=RES=CONF
GNY=60-                                2468-    -RX
SULFUR
REPORT ON PHOSPHORUS- SULFUR SYMPOSIUM
CHARLES, FL                           DX=080160, NORDIO=RES=CONF
GNY=60-                                2468-    -RX
SWIMMING=POOL
SANITIZATION OF SWIMMING-POOLS WITH HTM,
CANNER, AC                             DW=021659
LNF=59-    1237-    -RX

```

Figure 4. Sample printout of Term/Citation Index

```

ALE, JA
PREPARATION, PROPERTIES, REACTIONS AND USES OF NITROGEN TRI
CHLORIDE, SMITH, AB                     ALE, JA     CHEMICAL ABSTRACTS
1907-1952,                                BAL=52-    1150-    -RX
CANNER, AC
SANITIZATION OF SWIMMING-POOLS WITH HTM,
CANNER, AC                             DW=021659
LNF=59-    1237-    -RX
CHARLES, FL
REPORT ON PHOSPHORUS- SULFUR SYMPOSIUM
CHARLES, FL                           DX=080160, NORDIO=RES=CONF
GNY=60-                                2468-    -RX
LOO, CH
HETEROCYCLIC NITROGEN COMPOUNDS PREPARATION,
LOO, CH                                JNY=34-    017-    -LS
NONIONIC SURFACTANTS, MARKET-SURVEY
LOO, CH                                JNY=53-    102-    -MS
POOL, LM
AIR POLLUTION FROM NITROGEN DI OXIDE,
POOL, LM                                ACO=58-    139-    -LS
SMITH, AB
PREPARATION, PROPERTIES, REACTIONS AND USES OF NITROGEN TRI
CHLORIDE, SMITH, AB                     ALE, JA     CHEMICAL ABSTRACTS
1907-1952,                                BAL=52-    1150-    -RX

```

Figure 5. Sample printout of Author/Citation Index

section in the keywording operation. The test proceeds to the author section, sensing the first position of each of the three fields for authors. If a character is present, the entire block of positions for the field is picked up and transferred as a term to be keyworded with full citation into the work tape. The test continues on author fields until a field is located which is unoccupied.

The keyworded data are then alphabetically sorted with the complete citation immediately following each of the keywords. Such citations still contain the title, and the authors, followed by the description section and retrieval number. A printout of keywords, without citations, but with retrieval number is obtained for final proofreading. A comparison with a hold term glossary is completed to remove from the keyword list those terms which would occur repeatedly without being significant in the printed form of the index. These hold terms, however, are reserved as keywords in the subject file for future computer searching.

The edited keyword lists are returned to the Computer Center. Sorts are completed to remove the hold terms and their citations.

OUTPUT

The accepted keywords with their citations are printed as periodic indexes. Two separate indexes are prepared—one by term and the other by author—for the documents which have been entered in that particular period.

The sample printout (Figure 4) of the Term/Citation Index is illustrative of the KWOC format used in SWIFT. The term, or keyword, is printed in a position which permits easy scanning. The titles are indented and appear in alphabetic order on first part of title. Any additional descriptive information appears on the same line with the author's name.

The Author/Citation Index is illustrated in a sample printout (Figure 5). Here the citations appear under the name of each author in a similar format. Author's names are arranged in ascending alphabetic sequence, with citations in sequence on first part of title.

The output of the SWIFT program is used to update the subject file and the keyword file. From the latter we print the keyword/document number listing which is used for the proofreading of input. Term numbers are assigned and a term number dictionary is printed. At this time a term file is also updated.

Listings of the contents of the entire files may be printed upon request. Thus, a cumulative listing of bibliographic citations in sequence by document number may be prepared. Similarly, a cumulative keyword/document number listing may be printed.

As mentioned, we have the option to use the keypunched input for SWIFT to produce conventional 3×5 catalog cards.

SEARCH

Immediate access to the contents of the files may be gained by manually searching the printed periodic indexes of input to the system. The effort of manual searches can be reduced by using updated Annual Indexes.

More complex questions are computer searched. These search programs are put into operation only after the inquirer is interviewed by the Information Analyst to obtain complete insight into the nature of his request. The inquiry is then formulated with selected keywords and entered on a retrieval input form using term numbers and boolean logic operators, such as "and," "or," "and not" or "from-to" operators. The "from-to" logic operator affords one the opportunity to search physical or chemical property data over various ranges.

Searches are conducted on the master files updated with the information most recently processed by the SWIFT program. Responses to the inquiries are printed in a format which contains the question number, above and towards the left margin, followed by citation—i.e., the title, author, description and retrieval number—indented and below the question number. Confidential material may be withheld from the search responses.

To reduce the expense of printing cumulative SWIFT indexes, the bibliographic citations have been printed in document number sequence, and a cumulative keyword/document number dictionary has been produced as a companion to this cumulative bibliographic index.

Based on our experience with SWIFT we are developing new systems to meet changes in needs and equipment.

ACKNOWLEDGMENT

Acknowledgment, for their contributions to the development of the system described, is extended to S. Rock, J. Ogle, and J. Grald of the Squibb Scientific Computing Center, East Brunswick, N.J.; and to V. C. Pitts and J. T. Godfrey, formerly of Olin, New Haven, Conn.

LITERATURE CITED

- (1) Clark, B. F., *J. CHEM. DOC.* 6, 245 (1966).