

## A New Approach to Indexing Technical Reports in an Industrial Information Center\*

B. S. SCHLESSINGER,\*\* H. J. ACKERMANN, T. M. MANNING, and R. E. MAIZELL  
Technical Information Services, Olin Mathieson Chemical Corp.,  
275 Winchester Ave., New Haven, Conn. 06504

Received September 9, 1968

**The system described uses an unstructured indexing output which is stored on punched cards. The file has been used to generate printed indexes, in an associated key-phrase form, with several different approaches. Production of 5 × 3 card indexes with the IBM 870 Document Writer and the IBM 360 has been achieved. Suitability of the IBM 1800 for printed index approaches are compared, and future plans for the system are described.**

This work was designed to provide a system capable of handling a variety of indexing tasks with a simple indexing routine, followed by machine handling of indexed and coded information. Because our basic concern was with Olin reports, this was the first file chosen for indexing purposes.

The indexing instructions (Figure 1) are written to allow document indexing either by high-level clerical or professional personnel with minimal training in indexing. The indexing entries are key-phrases which indicate the content of the document. The indexing language is free form (no thesaurus), limited only by certain basic rules. Consistency is provided through editing by an experienced indexer.

A system using clerical indexers acquainted with Olin interests has proved very satisfactory. The indexer chooses key-phrase entries from the title, the abstract, and the body of the paper. In addition, he reviews information retrieval terms suggested by the report authors in accordance with standard instructions for Olin report writing.

Key-phrases chosen by the indexer are supplemented by author entries, title, document number, and additional descriptive material extracted by clerical help [such as date written (DW), period covered (PC), organizational division, etc.]. The material is then placed in a standard format compatible with material previously entered into the Olin information retrieval system by SWIFT routines, described in a previous publication.<sup>1</sup>

The compatibility in formats is designed to permit utilization of the complete punched card file with any future programs or new computers, without the necessity of writing special routines for special formats. Although the formats of the SWIFT and the present system are similar, it should be emphasized that the basic indexing in the SWIFT system was done by machine, whereas the basic indexing in this system is accomplished by clerical personnel with minimal training. In both systems, the basic indexing is checked by a trained indexer.

### Figure 1. Instructions to Indexers

Index entries should be made principally on the basis of the title of the report. Where the report title is not indicative enough of the subject matter, the abstract should be used. On occasion, it may be necessary to use the report itself to choose the entries.

The basic categories into which all material to be indexed falls are nine in number:

1. Biological or Chemical Entity. The entry here should be made on the basis of the author's terminology.
2. Preparation. An entry should be made at *A preparation*, where *A* is a chemical compound. A *synthesis*, *A formation*, or *A manufacture* are related entries which may be made.
3. Analysis. Determination, Detection, Separation, and other related ideas should be entered at *A analysis*. There is no distinction made between analysis of a substance and for a substance.
4. Uses. These should be specified if they are pointed out by the author as important or seem to be important from the Olin standpoint.
5. Biological Properties. Entered as given.
6. Physicochemical Properties. If studied, these should be entered. The beginning word should be one that would collect data together; e.g., magnetic properties of all types (magnetic susceptibility, magnetic induction, magnetic polarization) should begin with the word "magnetic."
7. Apparatus. For the name, depend on the author's terminology. Do not index apparatus just because it has been used in an experiment. Make the index entry only if it is a new apparatus, or an adaption of apparatus in use.
8. Check the information retrieval page supplied by the author to identify possible indexing terms you have overlooked.
9. Include all Olin Registry Compounds studied. Use the Olin Registry number as the index entry.

After coding, the material is keypunched onto tab cards, which serve as a data base for production of the desired index form. The keypunching is accomplished as with the SWIFT process described in an earlier publication.<sup>1</sup> The indexes generated from the punched cards have been 5 × 3 card files (using both an IBM Document Writer and an IBM 1800). The 1800, although not used in the past for this kind of processing, has proven suitable for this work.

The 5 × 3 cards produced via the 870 Document Writer and the IBM 360 provide three files. Each file is keyed to a document number which provides access to original

\* Presented before the Division of Chemical Literature, 156th Meeting, ACS, Atlantic City, N. J., Sept. 13, 1968.

\*\* Present address: Department of Library Science, Southern Connecticut State College, 501 Crescent St., New Haven, Conn. 06515.

documents, filed by document number in our Technical Reports Center.

The *document number file* contains cards with the document number and the title (Figure 2). This permits searching by document number to obtain the title of the document and to help determine whether the document is of interest.

The *report number file* cards (Figure 3) relate the number assigned to the document by the originating department or author to the document number assigned by the Technical Reports Center. Because the request for a document or reference to a document may be via report number, the file provides a valuable link for searching purposes.

The dual-purpose *author-subject file* is alphabetically arranged. This file permits search for a report by any author, as a separate author card is made for each author. A typical author card is shown in Figure 4. The presence of the title on the author card permits screening by the author's name for a specific report. The file also permits search by subject. The subject cards are printed as illustrated in Figure 5. Each key-phrase chosen for a document is printed at the top of the card. Up to 13 other key-phrases may follow. These help establish the content of the document from which a specific key-phrase was selected.

The 5 × 3 cards for the index files have been generated, as noted, by routines for the IBM document writer and via an IBM 360 program. A comparison of costs for the two methods of generation is given in Figure 6.

CASNH 60 1410 RR	DW-093060
REPT. NO. CASR-22-60	
DRY-ICE MANUFACTURE	

Figure 2. Document number cards

REPT. NO. CASR-22-60	
CASNH 60 1410 RR	DW-093060

Figure 3. Report number cards

SMITH, R. H.	
CASNH 60 1410 RR	DW-093060
DRY-ICE MANUFACTURE	

Figure 4. Author cards

DRY-ICE MANUFACTURE	
CASNH 60 1410 RR	DW-093060
CO <sub>2</sub> , SOLID-, MANUFACTURE	
SALTVILLE-PLANT	
DENSITY DETERMINATION	
X ANALYSIS	
Y ANALYSIS	

Figure 5. Subject cards

	5 × 3 Cards IBM 870 (\$1.50/Hr.)	5 × 3 Cards IBM 360-30 (\$50/Hr.)	Book Index IBM 1800 (\$27/Hr.)
IBM card sorting	\$80.00	\$5.00	\$5.00
Machine processing	\$104.00	\$18.00	\$13.00
Filing time	\$80.00	\$80.00	Machine filing
TOTAL COST	\$264.00	\$103.00	\$18.00

Figure 6. Costs of index production for 100 documents

The 5 × 3 card files are useful searching tools. However, they pose a problem in that many cards are generated, and hence substantial filing time is required. For a sample of 4620 documents processed by this system, the average number of cards per document was 9.6, consisting of 6.3 subject cards, 1.3 author cards, 1 document number, and 1 report number card. Filing time (including all sorting necessary) per 1000 cards (100 documents) is 16 man-hours.

To eliminate filing time, the IBM 1800 was programmed to produce a book print of the index entries generated. For the initial index, a two-volume print was used, the input again being the same tab card data base. The first volume contained key-phrase *vs.* document number; the second volume contained document number *vs.* title. The results are shown in Figures 7 and 8.

ELECTRIC CONTACT RESISTANCE	CNH	65	240	LS
ELECTRIC CURRENT DENSITY	CNF	49	2512	ER
ELECTRIC CURRENT EFFICIENCY	CNF	52	2538	ER
CNF 59	93 RR			
ELECTRIC DECOMPOSITION POTENTIAL	CNF	41	2465	ER
ELECTRIC DISCHARGE	ENH	59	22	RR
ONH 61 O	49 RR			
ELECTRIC INSULATION	XNY	65	4	MS
ELECTRIC MOMENTS	CNH	66	642	RS

Figure 7. Subject-document no. index

ANH 56	1364	PR
BUDGET AND PROGRAM FOR CHEMICALS R & D		
SMITH, RH		
CHEMICALS DIV., R & D, PC-1957		
ANH 58	1	LS
FERTILIZER BIBLIOGRAPHY		
AGRICULTURAL DIV.		
DW-1958		

Figure 8. Document no.-title index

In Figure 6, the comparison of index generation costs for this method are compared with card file methods previously reviewed. It is evident that the 1800-produced book index is much less costly (for 100 documents, about  $\frac{1}{6}$  the cost of catalog cards via 360 and  $\frac{1}{15}$  the cost of catalog cards via the 870), not including programming time. However, one must weigh the advantages of these lower costs (and large reduction in filing time) of the book index against certain disadvantages. These include possible restrictions in searching imposed by reduction of the context material (whereas the book print provides only the subject in a short phrase, the cards contain up to 12 additional subjects chosen from the document) and by the requirement for a two-volume look-up if the title is necessary for full understanding. One should also consider, if filing time is not a problem, the advantage of one index file, easily updated, which the catalog approach offers.

Notwithstanding these possible disadvantages, for many information centers, where relatively low-cost IBM 1130 and IBM 1800 machines are available as process control or scientific instruments, it may be well to consider the use of these computers for index preparation if the more favored (for information retrieval) IBM 360-type computers are not available. The experience derived from use, as well as the useful product, argues for such an approach.

The use of the simple indexing and reusable data base approach outlined here should be considered in any indexing situation where it is desired to achieve an economical, quick index of unsophisticated form, requiring minimal hardware and professional training.

Our future plans call for study of a book print of key-phrase-associated key-phrase index (Figure 9), which

DRY ICE MANUFACTURE-SALTVILLE PLANT-DENSITY  
DETERMINATION-X ANALYSIS-Y ANALYSIS  
CASNH60-1410-PR

Figure 9. Key-phrase-associated key-phrase book print entry

will combine the advantages of searchability of the catalog cards with those of the book index, and to compare the economics of such an index production with our previous methods. We also hope to apply our indexing and processing techniques to other files, perhaps even to production of a conventional book catalog. In each application, we will attempt to index and process so that the data base will be usable with any future data processing developments.

#### LITERATURE CITED

- (1) Ackermann, H. J., J. B. Haglund, H. G. Lindwall, and R. E. Maizell, *J. CHEM. DOC.* 8 (1), 14-9 (1968).

#### NEWS AND NOTES

##### National Auxiliary Publications Service

The American Society for Information Science has established a National Auxiliary Publications Service to preserve, in promptly retrievable form, the masses of supporting and illustrative material for articles in scholarly and research journals.

The new service represents an expansion and modernization of an Auxiliary Publications Service inaugurated in 1937 by the American Documentation Institute and operated by the Photoduplication Service of the Library of Congress. With the view of introducing and establishing this service on a broad national basis, the American Society for Information Science, formerly American Documentation Institute, has entered into an agreement with Research and Microfilm Publications, Inc., which will utilize new microform methods for making supporting material to articles readily available at low cost to editors and to the public.

According to Bernard M. Fry, project chairman of the Service for ASIS and Dean, Graduate Library School, Indiana University, the National Auxiliary Publications Service "represents an attempt to solve a problem of growing urgency for scientific and professional societies and other publishers of scholarly and research journals. This problem is how to make available the large amounts of supporting and illustrative data (tables, charts, photographs, bibliographies, and other source materials) in or attached to articles submitted for publication in these journals."

Manuscripts, tables, charts, bibliographies, and other source materials may be submitted to the National Auxiliary Publications Service by any editor of a scholarly, research, or abstracting journal. The document number, price, and details as to the availability of the deposited material are then published as a footnote to the article itself. As a rule, only material supporting a published article may be deposited. The material is microfilmed and made available at a nominal charge on microfiche.

There are over 100 scientific and professional journals and institutions which now deposit and participate in

the National Auxiliary Publications Service program of the American Society for Information Science. It is expected that, with the new facility and expanded program, the participation will greatly increase.

Material to be deposited with the Service should be sent to:

NATIONAL AUXILIARY PUBLICATIONS SERVICE OF THE  
AMERICAN SOCIETY FOR INFORMATION SCIENCE  
c/o CCM INFORMATION SCIENCES, INC.  
22 WEST 34TH ST.  
NEW YORK, N. Y. 10001

##### Sophar Joins ISI

Gerald J. Sophar has been appointed Washington representative for the Institute for Scientific Information. He is based at the Company's office (Suite 500) at 7315 Wisconsin Ave., Bethesda, Md.

Mr. Sophar joins ISI after several years as Executive Director of the Committee to Investigate Copyright Problems Affecting Communications in Science and Education, Inc., for which he now acts as secretary, and was principal investigator of the study, "The Determination of Legal Facts and Economic Guideposts with Respect to the Dissemination of Scientific and Educational Information as it is Affected by Copyright—A Status Report."

##### How to Use CA Indexes

The combined introductions to the volume 66 indexes of *Chemical Abstracts* are now available in booklet form from Chemical Abstracts Service. The 56-page booklet, which includes the introductions to the subject, formula, ring systems, and hetero-atom-in-context indexes, provides a brief, instructive review of how to use CA indexes. It also highlights the extensive indexing changes introduced with volume 66, the first volume in CA's eighth collective period. The booklet is priced at \$2.00 for single copies or \$1.00 per copy on orders of multiple copies for educational use. Order from Chemical Abstracts Service, Post Office Box 1378, Columbus, Ohio 43216.