OAK RIDGE NATIONAL LABORATORY
operated by
UNION CARBIDE CORPORATION
for the
U.S. ATOMIC ENERGY COMMISSION

ORNL-TM-525

57

THE ORNL LIBRARY KEY WORD INDEX PROGRAM

Ann S. Klein
Elizabeth B. Howard
Nancy B. Alexander

NOTICE

This document contains information of a preliminary nature and was prepared primarily for internal use at the Oak Ridge National Laboratory. It is subject to revision or correction and therefore does not represent a final report. The information is not to be abstracted, reprinted or otherwise given public dissemination without the approval of the ORNL patent branch, Legal and Information Control Department.

LEGAL NOTICE

This report was prepared as an account of Government sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, expressed or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission, or employee of such contractor, to the extent that such employee or contractor of the Commission, or employee of such contractor prepares, disseminates, or provides access to, any information pursuant to his employment or contract with the Commission, or his employment with such contractor.

ORNL-TM-525

Contract No. W-7405-eng-26

THE ORNL LIBRARY KEY WORD INDEX PROGRAM

TECHNICAL INFORMATION DIVISION
Ann S. Klein and Elizabeth B. Howard

MATHEMATICS DIVISION
Nancy B. Alexander

DATE ISSUED

APR 22 1963

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee
operated by
UNION CARBIDE CORPORATION
for the
U. S. ATOMIC ENERGY COMMISSION

ABSTRACT

The development of a Key Word Index system to index and retrieve reports originating at ORNL is reviewed. Points covered include input and output problems, the nonsignificant word list, machine logic vs. human logic, the importance of good titles, coding, and variation in format.

THE ORNL LIBRARY KEY WORD INDEX PROGRAM

INTRODUCTION

The need for a better system to index and retrieve the information in reports written by ORNL personnel had long been recognized by the Document Collection Section of the Library. Formal ORNL reports are indexed in the AEC journals Nuclear Science Abstracts and Abstracts of Classified Reports, but a time lag of three months or more can occur between the issuance of a report and coverage in these journals. Only a small percentage of the TM reports may be covered, and after the summer of 1961 no CF memorandums were covered since these are treated as internal Company reports. Access to indexes covering Laboratory reports was either too delayed to be effective or nonexistent.

In preparing a program to correct this situation, the Library felt the criteria for an index should be: it would be current, cover all types of reports issued by the Laboratory, and be accessible on the user's desk. No additional personnel would be hired, costs should be kept to a minimum, and consideration should be given to utilizing as much of the existing cataloging system as was feasible.

With these points in mind, Library personnel spent several months, late in 1961, in investigating index systems and writing to and talking with people working on the same problem in other companies. The Key Word Index appeared to meet the criteria without an expensive revision of the present system becoming necessary. Basically, preparation of this index involves taking the key words in a title and, through the use of a high-speed computer, sorting and arranging them for a printout of a quickly scannable alphabetical list. A compatible code is employed to make quick identification from the key word to the original article. In addition, a list giving a bibliographical description and an author index is available (see Fig. 1).

Andrea Crestadoro advocated the principles of this type of index, for use in the British Museum Library, in his book Art of Making Catalogs of Libraries published in 1856. In later years other names were applied, as Permutated Title Index, Rotational Indexing, and Key-Word-in-Context (KWIC).

This type of index has been successfully used in such journals as Biological Abstracts and Chemical Titles. P. H. Luhn at IBM and Robert A. Kennedy at Bell Telephone Laboratories have reported Key Word Indexes in use in their companies.[1,2] A permuted title index to the preprints of the International Conference on Scientific Information held in Washington, D. C., in 1958 was published by Systems Development Corporation, and Stanford University Press used this system in Index to Dissertations in Physics.

Mathematics Division personnel concurred on the feasibility of a Key Word Index, similar to the one used by Bell Telephone Laboratories, using equipment already available at the Central Data Processing Center at ORGDP. A member of the Mathematics Division was assigned to work with Library personnel as programmer and coordinator with the Central Data Processing Center. A letter was written to SHARE, an organization composed of IBM 704, 709/7090 computer users who have banded together in order to interchange information about computer programs developed by each user. The program used by Bell Telephone Laboratories called BEPIP was requested along with the master program which ran the Key Word Index program and supplied the accessory routines.

PRE-MACHINE PROGRAM DECISIONS

Before the program was ready for the computer, certain preliminary decisions had to be made. First, there was the choice of the material to be indexed. Since so many ORNL internal reports received no indexing of any kind and since such a long time interval occured before formal reports were picked up in abstract journals the decision was made for the pilot study to include only Laboratory generated reports. Next, there had to be a choice of the numbering system used to identify reports within the eleven spaces allowed by the program. The decision was made to not use a sequence number for Laboratory internal reports, but simply delete the letters CF from memorandums and ORNL from ORNL-TM's. Revision, addendum, supplement, and volume were written out after the title and indicated at the end of the numbering system. Since the deletion of letters does not give a unique report number, and as the numbering systems used by many installations employ more than eleven spaces, this choice of numbering system will not be satisfactory when the coverage of the index is expanded.

Key Word Index

```
/SONANCES, COULOMB EXCITATION, REACTIONS, CROSS SECTIONS, SPECTRA, FISSION FRAGMENTS  OPTICS - INFRARED, / ORNL-3268
WILCOX COMPANY LYNCHBURG,VIRGINIA TO DISCUSS REPROCESSING SPECTRAL SHIFT CONTROL REACTOR FUEL /ABCOCK-AND- 62-9-30
/-SOLUTION ANALYSIS. REACTOR PROJECTS ANALYSES. X-RAY AND SPECTROCHEMICAL ANALYSES. MASS-SPECTROMETRY OPT/ ORNL-3243
TRON DIFFRACTION AND X-RAY DIFFRACTION  COLORIMETRY  MASS SPECTROMETRY  MOLECULAR BEAM STUDIES /RATES. NEU ORNL-3320
/-27 TO JUNE-16,1962  CONFERENCE ON HIGH-SENSITIVITY MASS SPECTROMETRY AT ALDERMASTON AND CONFERENCE ON P/ 62-7-14
/E-9-TO-AUGUST-6,1962  ADVANCED COURSE ON NEUTRON CRYSTAL SPECTROMETRY, KJELLER,NORWAY AND NATO SUMMER SC/ 62-8-70
 EXTRACTION  CHEMICAL APPLICATIONS OF NUCLEAR EXPLOSIONS SPECTROPHOTOMETRIC STUDIES OF SOLUTIONS /SOLVENT TM- 18
BISMUTH IN MOLTEN BISMUTH-TRICHLORIDE                   SPECTROPHOTOMETRIC STUDY OF DILUTE SOLUTIONS OF  ORNL-3316
/RIUM-242  GAS CHROMATOGRAPHY  MOLTEN FLUORIDE ABSORPTION SPECTROPHOTOMETRY  HYDROFLUORIC-ACID POLAROGRAP/ 62-7-23
/SSIUM-43  PLUTONIUM POLYMERS  MOLTEN-FLUORIDE ABSORPTION SPECTROPHOTOMETRY  URANIUM AND PLUTONIUM COULOM/ 62-4-  7
CLEAR INSTRUMENTATION FOR SCINTILLATION AND SEMICONDUCTOR SPECTROSCOPY                                NU TM-137
 PHYSICS  DEFORMATION OF SOLIDS  METAL SURFACE REACTIONS SPECTROSCOPY OF IONIC MEDIA  X-RAY DIFFRACTION / ORNL-3313
                  PRELIMINARY INVESTIGATION OF A DUAL SPECTRUM MOLTEN SALT REACTOR CSMSR                 62- 1- 40
             AN ESTIMATE OF THE EFFECT OF NEUTRON-ENERGY SPECTRUM ON RADIATION DAMAGE OF STEEL            TM-299
. M/ HYDROGEN FIRED FUSED AND GROUND URANIUM-DIOXIDE FROM SPENCER-CHEMICAL COMPANY  SAMPLES V4, V5, AND V6 62-7-59
/  URANIUM - GRAPHITE FUEL PROCESSING  HEAT-TRANSFER FROM SPENT FUEL-ELEMENTS DURING SHIPPING   ZIRCONIUM/ TM-122
/CN FUEL PROCESSING  MODIFIED ZIRFLEX  HEAT-TRANSFER FROM SPENT FUELS DURING SHIPPING  EVAPORATOR-CALCINE/ TM- 32
/ES  URANIUM-GRAPHITE FUEL PROCESSING  HEAT-TRANSFER FROM SPENT REACTOR-FUELS DURING SHIPPING  SOLVENT EX/ TM-121
/DEJACKETING STUDIES  SHEAR AND LEACH  HEAT-TRANSFER FROM SPENT REACTOR-FUELS DURING SHIPPING  SORPTION O/ TM-112
      PERFORMANCE OF MECHANICAL EQUIPMENT FOR DEJACKETING SPENT SRE CORE-1 FUEL                          TM-367
Y OF PYROLYTIC CARBON COATED AND UNCOATED URANIUM CARBIDE SPHERES                              METALLOGRAPH TM-91
/OF THE VOID FRACTION OF STATIC BEDS OF 1.5-INCH-DIAMETER SPHERES PACKED IN A FULL-SCALE MODEL OF THE PBR/ 62-9-15
/C DISPLACEMENTS  X-RAY DIFFRACTION  ELECTRON MICROSCOPY  SPIN RESONANCE  SUPERCONDUCTIVITY  METAL SURFAC/ ORNL-3364
1961        TRIP REPORT  CONFERENCE ON NUCLEAR ELECTRONICS SPONSORED BY THE IAEA AT BELGRADE MAY-15-TO-19,  62-4-63
                                 SUMMARY OF HFIR HOT SPOT STUDIES                                        62- 1- 52
 RESONANCE INSTABILITY                          EFFECT OF A SPREAD IN CYCLOTRON FREQUENCIES ON THE CYCLOTRON 62-8-37
/NG ATOMS IN SOLIDS AND INTERATOMIC POTENTIALS  THEORY OF SPUTTERING  SEMICONDUCTORS  CRITICAL SIMPLICES / ORNL-3364
                     OR- GLS  A GENERAL FORTRAN LEAST SQUARES PROBLEM                                    TM-271
                OR- FLS, A FORTRAN CRYSTALLOGRAPHIC LEAST SQUARES PROGRAM                                TM-305
                                 A GENERALIZED LEAST SQUARES PROGRAM FOR THE IBM-7090 COMPUTER           ORNL-3259
/ HYDROCLONE STUDIES  KINETICS OF URANIUM ANION EXCHANGE SRE AND ZIRFLEX DEJACKETING  HETS AND HTU VALUE/ TM- 83
OR-8  FLUORINATION CORROSION  PLUTONIUM - ALUMINUM FUEL SRE CORE ONE FUEL  LOW-LEVEL WASTE /ORATORY  IN TM-  4
```

Bibliography

ORNL-3300 PUBLICATIONS, REPORTS, AND PAPERS FOR 1961 FROM ORNL-VOL.I
JUNE 26, 1962

ORNL-3301 PREPARATION OF CHARGE MATERIALS FOR ORNL ELECTROMAGNETIC ISOTOPE SEPARATORS
SHERIDAN, C. W. + GWINN, H. R. + LOVE, L. O.
JULY 19, 1962

ORNL-3302 GAS-COOLED REACTOR PROGRAM QUARTERLY PROGRESS REPORT FOR PERIOD ENDING MARCH-31,1962 EGCR PBRE
URANIUM-CARBIDE PARTICLES BERYLLIUM COMPOUNDS IRRADIATION EFFECTS HEAT TRANSFER
JULY 16, 1962

ORNL-3303 SIZE EFFECTS AND INTERSTITIAL IMPURITIES IN NIOBIUM - ZIRCONIUM(25-PERCENT) SUPERCONDUCTORS
SUPERCONDUCTING SOLENOIDS WITH METAL INSULATION
BETTERTON, J. O., JR. + KNEIP, G. D., JR.
EASTON, D. S. + SCARBROUGH, J. O.
AUGUST 28, 1962

ORNL-3305 A SURVEY AND EVALUATION OF URANIUM-233 FISSION YIELD DATA
FERGUSON, R. L. + OKELLEY, G. D.
JUNE 28, 1962

ORNL-3306 THE ANGULAR DISTRIBUTION OF FISSION FRAGMENTS FROM THE FAST NEUTRON-INDUCED FISSION OF URANIUM-234
LAMPHERE, R. W.
AUGUST 10, 1962

Author Index

TM-252	BLANCO, R. E.	62-9-36	BUCHANAN, J. R.
TM-376	BLANCO, R. E.	TM-106	BUCHANAN, J. R.
TM-377	BLANCO, R. E.	TM-173	BUCHANAN, J. R.
TM-420	BLANCO, R. E.	TM-93	BUCKHAM, J. A.
ORNL-3219	BLANCO, R. E.	TM-190	BUMGARNER, L. L.
ORNL-3293	BLANDER, M.	62-9-15	BUNDY, R. D.
TM- 2	BLANKENSHIP, F. F.	TM-228	BURCH, W. D.
TM- 72	BLEVINS, E. L.	61-12- 50	BURKE, O. W.
ORNL-3336	BLIZARD, E. P.	62-9-9	BURKE, O. W.
62-10-16	BLOCK, R. C.	TM-310	BURNS, J. H.
TM-86	BLOMEKE, J. O.	TM-229	BUSING, W. R.
TM-164	BLOMEKE, J. O.	TM-271	BUSING, W. R.
TM-210	BLOMEKE, J. O.	TM-305	BUSING, W. R.
ORNL-3356	BLOMEKE, J. O.	62- 2- 77	BUXTON, S. R.

Fig. 1. Key Word Index Arrangement.

The Library decided to issue a monthly index with semiannual and annual accumulations. After publication of an annual index, other issues for that year could be discarded, leaving only one source to search for a year's coverage. Besides the Key Word Index, each issue would contain a bibliographical listing by report number for quick identification, and an author index.

Originally a broad classification system was planned for inclusion in the index. This classification system divided all Laboratory reports into subject categories based on Laboratory interests. One card column was to be used to punch the fourteen subject category numbers. These numbers were to be listed in a separate index known as the Case Number Index. The idea was dropped because an extra printed line was necessary to carry this classification number in the bibliography, and economy of space was important in order to have the Key Word Index reproduced at ORNL. Also, this classification number caused numerous keypunching errors. Presently, the classification number is carried in columns outside the program and may be used to develop specialized bibliographies.

INPUT AND OUTPUT

As soon as the Bell Key Word Index computer program had been received from SHARE, changes had to be made so that the master program which runs the local computer coder would also be able to run the BEPIP code. The Bell Lab input/output codes had to be incorporated, along with some of their error detection codes before the Key Word Index could be used on some of the actual data.

Input and output requirements for the Bell program differ from other programs. In the case of input, the keypunchers, in addition to keypunching, are expected to arrange data in the proper form for title, author, and write cards. They sequence these cards for the program. Keypunching errors result when input does not follow a predetermined format. The keypunchers are not expected to catch spelling errors, and with this program a single such error may be multiplied ten times by the computer. Spelling is checked not only before and after keypunching but also after the computer run.

Output also caused problems. Many of the problems, especially those caused by spelling and hyphenating errors, can be edited with sissors. However, too much deletion conflicts with the programming of page length of the

permuted section. The entire output is reduced 40% for reproduction to provide optimal output. If there were no reduction, the index would either have to be turned lengthwise to use, or else the number of words per line would have to be cut in order to fit a standard size sheet.

The desire to reproduce this index resulted in frequent rejection of machine printout. Output with smudged print, extra symbols, dropped and missing letters, and lined paper was not acceptable. (The heavy green line cannot be deleted in reproduction.) The author index caused more problems because this index was listed separately from the rest of the program. The author index is now listed at ORNL on the IBM-870 and takes approximately ten minutes for a monthly index. Although a program is available for reversing the order of the author and report number in the author index, because of the chance of losing tapes and the need to have the whole program re-run, the present format is maintained. Document Collection Section catalog cards, with appropriate markings, were used for the keypunchers because these cards had an acceptable format for keypunching. Production time for the first issue is illuminating. Keypunching and computer time was about $1\frac{1}{2}$ months. Now keypunching requires about a day and a half, and the processing of the program requires about two days of which the IBM-7090 uses 1.5 minutes.

The greatly reduced production time resulted from learning three main points about running the program. These were (1) the computer's logic is more consistent than human logic; (2) methods used to handle mathematical and other types of programs do not provide output acceptable for reproduction; and (3) a simple program, which the Bell program is (as far as the library part is concerned), can be made extremely complex. To keep things simple and let the computer do the work is the hardest rule to follow.

NONSIGNIFICANT WORD LIST

The nonsignificant word list primarily contains noninformation words. It contains English articles such as a, the, of, and similar words that fasten sentences together. Such words are of no value in an index, and are stored in the computer's memory so they will not print out. The list also consists of words of low or no indexing value. The low indexing value of certain semi-technical words results from the nature of the work at the Laboratory. All forms of a word have to be stored because the computer scans and reads the word just as it is written.

The nonsignificant word list is used (1) to eliminate noise in the form of noninformation containing words and (2) as a solution to some of the problems connected with the hyphen. Words may be stored temporarily or permanently in the nonsignificant word list. Words in permanent storage are a regular part of the program instructions, and additions or deletions require the services of a programmer. Temporarily stored words can be added or deleted for each computer run. With both of these types of storage a single word can be deleted for just one entry. The word "plant" most of the time means one that is built, but in three reports last year the "plant" was the type that grew. The word was temporarily deleted from the nonsignificant word list for these reports. "Material," "product," "problems," "research," and "results" are words commonly associated with scientific research and consequently have little meaning as indexing terms. The nonsignificant word list slants both toward chemistry and reactors because a majority of Laboratory reports fall into these fields.

The original Bell Laboratories nonsignificant word list was used for the first run of the program, but plans were made to revise this list as soon as the program was running to better reflect word usage in ORNL reports. The Bell program had 585 permanently stored words and allowed space for twenty words of 6 or fewer characters, twenty words of 7 to 12 characters, and five words of 13 to 18 characters. At present there are 646 words programmed in permanent storage, fifty words of 6 or less characters, seventy-five words of 7 to 12 characters (most of the nonsignificant words are of this length), and twenty-five words of 13 to 18 characters can be added. The enlarged memory space is used for the temporary storage of words until the annual revision of the nonsignificant word list. Also this temporary storage can be used for special subject bibliographies where storage of technical words would be neccesary. Very few technical words are in the stored nonsignificant word list because of the broad subject coverage of Laboratory reports.

ORNL's nonsignificant word list is now based on lists from Bell Laboratories, Lawrence Radiation Laboratory at Livermore, and Chemical Titles. Words that occurred on the three lists and had no searching value in ORNL reports were selected for the ORNL list. The printout from Annual and Semiannual Laboratory Key Word Indexes was also checked and nontechnical words that appeared more than five times were deleted. The present size of the

nonsignificant word list does not affect computer cost, but a four-month count of words deleted by the program more than five times is revealing. There really are not many words that are truly nonessential. "Of" is the word that is listed most often (see Fig. 2).

MACHINE LOGIC VS HUMAN LOGIC

One of the problems of machine logic or program logic is controlling word printout with a dash. The 11 punch is used rather than the 8,4 punch, which causes a percentage sign to be printed. Some interesting CF numbers were invented before this was discovered. Also, in Laboratory reports there is wide use of the hyphen in adjective modifiers; for example, "radiation-induced," "radiation-inhibited." These dashes within the original article cause difficulty because they have to be checked and deleted if necessary. Use of the hyphen also affects the alphabetical arrangement of words in the Key Word Index. Other types of punctuation affect the alphabetical arrangement too; for example, the comma, period, parenthesis, and slash. Figure 3 illustrates alphabetization by the Binary-Coded Decimal (BCD) sort; followed by a blank, uranium followed by the hyphen as in uranium-carbide, and then uranium followed by a hyphen and a number as in uranium-235. This, incidentally, is the way isotopes are handled in the index. The lower part of the figure illustrated the BCD sort with numbers; for example, 316 prints out before 36.

Words that have been hyphenated or punctuated in order to have the computer read them as a single word cannot be subsequently added to the nonsignificant word list without re-keypunching, and this may cause sequencing errors. "Cycle," "element," "elements," "exchange," and "induced" are examples of words that have given trouble and are included in the nonsignificant word list in order to avoid hyphenating them. Too often the hyphen is forgotten.

Machine logic was defined when too many words were added after the titles. At first, meetings, symposia, and proceedings were grouped under the word "conference," which was added at the end of the title; literature searches and surveys were grouped under the word "bibliography," which was also added at the end of the titles. As additional words were added to titles of progress reports and trip reports, the part of the memory devoted to titles continued to fill up. Overloading continued until part of the memory devoted to program instructions was wiped out.

51	A	8	EXPERIMENTAL	28	PROCESSING
18	AN	13	FACILITY	57	PROGRAM
189	AND	7	FACILITIES	130	REPORT
8	AS	6	FAST	9	RESEARCH
29	AT	154	FOR	31	STUDIES
6	BETWEEN	30	FROM	10	STUDY
31	BY	78	IN	7	SYSTEM
6	DETERMINATION	19	LABORATORY	11	TESTS
30	DEVELOPMENT	11	MODIFIED	159	THE
46	DIVISION	20	NUMBER	12	THROUGH
10	EFFECT	345	OF	46	TO
15	ENDING	51	ON	9	TREATMENT
19	EVALUATION	25	PERIOD	11	USE
8	EXPERIMENT	29	PROCESS	25	WITH

Fig. 2. Occurrence of Nonsignificant Words in Key Word Index. (Four Months Count.)

```
RING YIELD FCR (111) - ALUMINUM PLANE
C PHENOMENA CN (111) - CCPPER SURFACES
E, AND OPTICAL ABSCRPTICN OF THE E-PRI
 SEPARATION OF AMERICIUM FROM OTHER EL
T AUGUST,1962  AMERICIUM-241  CURIUM-2
TWO            ANALYSES OF HEAT TRANSF
PROCEDURES AND ANALYSES OF SOL-GEL BAT
WATER AND MILK ANALYSES, CLINCH AND TE
ROGRAM FOR THE ANALYSIS OF SIMULATED M
OREIGN TRAVEL  ANALYTICAL CHEMISTRY PA
               ANALYTICAL-CHEMISTRY-DI
VER REPORT FOR AUGUST AND SEPTEMBER,19
SAL REPORT FOR AUGUST,1962
ROGRESS REPORT AUGUST,1962  AMERICIUM-
E/ TRIP REPORT AUGUST-2-TO-AUGUST-10,1
UROPEAN TRAVEL AUGUST-20-TO-25,1962 IN
 JAPANESE TRIP AUGUST-26-TO-SEPTEMBER-
 JAPANESE TRIP AUGUST-30-TO-SEPTEMBER-
 PERIOD ENDING AUGUST-31,1962
 PERIOD ENDING AUGUST-31,1962  PENETRA
E MEETING HELD AUGUST-8,1962 - REVIEW
ANIUM-CARBIDE  URANIUM - OXYGEN SYSTEM
ANIUM IN TOTAL URANIUM AND THORIUM ME
NS OF ENRICHED URANIUM AND URANIUM-235
JR RECOVERY OF URANIUM BY SOLVENT EXTR
A  THORIUM AND URANIUM EXTRACTION  COR
FULLY ENRICHED URANIUM IN TOTAL URANIU
               URANIUM RECOVERY FROM K
-CARBON COATED URANIUM-CARBIDE  URANIU
 HYDROLYSIS OF URANIUM-CARBIDES  PROTA
  REACTIONS OF URANIUM-CARBIDES AND TH
RED AND GROUND URANIUM-DIOXIDE SAMPLES
RSION HARDING  URANIUM-MONOCARBIDE LIQ
JT-CELL TESTS  URANIUM-THORIUM FUELS
ED URANIUM AND URANIUM-235 - THORIUM-2
STATIC BEDS OF 1.5-INCH-DIAMETER SPHERI
 REPORT NUMBER 14 OF ATR FUEL ELEMENT I
STRY-DIVISIONS 150-KV COCKCROFT-WALTON
JAL REPORT FOR 1961  ENVIRONMENTAL MON
POTASSIUM TYPE 316 STAINLESS STEEL COM
 REPORT NUMBER 36  PYROLYTIC-CARBON CO
SUREMENTS WITH 4.75-INCH-LONG MIDDLE S
 REPORT NUMBER 437 / FOLLOWING RUN-25
 REPORT NUMBER 439 /D, SINTERED AND GR
ORIA - URANIA, 6-PERCENT FULLY ENRICHE
```

Fig. 3. Alphabetization by the BCD Sort.

This incident resulted in re-programming for 800 characters to allow ample length for handling the title. Originally the program provided 400 characters. These extra characters provide for fuller indexing of progress reports, enough space being available for a brief abstract to be included.

COMPUTER AND PROGRAM EXPERIENCE

Perhaps the most significant misunderstanding resulted from the lack of specifications as to the limits of the program -- maximum length of title, number of cards per title, number of authors per document. When mistakes occurred, it was sometimes difficult to determine whether a card punching error had caused the fault or whether one of these unspecified limits had been exceeded.

As these limits were found, more programming changes were involved. Such changes can be very intricate, and tampering with codes written by others is a risky business. The exact locations of the machine instructions involved must be tracked down, every reference to these instructions must be found, and programming changes must be made to incorporate the new specification. These programming changes must fit into the available space within the computer and must affect only the part of the instructions to be modified. The above may not mean anything to anyone who is unfamiliar with computers not to people who have used computers for small programs or even large FORTRAN computer programs. This Key Word Index program is very sophisticated in its use of all available machine capabilities and is written in basic machine language. The sheer magnitude of the number of instructions involved would stagger most programmers. All but about five thousand locations of the thirty-two thousand available in the computer are used. Also, six magnetic tapes are used for scratch work and much reading in and out on these tapes is done to handle all the data which can be processed.

Modifications in the program were also made to change the nonsignificant word list. Rather than requiring programming skill, this involves an exact and tedious job. Two large boxes of cards must be checked for the elimination or addition of these words.

Besides incorporating changes a routine had to be established for card and paper handling. The likelihood of errors increases proportionally to the number of people handling the punch cards and papers. At least seven people now handle the data after the cards leave the Library.

IMPORTANCE OF GOOD TITLES

How valuable the Key Word Index is depends entirely on the information that is put into the computer. The Key Word Index can be no better than the titles that are indexed. A suitable title conveys the information contained in the original paper and is written in clear, concise English. Permuting a title will not alter ambiguity. Although being brief cuts costs, modifying phrases should be included, because the ability to have subject words in context is the basis of the success of the Key Word Index as a subject finding tool. Words should not be dropped from a title by an author simply because they are on the nonsignificant word list. The program will delete these words. Foreign language titles are translated into English, and the words "in French" or any other language are written after the title. (Foreign language equivalents of all nonsignificant words cannot be stored.)

There are certain things to avoid in writing titles, such as using a variation of words to describe the same process. This scatters a concept throughout the alphabetical listing. Newspaper language, trade journal style, and proper names should be avoided as they rarely convey meaningful subject content and often misrepresent the contents. Titles that end in a question mark should be avoided.

CODING

Titles are coded for the computer when the original words are written in non-computer or non-program acceptable language. This may mean that certain terms in the title cannot be key punched or represented on a printer. Superscripts, subscripts, the lower-case letters of chemical symbols, Greek letters, and the degree symbol are examples. Also the BCD sort used by the machine program may confuse retrieval of certain words in the Key Word Index. Chemical formulas would list at the end of a letter and follow the BCD number sort. A group of numbers such as are used to represent crystal levels would be buried among all the numbers listed at the end of the permuted index. This program follows the usual BCD alpha-numeric sort:

blank .) - , (+ A to R / S to Z 0 to 9

Consequently, hyphenated words can cause problems with coding titles, because a hyphenated word will list after the non-hyphenated word and can cause the same

term to be listed in two alphabets. (A reversal of this complication occurs when compounded words only list under the first term.)

Coding in titles is performed to eliminate the above problems and to establish consistent forms for scientific terms and formulas. The Library asked specialists in the fields concerned to develop meaningful formats for expressing scientific concepts. Chemical Titles is used as the standard for chemistry. The Library maintains a complete file of these forms and is grateful for continuing assistance of Laboratory personnel in developing this file.

Example of coding used in titles are shown in Figure 4. The (1) designates a space for the key puncher and causes the word following the space to be indexed. In chemistry, Roman numerals are used to denote metallic valence; a space, hyphen, space is used to show phase relationships; and a slash mark is used for an apostrophe. Chemical elements and Greek letters are spelled out, and in the case of sulfuric acid and beta rays, retrieval is by sulfuric and beta only. Note the splitting of words for indexing purposes.

In the physics and solid states field (Fig. 5), the BCD sort of the parentheses is used to list crystal planes and bombardment at the front of the permuted index before the alphabet. Atomic fractions are used to define intermetallic compounds.

VARIATIONS IN FORMAT

Program adaptation of the Key Word Index permits variations in input and output. There are several ways of numbering items within the Key Word Index. Figure 6 shows the Solid State Division Annual Progress Report numbering system. This system separates items in the body of the report from division publications. Figure 7 illustrates the numbering system for publications, reports, and papers from ORNL for 1962. B is for book, J is for journal, P is for paper presented at a meeting, R is for report, and T is for thesis. The first two numbers represent division. The 2 represents both the second volume of ORNL-3300 and the 2 of 1962. A sequence numbering system is used because the reports appear in random order. The author index also may be varied and used as a chemical compound index. The write card may be used for any type of information that does not have to be permuted, and consequently coding is not required, but brevity is suggested. Cross references are available, and the "see" and the "see both" forms are used. Every word except "see" and "see both" (which are in the nonsignificant word list) are permuted.

CODED	ORIGINALLY
8- (1) AZAGUANINE	8-AZAGUANINE
URANIUM (1) MONO (1) CARBIDE	UC
NICKEL(II) (1) OXIDE	NiO
BARIUM(II)	Ba^{2+}
ORTHO, ORTHO/ (1) DI (1) ALKYL	O, O'-DIALKYL
SULFURIC-ACID	H_2SO_4
BETA-RAYS	β RAYS
COPPER(II) (1) OXIDE (1) - (1) SULFUR (1) TRI (1) OXIDE (1) - (1) DEUTERIUM (1) OXIDE	$CuO-SO_3-D_2O$

Fig. 4. Chemistry Coding.

CODED	ORIGINALLY
SILICON-28 (1) (ALPHA, ALPHA) (1) SILICON-28	$Si^{28}(\alpha,\alpha)Si^{28}$
LITHIUM-7 (1) (HELIUM-3, N) (1) BORON-9	$Li^7(He^3,n)B^9$
THALLIUM-ACTIVATED (1) CESIUM (1) IODIDE	CsI(Tl)
GOLD(0.82) (1) - (1) MERCURY(0.18)	$Au_{0.82}Hg_{0.18}$
MAGNESIUM(0.67) (1) - (1) TIN(0.33)	Mg_2Sn
(III) (1) COPPER (1) SURFACES	

Fig. 5. Physics and Solid State Coding.

There are additional programming features available. The bibliography section may be printed separately and used for rapid announcement lists. A short line length is available for the bibliography, which results in larger printout when the program is reproduced.

The Library offers assistance in organizing and coding bibliographies and files, and in maintaining input for the computer. This is considered part of library service, so there is no code charge involved.

Within the next year the Library plans to have all keypunching done in the Library by Library personnel. This has the advantages of allowing access to equipment at all times using staff knowledge of library materials and methods, and eliminating verifying.

ADVANTAGES

1. Report titles are well endowed with significant words of indexing and retrieving value. Addition of words to a title results in increased information.
2. The most current words of a scientist's field may be found.
3. Once produced, the index may easily be updated and a continuous accumulation accomplished.
4. After the program is developed and running smoothly, the operation becomes an essentially clerical routine one.
5. Promptness of issuing a current index as well as a multi-aspect search system are provided.
6. A copy in book form can be made available on the user's desk.
7. It can match a large number of documents to a large number of readers in a short time.
8. Cross references from one term to another are available. This will take care of synonyms and variations in word usage.
9. Equipment and part numbers, test runs, and other numerical data can be easily identified.
10. By-product result, such as bibliographies.

DISADVANTAGES AND CRITICISMS

1. Efficiency is dependent on the choice of words an author uses in the title to characterize his work.

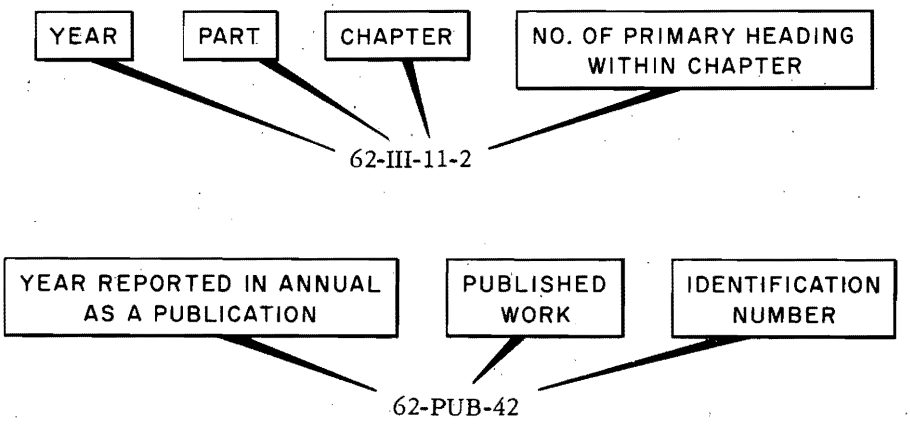

Fig. 6. Numbering System for Solid State Division Annual Report for Period Ending August 31, 1962.

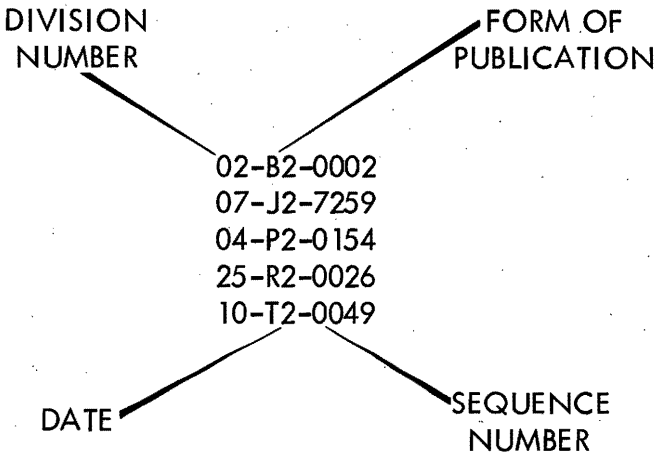

Fig. 7. Numbering System for Publications, Reports, and Papers for 1962 from Oak Ridge National Laboratory.

2. Concept scatter can occur as author's words are used instead of fitting concepts under predetermined subject headings.
3. There are space limitations in the identification code.
4. A new scanning format must be learned by the user.
5. This type of index will not take the place of deep indexing performed by human intellect.
6. Redundancy is inherent in this type of index.

CONCLUSIONS

After a year's experience with the Key Word Index, the Library considers the system usable, practical, efficient, and economical. The question of whether a machine generated product could serve as an adequate and useful index and as a subject approach to ORNL reports has certainly been answered in the affirmative.

The improvement in meaningful titles has begun to be noticeable. This is a result of talking with authors and getting them to become cognizant of the need for good titles. The value of an index, especially a mechanical index, is directly proportional to an author's skill and judgement in writing titles.

Other applications of a key word index have proved feasible. An index was prepared for one division's annual progress report, and work is nearing completion on a publication covering the published papers presented at meetings, journal articles, and reports originating at ORNL in 1962.

Other uses for this type of index will be tested. Expansion of report coverage and a bibliography are under consideration. As more experience is gained, more personnel trained in mechanized library methods, and more equipment is acquired, the Library will make available additional services and better coverage of the literature.

REFERENCES

1. Luhn, H. P., Keyword-in-Context Index for Technical Literature (KWIC Index). IBM Advanced Systems Development Division, Yorktown Heights, N. Y., Report RC-127 (August, 1959).

2. Kennedy, R. A., "Mechanized Title Word Indexing of Internal Reports." Bell Telephone Laboratories, Inc., Murray Hill, N. J. Presented to the Third Institute on Information Storage and Retrieval, The American University, Washington, D. C., February 13-17, 1961.

DISTRIBUTION

1-5. Nancy B. Alexander
6. P. S. Baker
7. D. S. Billington
8. C. A. Burchsted
9. H. P. Carter
10. W. B. Cottrell
11. J. H. Crawford
12. R. R. Dickison
13. A. C. Downing
14. Martha W. Gerrard
15. H. W. Hoffman
16. A. S. Householder
17-21. Elizabeth B. Howard
22. W. H. Jordan
23. C. P. Keim
24. F. Kertesz
25. R. F. Kimball
26-30. Ann S. Klein
31. R. H. Lafferty, Jr.
32. R. P. Milford
33. A. J. Miller
34. F. L. Miller (Y-12)
35. E. J. Murphy
36. J. S. Olson
37. R. B. Parker
38. S. K. Penny
39. Helen P. Raaen
40. M. E. Ramsey
41. J. A. Swartout
42. A. M. Weinberg
43. K-25 Central Library (K-25)
44-45. Central Research Library
46. Document Reference Section
47-56. Laboratory Records
57. Laboratory Records (R.C.)
58-72. Division of Technical Information Extension (DTIE)
73. Research and Development Division (ORO)