Experience Leading to the Introduction of an SDI System*

JOAN L. GALLAGHER
American Cyanamid Company, Organic Chemicals Division, Bound Brook, New Jersey 08805

Received June 24, 1966

Information services, including an SDI system for the total contents of nineteen journals, are described.

The Organic Chemicals Division of Cyanamid is located at Bound Brook, N. J., and employs approximately 400 professional technical people in Research and Development, Sales, Market Research, and Manufacturing. The Information Services area consists of the Library, a Patent Liaison Unit, the Literature Searching and Internal Report Indexing Staff, a Data Retrieval specialist, and Technical Files, our internal records department.

Each group has been actively engaged in supplying information to all areas within the Division, thereby fulfilling its responsibility by offering services to aid all departments in carrying out their respective roles.

But with the growth of Cyanamid, the information needs of its technical personnel have also grown. Management is not only aware of the information problems confronting the technical man, but also takes an active part in solving these problems by encouraging study and experimentation with new and more effective methods for disseminating information.

For example, during the past two years a semi-mechanized system was inaugurated for retrieving data from internal reports generated within the manufacturing department. Each report is indexed by subject and author, the data are keypunched on IBM cards and stored in two files: a random file for sorting and a vertical file for quick access to specific reports. A dictionary of cross-references is also available for use in conjunction with either file. Research and Development reports have not been included but are indexed more deeply and the entries filed in a standard 3 × 5 vertical file. Although this type of file may be considered "old fashioned" and wasteful of space, it allows for interfiling of index entries to reports issued at another of our research laboratories. Cyanamid has four major locations, only two of which use the same indexing procedure. A special committee is now studying better methods of indexing reports with the intention of establishing a universal system and thereby making information from all company research areas more readily available.

Further study of the information needs of the technical men on the staff showed that rapid retrieval systems would be advantageous in the area of research product testing. A Term-a-trex system was selected for use in storing test results obtained in research on plastics additives. The choice of Term-a-trex was made because of its ease of operation, its ready accessibility, its potential for expansion, its adaptability to data on hand, and, of course, its cost. It has served well as a storage and retrieval unit for this type of data and is now being considered for storing test data from other research product areas.

So far, I have mentioned only two problem areas: internal reports and research product testing data. Both areas form part of a communication complex necessary to the innovation of new products. However, the possibility of duplicating effort reported in both external and internal sources and the need for communication among all departments necessary to ensure product development has been of major concern.

A study was made of the communications pattern necessary to bring a product from the synthesis stage to the time when it becomes established in the line. No less than 60 key people were found to be involved in new product development from start to finish. One possibility to simplify this pattern was to centralize and expand storage and retrieval facilities and thus reduce the need to contact so many people and theoretically reduce the amount of work repeated because of lack of knowledge that the work had already been done.

Such a total information system would have to include:

a. numeric and technical data of all types;
b. analytical procedures, process instructions, legal clearances;
c. data appearing in outside published literature and patents;
d. data from internal technical reports and correspondence.

For a small, medium, *or large* information center this is an overwhelming project.

The problem now was to determine a suitable starting place and decide on one where greatest immediate benefit could be derived. After much discussion with technical staff members, the area of current published literature was selected as the problem area to be considered.

We are a small- to medium-sized library, and distribute Tables of Contents and publish internally, or have immediate access to, selected patent and literature abstract bulletins. Because of the similarities in interest with our Central Research Division at Stamford, Conn., we receive

* Presented before the Division of Chemical Literature, Symposium on Problems of Small Information Groups, 151st National Meeting of the American Chemical Society, Pittsburgh, Pa., March 25, 1966.

and make great use of selected abstract bulletins which they publish.

However, the inadequacies of titles and the possible errors of omission in selecting abstracts, prompted us to investigate SDI, or Selective Dissemination of Information, as one method for ensuring more complete current awareness.

The SDI program developed by IBM (*1*) was put into effect on a trial basis for 1965. The concept of SDI is shown in Figure 1.

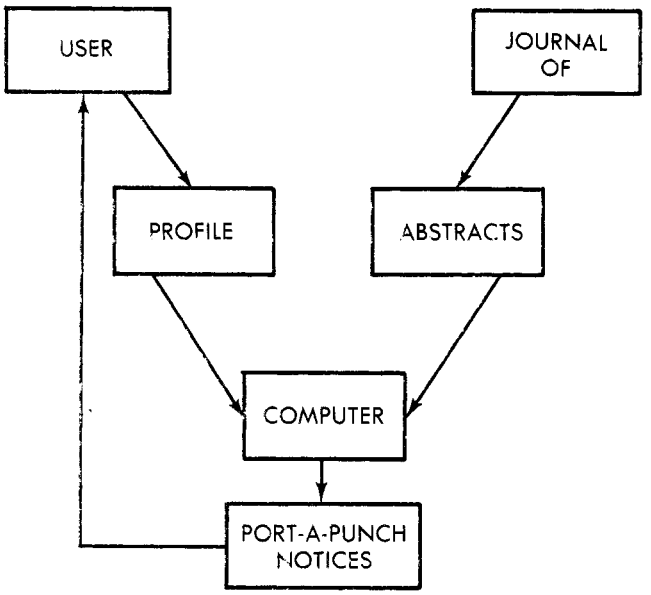

Figure 1. Flow concept of SDI service.

A profile or list of words describing a technical man's interest is prepared and, after keypunching, is converted to magnetic tape. At the same time, articles from current periodicals are abstracted, keypunched, and converted to tape. Several times a month, or week if the need arises, both profile and abstract tapes are compared by computer. Each profile is compared to every abstract and when a predetermined number of words in a given profile match those in an abstract, the abstract is considered to be of potential interest.

The profiles were prepared according to instructions outlined in the IBM manual (2). Figure 2 shows a sample profile.

BG 0412341234561 Name Location Hit Level +4

% 4 ☐	ACRYLOYL CHLORIDE	/ 4	COTTON
% 4 ☐	1,4 ADDITION REACTION	% 4	COUPL
# ☐	COTTON EFFECT ☐	% 4	DYE
% −9 ☐	COUPLING CONSTANT	% 4	HYDROXYMETH
% −9 ☐	ORGANOMETALLIC COMPOUND	% −3	POLYMER
#	UNCI	% −9	SPIN

Figure 2. Sample profile showing word types, weights, and keywords.

% indicates root term; / indicates exact term; # indicates not exact term; ☐ indicates a consecutive word entry.

Each man was given a choice as far as his profile terms were concerned. They could be "exact" terms and as such must appear exactly as they are written before the computer will recognize them. They could be "root terms" and coded so that any words beginning with the letters shown would be recognized. Two other word types, "must terms" and "not terms", were used when necessary. Weights have no effect with must terms. As the name implies, the abstract must contain this term before a printout can occur. A "not term" is equivalent to a weight of minus infinity and, therefore, negates any abstract containing that term.

The weighting of each term allowed clearer definition of the parameters by which articles of particular interest would be directed to a subscriber. These weights are positive or negative numerical values that express the importance of each term in relation both to the subscriber's interest and to every other term in the profile. Negative weights are assigned to those terms which frequently occur with positive terms in articles of no interest. At the same time, an arbitrary hit level is also assigned to each profile. This level is a positive number that must be reached by comparing the terms in an abstract *vs.* those in the profile and algebraically combining the weights of the profile terms. If a level equivalent to or above the hit level is reached after complete matching, the abstract is printed out on a two-part card and mailed directly to the subscriber.

The two-part card, (the left side bears the abstract and the right side is a detachable response card) serves three major purposes:

a. It can be used to request a copy of the entire document.
b. It can be used to comment freely on any aspect of the system.
c. It can be used for machine tabulation of statistics valuable in evaluating the effectiveness of the program.

When the profiling was completed, a cross-referenced list of every term contained in all profiles was prepared for each abstractor.

The pilot run was carried out with 30 men from all areas of research and development who ranged in academic level from B.S. to Ph.D. The reactions of the participants have been very good and we have added 45 new subscribers as of June, 1966.

Since no additions were made to the information service staff, the number of journals was arbitrarily limited to eleven for the trial period. The selection of journals was based on the premise that SDI could not and should not attempt to be a cure-all for a technical man's information and data problems. For example, it will never replace a chemist's necessity and responsibility to read, wherein by sheer serendipity he could be sparked with an idea from a totally unrelated article. But rather it should be looked upon as a supplement to his normal reading habits and hopefully bring articles of interest to his attention from journals that he would not have time to see on a regular cover-to-cover bases.

With this in mind, only journals of a general chemical nature were selected for abstracting. This would ensure some articles of potential interest to everyone regardless of

the product area in which he was involved. We also assumed that each technical man would, out of necessity, see the periodicals peculiar to his own product area and, therefore, included no subject-categorized journals dealing only with such subjects as dyes, plastics, textiles, etc.

The following journals were selected for abstracting:

Angewandte Cheme, International Edition
Journal of Applied Chemistry (London)
Journal of General Chemistry of the U.S.S.R.
Journal of Heterocyclic Chemistry
The Journal of Organic Chemistry
The Journal of Physical Chemistry
Journal of Polymer Science, Parts A,B,C
Tetrahedron
Annalen der Chemie
de la *Bulletin société chimique de France*
Canadian Journal of Chemistry
Chemical Engineering Science
Chemische Berichte
Chemistry & Industry (London)
Collection of Czechoslovak Chemical Communications
Industrial and Engineering Chemistry
Inorganic Chemistry
Journal of Organic Chemistry, U.S.S.R.
Journal für praktische Chemie

The first nine titles were abstracted on a cover-to-cover basis during the trial period. The last ten were added in January, 1966. Each abstract contained an average of 150 words and was enriched by addition of an average of 10 descriptors taken from the whole text of the article. Since the computer scans the complete abstract text, all words—*i.e.*, authors, title, journal title, abstract and descriptors—are included in the matching process.

The preparation of input presented some difficulties. One peculiarity of the program imposed certain restrictions on abstracting procedures. Profile terms of two or more words must be coded to be read as one word (Figure 2). Unfortunately, consecutive word terms in an abstract also require the same special coding in order to match this type of profile term. As a result, use of chemical names as profile terms must be very carefully handled. The clerical work involved in keypunching and verification also added to the input problems. The costs involved have been very high. For these reasons, such a system could hardly be considered a solution to a small information center problem.

We have learned a great deal from experimenting with this system and have been encouraged to do further studies on existing SDI programs and to evaluate their possible application to a centralized system designed to serve all Cyanamid locations. The program described by Sage, Anderson, and Fitzwater (*3*) employs a more sophisticated weighting mechanism than the IBM program and does not require special coding of consecutive words. This program also follows Boolean logic (A and B, A or B, A but not B) and theoretically, should provide a better means of eliminating false drop.

However, the major obstacle to be encountered by a small information center in the effective operation of any SDI program is the time and staff required to completely prepare abstracts for conversion to tape.

It is our belief that the time has come when some kind of program offering selective abstracts to a technical man is necessary. We also feel that such a program would be certainly more economical and best handled if abstract tapes were made available by a centralized commercial abstracting service. If abstract tapes and computer programs were made available to a large chemical population, the cost could be shared and all subscribers would reap benefits they might not otherwise afford if such a program were established on an internal basis. It could then be a practical solution to the current awareness problems of a small information group.

In developing our thoughts along this line, we contacted Chemical Abstracts Service to discuss this as a service they might offer. They are enthusiastic about it and have suggested that Chemical Abstracts Service and American Cyanamid Company join forces to investigate the possibility of establishing a centralized SDI system or an abstracting service that would supply a program and compatible abstract tapes to any chemical company with an interest in SDI.

Other companies may look into such a service for use within their own confines by contacting P. K. Reily, Director of the Marketing Division at Chemical Abstracts Service in Columbus, Ohio.

LITERATURE CITED

(1) General Information Manual, "Selective Dissemination of Information," IBM Technical Publications Department, White Plains, N. Y., 1962.
(2) "Selective Dissemination of Information System 1401-CR-01X," IBM Data Processing Division, White Plains, N. Y.
(3) C. R. Sage, R. R. Anderson and D. R. Fitzwater, *Am. Doc.*, **16**, [3], 185 (1965).