It can be expected that authors, as they become more aware of the extent to which chemical title-announcement services are used, will title their papers more carefully. Titles have become recognized as an aid in information retrieval by government and industrial information centers.

LITERATURE CITED

(1) Freeman, R. R., Godfrey, J. T., Maizell, R. E., Rice, C. N., Shepard, W. H., *J. Chem. Doc.*, 4, 107 (1964).

Internal Alerting with Keyword-in-Context Indexes*

ANN KLEIN HAAS
Technical Information Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee
Received March 4, 1965

Keyword-in-context (permuted keyword) indexes of documents are being increasingly produced for alerting and retrieval of information. Publications such as *Chemical Titles* and *The Biochemical Titles Index* have done much to establish standards of terminology and format and to assure reader acceptance of internal keyword indexes.

Keyword-in-context indexes achieve a balance among computer and program requirements and the individual needs within an organization. Thus, chemists are alerted to new developments within an organization by keyword indexes that are tailored to specific research programs.

Many companies start by using a keyword-in-context index to cover their own internal reports, and then discover that the same computer program or form of input may be modified to yield bibliographies and rapid-announcement bulletins. Indeed, the production of internal permuted keyword indexes may be part of an integrated program which also yields input for coordinate index systems and book catalogs.

While a permuted keyword index can be produced manually, this method is not economical. If an organization has access to a computer, the availability of several computer programs and the economy of producing permuted keyword indexes make them attractive for local use.

The internal keyword-in-context indexes to be discussed here will illustrate methods used to overcome the inflexibility of a computer program, variations of expressing chemical concepts in titles, and the problems of using computer printouts. In order to alert chemists to new developments, the new information has to be readable, easy to locate, and quickly discovered. Tailor-made internal keyword-in-context indexes achieve effective alerting, provide information retrieval, and produce management statistics by using a variety of formats and computer programs.

* Research sponsored by the U. S. Atomic Energy Commission under contract with the Union Carbide Corporation; presented before the Division of Chemical Literature, 148th National Meeting of the American Chemical Society, Chicago, Ill., Aug. 31, 1964.

GENERAL FORM AND ADVANTAGE OF KEYWORD-IN-CONTEXT INDEXES

Keyword-in-context indexes are indexes of the significant words or keywords in titles or sentences. To prevent entries for nonsignificant words, each word in the title is computer matched against a word list (stored in the computer memory) that contains primarily noninformation-containing words. The number of nonsignificant words used in the "stop list" depends on the specialization of the index. The more specialized indexes have longer lists of nonsignificant words because more technical words have to be included. The Oak Ridge National Laboratory (ORNL) indexes cover all the sciences, use 490 words in permanent storage, and provide space for an additional 150 words in a temporary storage. This allows a variation in vocabulary for the specialized indexes for chemists.

The length and form of the indexing sentence varies in the permuted part of the index, and all the keyword-in-context indexes have the snap-back or wrap-around feature. *Chemical Titles* (*1*) prints a 60-character title line, and fills in any remaining space with words from the beginning or end of the title. This IBM form of the permuted keyword index may be printed by photo-offset methods without reduction; with reduction, two columns per page are possible. One disadvantage of the IBM form for internal use is that it has no flexibility in its nonsignificant word list. The computer program does not provide for deletion or inclusion of a keyword for a single use. The form of the bibliography listing is in the common form of author first followed by title and reference.

The Bell Telephone Laboratories (Bell Laboratories) computer program has room for 105–106 characters of the title per line, and blank spaces are filled with words adjacent to the keyword (*2*). Use of the Bell Laboratories program causes less title omission and loss of keyword phrases (*3*) because of the longer line length, and it provides greater flexiblity in vocabulary control because of the opportunity to include words in a temporary stop list. It is possible to use segments of the output tape

to print specialized portions of the keyword index. The use of this feature in "Indexes to Conferences Published in *Nuclear Science Abstracts*" permits separate indexes for titles and cities, dates, numbered conferences, and report numbers (*4*). All these indexes were generated from the same output tape. The title must appear as the first element of the bibliography listing and the order of author and reference is optional.

Another variety of the keyword index involves printing the keyword at the left of the title entry line, with the *complete* title then printed under it, indented or centered in conventional-index-appearing form. The document identification number is in the usual location at the right of the page. This "keyword-out-of-context" form is used as an aid to readability, and is also usually derived from an IBM form computer program.

Most keyword index programs used for internal reports in various organizations use either the IBM or Bell Laboratories form.

Part of the economy of permuted keyword indexes is due to the fact that the same deck of cards punched in the form of the bibliography part of the indexes can yield a permuted index and several author-like indexes. It is this feature of the index program that makes possible its applications for alerting and information retrieval.

The document identification number may be used to give statistics, code publications, or define broad subjects. The author index may be altered, by adjusting the input, to yield special isotope indexes or report number cross indexes. The format of the bibliography section may be varied to include information necessary for a complete citation, or abstract, or may be eliminated entirely.

Computer printout of permuted keyword indexes may be made more readable by the use of special techniques when preparing the material for reproduction. The use of various sizes of large boldface (*e.g.*, from Vari-Typer-Headliner) for section headings adds to the readability of the indexes. At ORNL the bibliography is reproduced at 65% of the original size and the author index at 75%. Esso Research and Engineering Company, which reproduces its permuted index at 60% of the original size, uses a screen tint to print all the material preceding (to the left of) the keyword in the permuted index on a gray background (*5*), to speed the location and scanning of keywords.

TAILOR-MADE ORNL INDEXES

A general-purpose computer program deck was designed for ORNL indexes that would be suitable for running all the internal indexes. The Bell Laboratories computer program was modified to allow the equivalent of a short abstract to be permuted along with the title of the report (*6*). This feature is used for deeper indexing of progress and trip reports. Since titles of progress reports generally are not descriptive of their contents, words or phrases are added to identify in greater detail the material covered.

Although the research program at ORNL is so broad that almost every scientific field is covered in ORNL publications (about 25% of the included titles are from biology, 30% from chemistry, and 25% from physics), the most avid users of keyword indexes are those scientists who work in chemistry and related fields. The problem exists, then, as to how to alert the chemists to new developments in their fields when the pertinent titles are mixed among the titles from various other scientific fields. This is attempted in several ways: (1) *Chemical Titles*, which the chemists read, is used as the authority for editing chemical terms; (2) the ORNL nonsignificant word list is slanted toward chemistry and related fields; and (3) a special arrangement of chemical terms in the permuted part of the index is achieved with punctuation symbols.

Three types of tailor-made indexes are published at ORNL and illustrate the relationship among program requirements, organization of input information, and the demand for a special-purpose index. "Publications, Reports and Papers for 1963 from Oak Ridge National Laboratory" (*7*) is a yearly compilation of all papers in books, journals, reports, theses, and presented at meetings by ORNL staff members. This index presents several input problems, because titles for its 1920 entries involve three series of file numbers and are collected in random order during the year. Since management uses this annual publication for statistics, all titles have to be identified by type of publication and ORNL division. Repetition of titles is very noticeable in the permuted index, so each title is listed only under the division of the first author. Divisions represented by coauthors when such authors are not in the same division as the first author are indicated at the end of each division listing. Also, since this document receives international distribution and serves as a guide for ordering reprints or for finding an article in the open literature, suitable references are included with each entry.

Figure 1 shows a listing from the bibliography section of the index. The entry identification number is coded so that the first two numbers designate one of 30 ORNL divisions. The hyphens are aids to readability. The letter is a mnemonic designation; B, indicates books; J, journals; P, papers presented at professional meetings; R, documents issued by ORNL or government agencies; and T, theses. The last four numbers are accession numbers from three different files. An author index and a permuted index are included. Putting this annual listing on the computer speeded publication by two months, although the number of entries increased 45%. This index also is used for collecting material for bibliographies because the

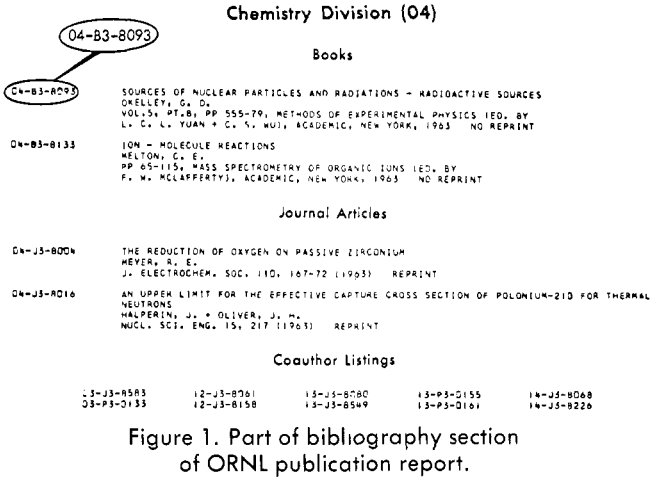

Figure 1. Part of bibliography section of ORNL publication report.

permuted index provides a subject approach to all unclassified ORNL-generated literature.

"Publications on Isotopes at Oak Ridge National Laboratory, 1946–1963" (*8*) identifies the radioisotopes and enriched stable isotopes that were used in research during the past 16 years. This bibliography of about 1700 references is arranged by broad subject category and by accession number. The main purpose of this index is to retrieve data to be used as input for a coordinate index system; consequently, besides the bibliography, permuted index, and author indexes, isotope and report number indexes are needed on punched cards. These cards are sorted and separated from the author cards. The Bell Laboratories computer program yields these three indexes together.

The permuted index program provided the impetus for publishing "Indexes to the Oak Ridge National Laboratory Master Analytical Manual, 1953–1963" (*9*). Since these indexes were to be updated yearly, a system was needed with the program to assure rapid updating of the approximately 550 methods. The complete history of each analytical method, current status, evaluation of all methods, and compatibility of both the published sections of the manual and the unpublished sections used in the ORNL Analytical Chemistry Division are provided by this index. Since each entry in the permuted section refers directly to the method number under which the full method is located, cross-reference indexes are needed. All numbers used to identify the various methods follow the general arrangement of the analytical research and process Laboratories at ORNL, and probably are unfamiliar to people not connected with the Analytical Chemistry Division. The keyword-in-context part of the index eliminates the need for intimate knowledge of the numbering system. Additional words added to the title for better identification are enclosed in spaced parentheses. An entry for the bibliography section of the indexes illustrates how each method is described (Figure 2). The order of entry was planned to permit updating with minimum keypunching of new cards. This index serves as an aid in locating the analytical method according to subject analyzed, property determined, general and specific type of method used, reagent, and instrument employed.

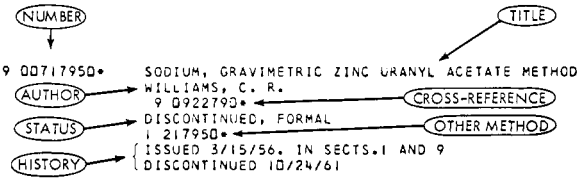

Figure 2. Typical bibliographic entry for analytical index.

Pertinent to alerting is a third type of ORNL tailor-made index produced with the same basic computer program—the "very rapid announcement index." This weekly list of ORNL papers cleared for publication and/or presentation at meetings informs ORNL management and scientific personnel of research months before actual publication. The permuted index part of the computer program is not run; consequently, except for following the *Chemical Titles* rules for writing out the names of all chemical compounds, titles are keypunched without editing. The index is reproduced in two columns per page. This format speeds reproduction, and the spacing allows cutting and pasting on 3 × 5-in. cards for personal files. Total production time is about one day for 25 to 45 items.

COMPUTER PROGRAMS AND ALERTING METHODS USED BY OTHER INSTALLATIONS

Bell Telephone Laboratories used the permuted keyword index to cover their internal technical reports. Permuted keyword indexes also are produced for talks and papers presented by Bell Laboratories personnel, specifications, and bibliographies. A reproduction at 80% of original is used in the bibliography section of "Laboratories Talks and Papers." A blank card which will later list the journal reference is included with the card set of each entry (*10*). The Internal Technical Reports index of approximately 200 reports per month includes a case number, broad subject, and sorted author indexes, and cross references in the permuted index. The production of the Bell Laboratories internal index is machine oriented because, except for following the usual *Chemical Titles* conventions, very little editing is done. Consistency of the keypunched product is checked on the IBM 101.

Esso Research, however, does extensive editing of the titles of 2000 reports per year, and achieves a measure of deep indexing in its internal permuted keyword indexes. Multiple sentences are written for reports, and titles are only used where they describe the contents of the report. An average of three sentences is used per document, and the vocabulary is controlled by hyphenating, inserting spaces to permit generic entries, and use of "see" and "see both" entries. A "Monthly Technical Reports Index" is produced for alerting, then entries are cumulated annually.

Lockheed Aircraft Corporation uses an IBM-form permuted keyword computer program written for the IBM-7094 computer (*11*) and modified to produce its keyword-out-of-context index. The whole title appears in the center of the page, with the second and third lines of the title indented further. Additional words that are added to the title are separted by asterisks from the main portions of the title. The identification numbers and report location symbols are at the right of the page, and the keywords are at the left. The usual nonsignificant word list is used. This form usually replaces the permuted index portion; however, Lockheed is able to eliminate the bibliography part of the index by using this form.

The Ernest O. Lawrence Radiation Laboratory uses an IBM-1401 computer to produce a keyword-out-of-context index (*12*). In the bibliography section of its index, a complete citation is given for each report, and the parts of the entry not to be included in the index are indented. Advantage is taken of the need to write the computer program in several steps to provide a program with several options in format. The identification number and nonsignificant word list may vary in length, and the total length of an entry in the index may be either 100 or 128 characters. The second line of an entry is indented. This format provides a very readable index.

The permuted keyword indexes discussed here were chosen to depict the general state of the art. Both the

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-occurence 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.

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

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