SIDAR: Selective Information Dissemination and Retrieval

By K. D. OFER
Chemicals and Phosphates Ltd., Haifa, Israel
Received July 8, 1963

Chemicals and Phosphates Ltd. employs about 100 professional people having widely diversified technical interests. The manual system for dissemination of current information and literature searches taxed the resources of the Technical Information Department (one chemical engineer and one clerical assistant) to the limit, and with an expansion of the company the introduction of a mechanized system became mandatory. At about the same time management decided to use an IBM 1401 DPS (a card orientated system with 4 k core storage and, for the time being, no further storage facilities) for various commercial purposes, with some machine time left for information work. A survey showed that the only information system existing for this machine configuration is an inverted file system[1] and that all other programs require magnetic tapes or disks for their operation. It was felt than an inverted file would facilitate retrieval of information for only a limited number of documents and would leave the problem of dissemination unsolved; therefore, it was decided to develop a new system. This paper describes the salient facts about its conception.

The system adopted is a modification of the SDI (Selective Dissemination of Information) first proposed by Luhn[2] and implemented by Sowarby[3] for the 1401 with tapes. Keywords were selected from the A.I.Ch.E. Thesaurus and assigned six-character codes (the coding and printing of the dictionary can be done by the 1401). Role-indicators similar to the ones introduced by Morse[4] and Holm[5] are coded as one digit. For the negation of the keyword a zero punch is added to the role-indicator; for instance if "WATER" is the code for water, and 5 indicates the role as a solvent, medium, etc., then WATER V means "non-aqueous solvent."

For reviews covering all aspects of a subject, the role digit is left blank. For each document a keyword card with the serial number of the document and up to 10 codes with role-indicators is punched, as well as a bibliography card of the following format.

Card column	
1	Card code
2- 7	Serial number
8-11	Journal code (as in *Chemical Titles*) or country (for patents)
12-19	Volume, issue, and page, or patent number
20-27	Abstract reference
28-37	Author (two initials and name)
41-80	Title, abbreviated if necessary

Cards with a similar lay-out are prepared for books. The punching can be done from work sheets or from writing on the card itself as suggested by Kirschner.[6] The addition of abstract cards is optional.

Different approaches were thought expedient for the dissemination and for the retrieval function. While, for the former, a certain amount of material which does not exactly correspond to the user's interest would cause no harm (and could even do some good, such as stimulating interest in related fields), only strictly relevant documents should be selected for a retrieval search. It was therefore decided to select current information for dissemination when a single keyword of the document matches a user's interest profile. On the other hand, selection in a retrospective search is based on a complete match of keywords, roles, and Boolean operatives of the query to the document.

DISSEMINATION

The dissemination procedure is effected in two runs, SELECT and NOTIFY (*cf.* Fig. 1). For SELECT the interest profiles are loaded into the 1401 from profile cards (up to 12 keywords per card); the merged bibliography and keyword cards are read in and accession lists are printed in order of serial numbers (Fig. 2). The keyword cards are then compared with the profiles. For each match a selection card is punched, with a format identical with the bibliography card and the identification of the user added in columns 38-40. The profile cards also contain an indication if it is desired to send duplicate notifications to the user's supervisor or to others; in this case an additional selection card is punched for them. A document may not match any of the profiles; this may be due to an error in coding or because the document relates to a field of possible future interest. In these cases a selection card is generated with the Technical Information Department as the interested party. The keyword and bibliography cards are filed.

The selection cards relating to papers from current periodicals are sorted by journal and merged with journal master cards indicating name of periodical, volume number, issues per volume, and identifications of individuals who have placed "standing orders" for this periodical. These decklets, with the remaining selection cards interspersed in random order, constitute the input for the NOTIFY run. In this run, author and subject cards are printed for each document (Fig. 3). For patents a numerical patent number card of similar format is generated. For each book three author and three subject cards are prepared (the catalog is kept in three different locations); stubs on the above cards serve as lending records. For each selection card a notification card with response stub is printed; the wording is slightly different for papers from periodicals (Fig. 4) and from abstracts (Fig. 5). In the same run, routing lists for periodicals

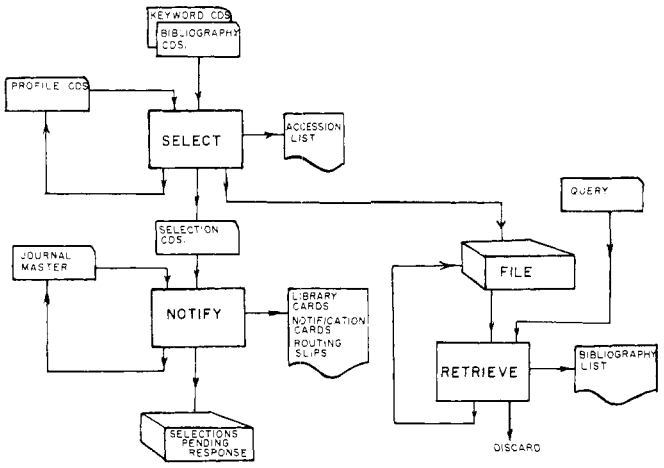

Fig. 1.—General flow-sheet.

```
1234 50  138 1537 MENTOR   N  MACHIAVE IL PRINCIPE /THE PRINCE/ ENGLISH TRANSL.
1235 USA  3654321 58024688     S.A.I.    GRANULAR POTASSIUM METAPHOSPHATE
1236 JACS  841035 58        L  BORGIA       STARCH PHOSPHATE & NUTRITION
1237 JACS  841036 58           AB SMITH       NEW WAYS IN INORGANIC CHEMISTRY
1238 JACS  841040 58           WC WHITE       FERTILIZER USE IN THE FAR EAST
1239 JACS  841262 58           BC BROWN II KINETICS OF THE WATERGAS REACTION
```

Fig. 2.—Accession list.

```
S.A.I.          1235                                    USA  3654321

GRANULAR POTASSIUM METAPHOSPHATE                        1235   AS

USA  3654321      58024688                              S.A.I.

AS                              C&P LTD. NORTH          58024688
```

Fig. 3.—Author–subject–patent number card.

```
TO DSG    FROM TECH.INFORMATION DEPT., C&P LTD.  |  TO TECH.INF.DEPT.

   STARCH PHOSPHATE & NUTRITION                |     FROM DSG
L  BORGIA                    58      1236      |       1236
                                               |
                                               |   INTERESTED
                                               |
                                               |   NOT INTERESTED
                                               |
                                               |   COMMENTS OVERLEAF
                                               |
                                               |

THE JOURNAL WILL BE SENT TO YOU.PLEASE RETURN IT AS SOON AS POSSIBLE
PLEASE DELETE WORD/S/ NOT APPLICABLE AND RETURN STUB
```

Fig. 4.—Dissemination card for journal article.

```
TO HB     FROM TECH.INFORMATION DEPT., C&P LTD.  |  TO TECH.INF.DEPT.

GRANULAR POTASSIUM METAPHOSPHATE               |     FROM HB
  S.A.I.    USA   365,432 CA 58- 024688   1235  |       1235
                                               |
SPACE AVAILABLE FOR ABSTRACT,NOT ONE           |
NOT TWO                                        |   INTERESTED
NOT THREE                                      |
NOT FOUR                                       |   NOT INTERESTED
NOT FIVE                                       |
NOT SIX                                        |   COMMENTS OVERLEAF
NOT SEVEN                                      |
BUT EIGHT FULL LINES                           |   REQUEST COPY
                                               |

  THE ABSTRACT CAN BE SEEN IN THE LIBRARY      |
PLEASE DELETE WORD/S/ NOT APPLICABLE AND RETURN STUB
```

Fig. 5.—Dissemination card for abstract.

are printed; they include all selection cards relating to that particular journal with the addition of standing orders. The form is attached to the periodical; the stub serves as loan register. The journal master card for the next issue is punched, with the issue number (and, eventually, also the volume number) ammended. The possibility that several issues of a periodical are processed at the same time was taken into account, as well as the case of periodicals intended for a single individual only; the routing slip for a whole volume is prepared at the same time. The selection cards are kept until the response stubs are returned.

RETRIEVAL

For retrieval of information from the accumulated keyword file, question profiles are punched on cards which are loaded into the 1401. In addition to negated keywords, negated roles (everything on the subject *except* this role) are permissible. Groups of keywords enclosed in parentheses (%. . . . . /) are considered to be connected by the Boolean "OR"; all others by "AND." By using a sense switch the output can be chosen in the form of a printed list or as punched cards with listing. In both cases the question profile number and the serial number of the document are given; complete bibliographical particulars can be obtained by collating these cards with the file of bibliography cards or by visual searching of the accession lists.

The coding of keywords and the elimination of duplicate profiles by the inclusion of the "supervisor" in the individual's profiles reduce the storage capacity required to about 45% of that for the unmodified system. In a 4 k machine, over 3200 storage positions are available for users' profiles, *i.e.*, 60 profiles with an average of 8 keywords each or 100 profiles averaging 4 keywords each. This means that for the information services of a smaller company, the core storage of the 1401 is of sufficient capacity and brings an efficient and economic dissemination and retrieval system within the reach of such companies.

REFERENCES

(1) M. Newkirk, "Inverted Card File—1401 Information Retrieval System," IBM Corporation, Charleston, West Virginia, 1401 General Program Library 10-31006, April, 1962.
(2) H. P. Luhn, "Selective Dissemination of New Scientific Information with the Aid of Electronic Data Processing Equipment," IBM Advanced System Development Division, Yorktown Heights, N. Y., Nov., 1959.
(3) A. J. Sowarby, "Selective Dissemination of Information (SDI) for the 1401 Tape System, the 650 Tape System, and Fortran 11," ref. 2, 1401 General Program Library 10-3-004.
(4) R. Morse, *Chem. Eng. Prog.*, **57**, 55 (1961).
(5) B. E. Holm, *ibid.*, **57**, 73 (1961).
(6) S. Kirschner, *J. Chem. Educ.*, **38**, 526 (1961).