

requirements of the computer program and the demands of the user dictated the format of the tailor-made indexes used by the several organizations mentioned. These indexes are oriented to the demands of the computer program which in turn is closely tied to the individual computer. Thus, the various forms of keyword indexes are an attempt to provide flexibility in alerting chemists with a rather inflexible machine medium. Despite this the various tailor-made indexes cater to the needs of the company chemist by quickly providing him with information that is readable, easily isolated, and, in some cases, subject controlled. This control may be accomplished by use of punctuation symbols, the choice of words in the "stop list," pre-editing according to *Chemical Titles*, and the arrangement of the elements in the bibliography.

## LITERATURE CITED

- (1) Freeman, R. R., Dyson, G. M., *J. Chem. Doc.*, **3**, 16 (1963).
- (2) Kennedy, R. A., *ibid.*, **2**, 181 (1962).
- (3) Youden, W. W., in "Automation and Scientific Communication: Short Papers," H. P. Luhn, Ed., Part 2, American Documentation Institute, Washington, D. C., 1963, pp. 331, 332.
- (4) Clark, W. E., Greene, A. G., Pflueger, M. L., Ed., "Index to Conferences Abstracted in *Nuclear Science Abstracts*," TID-4043, 1963.
- (5) Weil, B. H., Esso Research and Engineering Company, personal communication, 1964.
- (6) Klein, A. S., Howard, E. B., Alexander, N. B., "The ORNL Library Key Word Index Program," ORNL-TM-525, 1963.
- (7) Johnsson, K. O., Klein, A. S., Ed., "Publications, Reports, and Papers for 1963 from Oak Ridge National Laboratory," ORNL-3300, Vol. 3, 1964.
- (8) Lafferty, R. H., Jr., Baker, P. S., Gerrard, M. W., "Publications on Isotopes at Oak Ridge National Laboratory, 1946-1963," ORNL-IIC-1, 1964.
- (9) Raaen, H. P., Klein, A. S., Ed., "Indexes to the Oak Ridge National Laboratory Master Analytical Manual, 1953-1963," TID-7015 (Indexes), Rev. 1, 1964.
- (10) Kennedy, R. A., Bell Telephone Laboratories, personal communication, 1962.
- (11) Bauer, C. K., *Spec. Libr.*, **55**, 137 (1964).
- (12) Kennedy, J. H., "IBM 1401 Computer Produced and Maintained Keyword-Plus-Title (KWPT) Index for Reports, Photographs, and Specifications," UCRL-7556, 1964.

## A Computer-Based Alerting System for *Chemical Titles*\*

C. N. RICE

Research, Development, and Control, Eli Lilly and Company, Indianapolis, Indiana

Received May 11, 1965

The selective alerting of an individual to current information by machine methods has been described by Luhn (1, 2) and by others (3-8) under the name of selective dissemination of information or SDI.

In general, indexing of reports and papers for SDI is performed in depth with keywords drawn from the text of an article, from an abstract of it prepared for the system, or from the title. The keywords are listed in a thesaurus of indexing terms, or, following analysis, are added to it, and represent preferred terms describing given concepts.

Because indexing is in depth, specific and restricted subject matter can be retrieved and disseminated. The co-occurrence of several keywords serves to restrict the subject field to significant citations. One served by an SDI system may make his requests quite specific, thus restricting the papers that will be provided him to his main interest, or very broad, so that in addition he receives citations to papers that bear tangentially upon his main interest. A match may be required between perhaps as many as eight or nine keywords in the requester's profile list and in the keyword index list of a paper. This represents a statistical criterion for retrieval, which may be relaxed or tightened according to the desires of the requester.

\* Paper presented before the Division of Chemical Literature, 148th National Meeting of the American Chemical Society, Chicago, Ill., Aug. 31, 1964.

Any such system, including the somewhat different one to be described in this paper, will fail to retrieve significant papers. Language is not sufficiently well understood or used, and the indexing principles and practices used in these systems are not sufficiently rigorous to assure retrieval of all information relevant to a given concept. One must therefore accept the fact that he is going to miss important references. But this will be true whether he uses machine or manual search procedures.

There is another price for the advantage one obtains from an SDI system. In order to have great discrimination in getting at desired references, much indexing effort is required. In any system there is a compromise between indexing effort and search effort. If indexing is careful and complete, the cost of input goes up. Since keywords are chosen easily and require only clerical operations, they are favored for indexing purposes as a means for keeping costs down. Index words are more descriptive of concepts and procedures in a paper, but greater care must be exercised in their choice.

The chief virtue of an alerting service is its timeliness. Since human indexing requires time, there develops the need in an alerting service for a trade-off between timeliness and high relevance. The popularity of keywords and the appeal of auto-abstracting and auto-indexing occur, in part, because of the need for expeditious handling of

references. We need not argue the merit of a selective dissemination system on any other basis than that it provide effective, timely service.

#### A CURRENT AWARENESS PROGRAM FOR CHEMICAL TITLES

Mr. Robert Freeman has described the use of computers for alerting biologists and chemists to scientific papers in *Chemical Titles* (9, 10). Since keywords were available only from titles and not from abstracts, or texts of the papers, the system described is unlike SDI in certain respects: requests for information cannot be made as specific; no statistical criteria exists for the choice of a paper.

The present system at Eli Lilly and Company has evolved from the system described by Mr. Freeman. No change has been made in the logic of the computer search. There have been several changes in the manner used to present the data, however. For a given journal, the scientist served may now request citation of all papers listed in the bibliography section of *Chemical Titles*, or he may ask for all papers of a given author found in the author index. These, combined in a logical search request with his keyword vocabulary, give him a more thorough, and at the same time, a more discriminating retrieval of significant *Chemical Titles* citations.

Magnetic tape records of *Chemical Titles* indexes and bibliography were obtained by Eli Lilly and Company from Chemical Abstracts Service under an agreement for an experimental trial of a computer-based current awareness service to members of the Lilly scientific staff. The decision of the American Chemical Society to provide the information from *Chemical Titles* as magnetic tape records in 1965 was based upon the results obtained in our experimental trial and a similar trial with Olin Mathieson Corporation, whereby Chemical Abstracts Service performed the computer search.

#### REQUESTER VOCABULARIES

The individual served is asked to describe his scientific interests with a set of words that would, in his opinion, appear in the titles of interesting papers. The set of words he provides is checked manually against the indexes of *Chemical Titles*. Some words may be spelled differently by various authors, or there may be different British and American spellings—as for the word *hemolytic*. Some of his vocabulary words may be divided as a matter of *Chemical Titles* editorial practice so that the word will appear at several places in the index, for example, so that *cholesterol* may be found under *sterol* as well as *cholesterol*. Word roots may be substituted for the vocabulary word suggested by the requester in order to assure retrieval of what he has in mind. The editorial modifications are submitted to the requester for his approval, and the edited list of words is incorporated into the request file. In the request file, each word or word root is identified with the individual and his department number. The file is alphabetized by keyword.

Our staff members are encouraged to experiment with their request vocabularies. Each individual may add to,

or remove words from, his request vocabulary any time he wishes. The effect of his changes can be seen by comparing the parallel results obtained with the altered and with the unaltered request. Modifying the vocabulary to improve the relevance of citations becomes a sort of game of wits between the requester and the universe of authors who conspire to trip him up. As an additional aid, members of the scientific staff may divide their vocabularies into groups representing prime interests, on the one hand, and secondary interests, on the other. Those who have followed these suggestions have developed superior vocabularies and have weeded out much unrewarding and peripheral material.

Provision is made for simple Boolean inclusion, exclusion, and negation: A together with B; A or B; A but not B. Each logical request, assigned a unique number, has as many punched cards as there are terms in the request.

Upon arrival of each issue of *Chemical Titles* on magnetic tape, the following series of operations is undertaken. The magnetic tape records are edited on the IBM-1401 computer to assure that the records are properly labeled, blocked, and free of transcription errors. The requester vocabulary file on punched cards is sequence checked and edited on the 1401. The indexing keywords in the index are matched against the request keyword file with the 1401. When a match is achieved, the keyword, journal code, requester, and his department number are written on magnetic tape. This working tape is sorted by journal code with the IBM-705 computer. The journal codes on the sorted working tape and the journal codes of the bibliography section on tape are matched on the 1401. When a match is achieved, the journal code, authors, full title, the requester, and his department number are written on magnetic tape. The resulting tape record is sorted on the 705 by department number and requester. A table look-up is used to convert the four-character journal codes to the accepted American Chemical Society abbreviations for journals, and these abbreviations and the full citation are line printed by the 1401 into individual lists to be sent to a specific requester.

In addition, the table of contents for a given journal, as listed in *Chemical Titles*, can be printed. Only a few of our scientific staff take advantage of this capability.

For forty-three individuals with 1,250 requests and 2,100 keywords, 4.0 hours of IBM-1401 and 1.4 hours of IBM-705 time were required per issue of *Chemical Titles* in July 1964. The 1401 used has 4,000 positions of core memory and the 705 has 40,000. A 600 line per minute printer and 729 Model III tape drives with 556 characters per inch on tape are used. The computer programs being used, however, can be significantly improved.

A number of intangible factors having relation to comparative costs of alerting services should be considered. A significant cost would be incurred for our scientific staff to go to the library, take the time to look through *Chemical Titles*, and list the papers of interest found there. In our experience, *Chemical Titles* is not used in the library except by reference librarians, literature searchers, and a very few others.

The selected titles from the current magnetic tape records of *Chemical Titles* are on each man's desk by the time the printed edition arrives in the library. It is impossible to say whether the promptness of the service has

placed one of our staff in a favorable position to exploit an idea before someone else does, or to avoid effort which he might otherwise duplicate. Unfortunately, it is quite difficult to obtain objective data by which to evaluate such a service, and testimonials are simply not adequate or convincing.

There are a number of ways in which *Chemical Titles* itself can be improved for machine searching purposes. There is absolutely no value in a wrap-around index for a machine search, but there is an advantage in a word-sequencing code applied to title words. Information can be provided to the customer much more compactly in such a form, while the usefulness of word sequence for certain keyword matching purposes is retained.

During the experimental trial period, titles from *Chemical Titles* were provided in several different formats. At the start, the citations were grouped under the requester keywords that had caused their selection. Each participant was provided with his own desk copy of *Chemical Titles* for the first three months of 1964. There was much complaining from the experimental group during this period. From April through June, 1964, the participants were provided with the full bibliography sheared from *Chemical Titles*. To the front was stapled a list of journal codes that had been selected by use of the requester vocabularies; that is, everything was done for the participants except the final step of lifting the citation out of the bibliography and printing it for them. The idea of this was that each individual could look up the titles to which he was being alerted, and afterward he would have the bibliography for browsing purposes. This concept was suggested by Mr. Al Warheit of IBM San Jose Laboratories, and appeared to be a good idea. The participants did not like it. The print sizes of the list and of the bibliography were quite different; the journal code sheets were monotonous; thumbing back and forth was a nuisance; and transcription of the reference was necessary if it were to be kept. Many turned the clerical chore over to assistants. Next, a partial table of contents for those journals in which titles containing the requester's keyword appear was printed. The selecting keyword was printed with the full title, author, volume, page, and year. After October 1964, each selected reference with all its attendant information was printed on a separate card and sets of cards were distributed. The participant keeps the cards of interest in his card file and returns card stubs as an indication of his action.

There has not been unanimous support from the scientific staff. A few prefer the printed edition of *Chemical Titles*. Some have been enthusiastic about the idea of receiving selected titles, but feel that their biochemical or biological interests are not covered adequately by *Chemical Titles*. Others, who have been unwilling to put in the effort required to obtain a good keyword vocabulary, or have been unwilling to restrict the scope of their interest, have been inundated by references, and have thrown up their hands in despair. Some subject fields are not as specific and well-defined as others and this has created some difficulties in keyword building for certain individuals.

In summation, an automatic literature alerting service must be timely to be of value. If this is to be possible, information must be rapidly converted to machine-processible form. Costs should be reduced by eliminating secondary transcription of the information. A trade-off will be required between indexing and analysis, on the one hand, and relevance of the references supplied, on the other. Users of alerting services should provide the feedback to the service that will automatically assure more relevant future alerting. Authors and editors have the responsibility to provide more meaningful titles and terms of annotation. Since authors are also users of scientific literature this is not too much to ask. The chemical profession will profit from a degree of vocabulary control obtained from a commonly accepted dictionary of terms.

#### ACKNOWLEDGMENT

The attitudes expressed in this paper have been formed through many informal discussions about automatic literature alerting services. For these discussions I wish to thank many friends and associates, especially Mr. Robert Freeman, Dr. Peter Bernays, Mr. Kenneth Zabriskie, Dr. Malcom Dyson, Mr. William Muirhead, Mr. Max Marsh, and Dr. R. E. Crabtree. For provision of the computer programs I wish to thank Mr. William Shepherd and his associates. I am grateful to the Chemical Abstracts Service of the American Chemical Society for making these trials possible. The patience and cooperation of the Lilly scientific staff who have made our program successful is also much appreciated.

#### LITERATURE CITED

- (1) Luhn, H. P., *IBM J. Res. Develop.*, **2**, 314 (1958).
- (2) Luhn, H. P., *Am. Doc.*, **12**, 131 (1961).
- (3) Hensley, C. B., "SDI Bibliography," Advanced Systems Development Division, IBM, Yorktown Heights, N. Y., Sept., 1962.
- (4) Barnes, A. B., Briggs, A. A., Gauss, J. A., Resnick, A., "SDI-5, An Advanced System for Selective Dissemination of Information," ACM's 19th National Conference and Data Processing Exhibit, Philadelphia, Pa., 1964.
- (5) Dennis, B. K., Battelle Memorial Institute, personal communication, 1964.
- (6) Koriagin, G. W., "Library Information Retrieval System," 141st National Meeting of the American Chemical Society, Washington, D. C., March 1962.
- (7) Westmoreland, M. R., Douglas Aircraft Company, Missiles and Space Systems Division, personal communication, 1964.
- (8) Anon., "The MEDLARS Story at the National Library of Medicine," U. S. Department of Health, Education, and Welfare, Public Health Service, Washington, D. C., 1963.
- (9) Freeman, R. R., *et al.*, "Automatic Preparation of Selected Title Lists for Current Awareness Services and as Annual Summaries," 144th National Meeting of the American Chemical Society, Los Angeles, Calif., April 1963.
- (10) Freeman, R. R., *J. Chem. Doc.*, **4**, 107 (1964).