SUMMARY

After one year of operation of the consolidated system, the value of the Report Index has been demonstrated in several ways. Of primary importance is the ability to retrieve reports from many departments in answer to a single query of the consolidated file. The number of searches per inquiry as a result of consolidation has been reduced from 2.8 to 1.3. This ratio is still greater than 1.0 due to the forwarding of inquiries to other information centers within and outside Du Pont. Consolidation has resulted also in standardization of techniques and methods, improved manpower utilization, simplified and more efficient computer programming and services, better coordination of research and development work for improvement of information handling, and better over-all economics.

ACKNOWLEDGMENT

A project of this magnitude obviously involved the work of many people. We gratefully acknowledge the direction provided by Carleton C. Conrad, Manager, and James G. Van Oot, Assistant Manager, of the Central Report Index. We also acknowledge the contributions of Robert L. Taylor to the microfiche program, of James E. Crow as a System Consultant, and of Bart E. Holm, Manager of the Development Section, who directed the development studies.

LITERATURE CITED

(1) Van Oot, J. G. *et al.*, "Links and Roles in Coordinate Indexing and Searching: An Economic Study of Their Use, and an Evaluation of their Effect on Relevance and Recall," J. CHEM. DOC. **6**, 95 (1966).

(2) Walker, J. F., "The Singularity Sub-Link—A New Tool for Use in the Storage and Retrieval of Information," *ibid.*, **4**, 45 (1964).

(3) Costello, J. C., Jr., "Storage and Retrieval of Chemical Research and Patent Information by Links and Roles in Du Pont," *Am. Doc.* **12**, 111 (1961).

(4) Herner, S., F. W. Lancaster, and W. F. Johanningsmeier, "Investigation of the Effect of Roles and Links on the Performance of a Mechanized Retrieval System," presented before the Division of Chemical Literature, 148th Meeting, ACS, Chicago, Ill., September 1964.

(5) Hyslop, M. R., "Role Indicators and Their Use in Information Searching—Relationship of ASM and EJC Systems," American Documentation Institute, Proceedings of the annual meeting, October 5–8, 1964, Philadelphia, Pa., **1**, 99 (1964).

(6) King, D. W., "Evaluation of Coordinate Index Systems During File Development," J. CHEM. DOC., **5**, 96 (1965).

(7) Lancaster, F. W., and J. Mills, "Testing Indexes and Index Language Devices: The ASLIB Cranfield Project," *Am Doc.* **15**, 4 (1964).

(8) Montague, B. A., "Testing, Comparison, and Evaluation of Recall, Relevance, and Cost of Coordinate Indexing with Links and Roles," *ibid.*, **16**, 201 (1965).

(9) Sinnett, J. D., "An Evaluation of Links and Roles Used In Information Retrieval," ML TDR 64-152, Air Force System Command, Wright-Patterson Air Force Base, Ohio, July 1964.

(10) Schirmer, R. F., "Thesaurus Analysis For Updating," J. CHEM. DOC. **7**, 94 (1967).

(11) Hoffman, W. S., "An Integrated Chemical Structure Storage and Search System Operating at Du Pont," *ibid.*, 8, 3 (1968).

(12) Crow, J. E., "Microforms and Technical Information," presented before the Division of Chemical Literature 154th Meeting, ACS, Chicago, Ill., September 1967.

A Mechanized Information System for Many Outputs from One Input*

HERMAN SKOLNIK and RUTH E. CURTISS

Hercules Incorporated, Hercules Research Center, Wilmington, Del. 19899

Received December 1, 1967

A highly mechanized information system, oriented to the needs and requirements of an industrial community of scientists and engineers, is described. From a single input of IBM cards, current awareness bulletins are produced from the cards by means of an IBM 870 and cumulative printouts for retrieval are produced by means of an IBM System/360 from magnetic tapes prepared from the cards. The information system is designed to give total cumulative printouts and also selective printouts of information related to disciplines and missions of science, research and development projects, and individual and group interests.

In a broad sense, the primary needs of today's scientist, as it was for his forbears, are mechanisms for being aware of the current literature within his area of activity and for retrieving from the total literature that which he requires to be knowledgeable in his area of activity (*4*). Because of the size, complexity, and growth of the chemical literature, today's scientist cannot afford to be tolerant of information systems and services that do not give him the maximum number of documents within his area of interest and the absolute minimum outside of his area of interest. On the other hand, the economics of information systems and services dictate a design for a community of scientists, not for an individual scientist.

* Presented in Symposium on "Automation of Information Operations," ACS Middle Atlantic Regional Meeting, Philadelphia, Pa., February 1, 1968.

This paper describes the design of an information system that yields from a single input many outputs for the following varied needs and requirements of individual and groups of scientists within a large community of scientists:

Awareness bulletin arranged by
- Discipline and mission of science
- Subject or subject area

Retrieval from the total literature in the system by
- Subject or subject area
- Discipline and mission of science
- Research and development project
- Individual or group interests
- Author, company, document, etc., viewpoints

OBJECTIVES OF AN INFORMATION SYSTEM

The objectives of an information system must match those of the community of scientists it is to serve. In an industrial research and development environment, the scientists need accessibility to the pertinent literature, such as books, journals, patents, company reports, government reports, in those areas defined by the company's products, processes, and research and development programs. To have this pertinent literature available requires that information scientists be aware of the total literature resource from which the pertinent literature may be selected.

Document selection is like flood control. Without it, we would be inundated with more literature than we want or could possibly handle. Even *Chemical Abstracts* must exercise selectivity of documents, and thus, to cover the chemical literature within its objectives, it has selected a little over 11,000 journals from the tens of thousands currently published (*6*). *Chemical Titles*, on the other hand, finds that a selection of 750 journals satisfactorily meets its objectives of literature coverage (*6*). Likewise, a community of scientists in an industrial research and development environment, will have most of its journal literature needs met by a coverage of 500 to 1000 journals (*2, 3*). Indiscriminate accessioning and housing of documents is a great disservice to a community of scientists by diluting pertinent documents and increasing the costs.

Although many information scientists are quite aware of the need to exercise selectivity in the obtaining and housing of documents, only a few seem to realize the need for selectivity of documents that go into communication and retrieval systems, and the need for selectivity in assignment of classification and index terms for communication and retrieval systems. For example, a chemist in practically any environment requires the availability of the *Journal of the American Chemical Society,* but he does not need to have every paper in every issue brought to his attention or put into a retrieval system for his future reference. Which papers are communicated and put into a retrieval system in an industrial research and development environment are determined by the objectives of that environment.

Furthermore, how a document is classified and indexed for a community of scientists is determined by the relevancy of the document to the objectives of the community of scientists, and not by the objectives of the author of the document.

A classification or index term, however, does not tell a chemist whether or not he needs to read the document. An information system designed to communicate and retrieve documents by classification and index terms might be relatively economical for input, but expensive for users who are directed to documents they do not wish to read. Although it is true that many titles of documents are indicative of the contents, an inspection of Tables of Contents of even the better scientific journals shows that titles too often are poor indicators of the contents, particularly as the contents may relate to the objectives of a community of scientists. Consequently, a good information system which is responsive to the objectives of a community of scientists must be both document selective and document-content selective. These requirements thus determine the input to the information system.

INFORMATION SYSTEM INPUT

The input is determined by the objectives of the information system *vis-à-vis* the needs and requirements of the community of scientists the information system is to serve. In our environment, these needs and requirements are defined by the following viewpoints:

- Discipline of science, such as organic chemistry or analytical chemistry
- Mission of science, such as food chemistry or pesticides
- Subject, such as *propylene,* and subject area, such as *Polymer: Olefin*
- Research and development program
- Individual or group interests
- Author, company, and document identification

Furthermore, these viewpoints must be organized for timely communication and optimum retrieval, and, in these uses, must be given to the users in a form and format they approve of. Consequently, the input must be such as to

- Give readable outputs containing the following information for each document in the system:
  - Subjects and subject areas which relate the document to the user
  - Title of document
  - Abstract which relates the document to the user
  - Reference citation, author, author's location, etc.
- Give total printouts in alphabetical order by subject, author, company, etc.
- Give selected printouts in alphabetical order by subject, author, company, etc., for
  - Discipline of science, such as analytical chemistry, or mission of science, such as chemical propulsion.
  - R and D project or individual or group interest

Assignment of subjects and the preparation of abstracts which relate documents to the needs of a community of scientists can be done only by one who is knowledgeable in the scientific areas and who is aware of the information needs of the scientists. He needs such knowledge also for selecting those documents that go into the system. Within this perspective, it is relatively simple to assign codes which relate the document to a discipline and mission of science, to an R and D project, and to an individual or group interest.

FORM AND FORMAT OF OUTPUT

Form and format requirements are quite different for different needs. Although it is convenient and economical

to have a single form and format, communication of current literature, such as a journal literature bulletin or a patent literature bulletin, requires a form and format that is highly readable. Factors affecting readability are: use of upper and lower case, length of line, indentation, use of common punctuation marks, and page size. The need for producing 5 × 3-inch cards for personal files is considerably less demanding in the readability requirement. Least demanding in the readability requirement is the need for cumulative printouts for retrieval. Cumulative printouts for retrieval, however, must be computer produced because of their size and complexity, and consequently their form and format will be controlled by the limitations of the computer and its software.

Inasmuch as the three basic needs—bulletins, 5×3-inch cards, and cumulative printouts for retrieval—are tied to the same input in our system, the different forms and formats are interrelated and are thus constrained by the compatibility of the different machines used. The controlling factor in our case is the IBM System/360 (7), located some miles from the base of operations and available to us on a scheduled basis. The computer is not available for the production of weekly bulletins, nor would it be economical for us to so use the computer—i.e., as our printing plant.

MACHINE REQUIREMENTS AND OPERATIONS

When we set up our computer-based information system in 1963 (the computer then was an IBM 1401-7070 combination), the IBM 870 Document Writing System was the only equipment available which allowed us to produce bulletins with upper and lower case and whose input was compatible with the computer. Other advantages of the 870 have been: low rental cost (about $150 per month for the basic system), relatively small space requirements, an adequate typeout speed of 105 words per minute, easy programming, and a three-way operating control.

The basic 870 System consists of two parts: a control unit and an electric typewriter. The only major difference between the control unit and an IBM keypunch is the addition of a plugboard or panel which, through wiring plus a program card, activates the control unit—e.g., for reading cards—and activates all of the typewriter functions of ON, OFF, TAB, CARRIAGE RETURN, and CASE SHIFT. The 870 is, as its name implies, only a system for producing typewritten documents. It does not have the capability to calculate, to sort and merge, or to rearrange data. It is, however, something more than an automatic typewriter as input can be done external to the system and there is a three-way operating control for producing typewritten documents.

The three-way operating control for typeouts is obtained:

With each input card by the keypunching of special characters to activate typewriter functions—e.g., using
- ☐ to turn ON the typewriter
- $ for CARRIAGE RETURN and
- # to give UPPER CASE letters (Figure 1, CARD A)

produces the typewritten line:

MIDDLE ATLANTIC REGIONAL MEETING

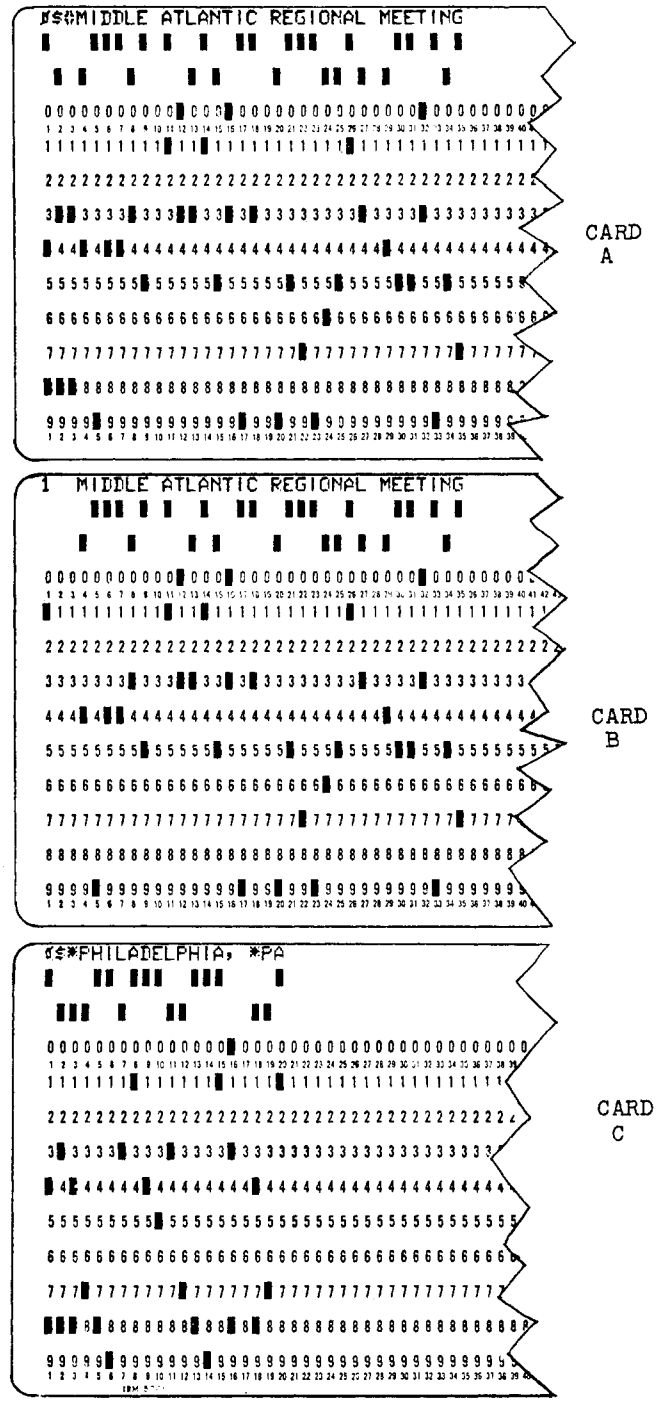

Figure 1. Card inputs

With each input card by keypunching in column one a number which, through wiring of the plugboard plus a program card, activates all required typewriter functions; e.g., using a 1 in column one (Figure 1, CARD B), and the appropriate wiring and program card, produces the same typewritten line as above:

MIDDLE ATLANTIC REGIONAL MEETING

With a combination of the above two methods—e.g., using method 2 for "Middle Atlantic Regional Meeting" (Figure

1, CARD B) and using method 1 for "Philadelphia, Pa" (Figure 1, CARD C) where

⌑ turns the typewriter ON
$ makes the typewriter CARRIAGE RETURN, and
* shifts the typewriter to the UPPER CASE for a single letter, produces the typewritten lines:

MIDDLE ATLANTIC REGIONAL MEETING

Philadelphia, Pa.

Input card control for upper and lower case, unless it can be used in a constant card column and thus readily programmed for omission on transferring to magnetic tape for 360 processing, would yield an awkward 360 printout unless excessive computer time is used to search out, delete, and close up the deleted space of the characters. Consequently, we use only card column one control for lines of all upper or all lower case. Column one control also allows the first letter of a line to be in upper case and subsequent letters in lower case. Thus, although the 360 printout is only in upper case, the 870 typeout is in upper and lower case as shown in Figure 2. Card column one control is also used to activate the other typewriter functions of CARRIAGE RETURN, TAB, and turning the typewriter ON or OFF.

In the 870 System card column one control is done by wiring a combination of three features in the plugboard (*8*):

Column One Only

Format Selector

Transferred Level

The number, or Format Selector, punched in column one specifies that each card with this number is to be typed out in the same format; the Format Selector interrelates the number punched in column one with the Transferred Level which is then wired to the desired typewriter function hub.

The bulletin output format of Figure 2 requires the following combination of instructions to the 870 typewriter:

	Typewriter On	Carriage Return	All Upper Case	All Lower Case	Initial Upper Case	Tab
Subject	Yes	Yes	Yes			
Title	Yes	Yes	Yes			Yes
Abstract	Yes	Yes		Yes		Yes
Company / author /	Yes	Yes			Yes	
Reference	Yes	Yes			Yes	

or a similarity of instruction in only Company / author / and Reference. Four different Format Selectors are thus

SUBJECT
  TITLE OF DOCUMENT
  TITLE OF DOCUMENT CONTINUED
  abstract for document
  abstract for document continued
Company / author / author /
Source

Figure 2. Bulletin output format

CARD COLUMN →	FORMAT SELECTOR 1	870 WIRED COMMANDS 2, 3	4–10, 11–12	INFORMATION INPUT 13–80
	1		[a]	SUBJECT
	2			TITLE
	3			ABSTRACT
	4			COMPANY / AUTHOR / AUTHOR /
	5			SOURCE

[a] In Format Selector 1 only, columns 4–10 are to code for:
subject or subject area
mission of science
research and development project
individual or group interests
and columns 11 and 12 for discipline classification

Figure 3. Input card sequence

required to give the desired output. Any of six available Format Selectors in the 870 could be assigned to each different output pattern. For simplicity in keypunching, however, we assign them in order of input or 1 for subject, 2 for title, etc. Although Company / author / and Reference lines require the same typewriter commands, separate numbers (4 and 5) are assigned to differentiate them in 360 programs.

When a diversified community of scientists is receiving the same awareness bulletin, it is advantageous to produce the bulletin in sections so that the reader can select, if he wishes, only that part which pertains to his discipline. Such a breakdown can be done in many ways. An easy, workable system is a one- or two-letter mnemonic designation such as A for analytical, CO for organic chemistry, or PC for polymer chemistry. This designation serves not only to break down a bulletin but to obtain computer printouts by the various disciplines.

Retrieval from the total literature of a system by subject area, mission of science, R and D project, etc., is handled through a specific code assigned to each document. The Format Selector (column one) assignments with the codes for subject area, etc., the discipline classification, and the input of subject, title, abstract, company and author, and source, and the input card sequence are shown in Figure 3. Either 026 or 29 IBM keypunches can be used.

When our computer facility was converted to System/360, extensive evaluation was made of the IBM 1050 Data Communication System as a replacement for the IBM 870 because the 1050 output keyboard contained the many special characters which were available with the 360. Replacement was not practical, however, because there is nothing comparable to the 870 Format Selector (card column one) for control and because the keypunchers would be required to convert all special characters to a three-column punch (code for shift upper, number equivalent to character on 1052 keyboard, code for shift lower) whenever a 29 punch was used for input.

SUMMARY

Because information in a document can have different meanings and values to different members in a community of scientists, an information system serving a community of scientists needs to be responsive to the differences among the members as well as to the over-all objectives. The

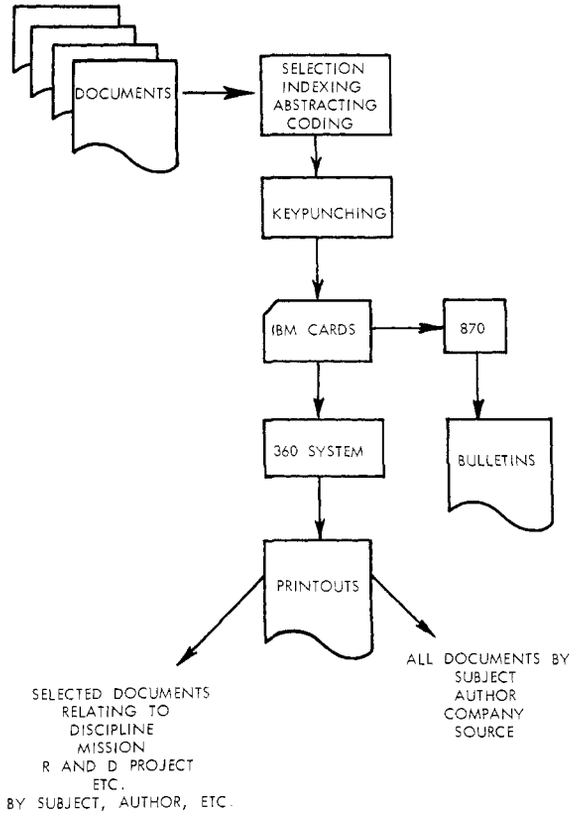

Figure 4. Flowchart of information system

differences among the members and the over-all objectives of a community of scientists constitute an array of variables, such as disciplines and missions of science, R and D projects, and individual and group interests. Within these differences and objectives, the information needs of the scientists are resolved by two mechanisms: current awareness bulletins and retrieval systems.

Traditionally, these two mechanisms, awareness and retrieval, have been separate operations (*1*, *5*); also, documents processed from different viewpoints have been handled as many times as there were viewpoints, or alternatively, the documents have been indexed and abstracted from the viewpoint of the documents without relationship to the needs and requirements of the community of users.

The uniqueness of the information system described in this paper is the mechanized flow from a single input to multiple information products as shown in Figure 4. This mechanized flow has been made possible by the combining of an IBM 870 Document Writing System with the IBM System/360. The objective of this paper has been to treat the information system from the perspective of its uniqueness. Subsequent papers will detail the operating aspects of the machines and of the information system for specific needs and requirements of the community of scientists it serves.

LITERATURE CITED

(1) Friedenstein, H., "Alerting with Internal Abstract Bulletins," J. CHEM. DOC. **5**, 154–7 (1965).
(2) Skolnik, H., "The Hercules Literature Chemist," *Hercules Chemist*, No. 41, 7–9 (February 1961).
(3) Skolnik, H., Chap. 7 in "Vistas in Information Handling," Vol. 1, edited by P. W. Howerton and D. C. Weeks, Spartan Books, 1963.
(4) Sorrows, H. E., "Industrial Technical Intelligence," *Research Management* **10**, 217–27 (1967).
(5) Strauss, L. J., I. M. Strieby, and A. L. Brown, "Scientific and Technical Libraries," Chap. 10, Interscience, 1964.
(6) "CAS Today," Chemical Abstracts Service, 1967.
(7) "IBM System/360 Principles of Operation," 6th ed., IBM.
(8) "Reference Manual. IBM 870 Document Writing System," IBM (November 1961).

Error Control in a Computerized Coordinate Index/Document Retrieval System*

J. L. HOLLOWELL†
Marshall Laboratory, F. & F. Dept., E. I. du Pont de Nemours, Philadelphia, Pa.
Received December 8, 1967

A novel technique has been developed for making substantial reductions in indexer, clerical, and keypunch-derived errors. In use for over 2 years in a medium-sized document retrieval system, real benefits included shortening of "clean-up" time after computer up-dates, less noise in the system for surer searches, and shortened keypunch time. The retrieval system comprises a term coded thesaurus with automatic generic posting, a term-document search dictionary with extensive link and role usage, and a doc-term file. Both machine and manual searches are made.

Error control in a computerized information and document retrieval system is often a significant and onerous problem. This paper describes a group of techniques, some novel, some otherwise, which have been used successfully for the past 2½ years to make substantial reductions in indexer, clerical, and keypunch-derived errors. These techniques have also simplified and speeded up several of the basic processes for inputting and processing information in the system, especially shortened keypunch time, shortened post-update "clean-up," and reduced "noise" in the system.

These techniques of error control have been applied

*Presented in Symposium on "Automation of Information Operations," ACS Middle Atlantic Regional Meeting, Philadelphia, Pa., February 1, 1968.
†Present address, D-7012, F. & F. Dept., E. I. du Pont de Nemours & Co., Inc., Wilmington, Del. 19898