11 Keyword-in-Context Index for Technical Literature (KWIC Index)

H. P. Luhn

1. Introduction

Specialized indexes to technical literature are an established means for directing engineers and scientists to sources of information pertinent to their current interest. Whatever the specific purpose of an index may be, a substantial amount of intellectual effort is required to compile it. In many cases, the time presently required for compiling and updating an index interferes seriously with its usefulness at the instant of publication. This is particularly true of bibliographical indexes to material currently being published in such media as technical journals, magazines or technical governmental, institutional and private industry reports.

The accelerated pace of scientific developments in recent years has accentuated the perishable nature of new information. As a result there is a pressing demand for speedier communication in this area. It appears doubtful that this demand can be satisfied without breaking with some of the standards conventionally applied to the compilation of literature indexes.

In what follows the relationship between user and index is examined, and it is shown that for new information, which as it appears is only a fraction of the total information accumulated in an area, relatively rough clues can answer the user's needs. It is then argued that such clues can be generated entirely by machine in the form of a series of extractions each containing a significant, or key, word as its nucleus. Samples of indexes compiled entirely by machine methods are presented in support of this argument.

Note: This work was performed in the Advanced Systems Development Division, International Business Machines Corporation, Yorktown Heights, New York. Originally released as ASDD Report RC-127, August 31, 1959, and delivered before the Division of Chemical Literature at the 136th Meeting of the American Chemical Society, this paper is reprinted with the permission of IBM: a section on "A Derived Code for the Identification of Bibliographical Items" is omitted here.

2. Dissemination v. Retrieval

In the area of communication served by technical literature, the two main functions being performed are the dissemination of information on the one side and the retrieval of information on the other. A publication, when issued, serves to broadcast new information. After the publication has fulfilled this purpose and has been retired to the Library and properly stored, it serves as a potential reference in the process of information retrieval. In the first case its news aspect is predominant, while in the second its historical aspect is predominant.

It is here argued that by means of a rather few clues an expert can judge whether an article touches upon his field of interest and adjust himself momentarily to whatever new information may be furnished. In the case of information retrieval the same expert expects that the information furnished be adjusted to him, that is, to his rather specific interest at the moment.

Because of the difference in attitude in these two cases it is here proposed to consider two types of indexes, namely a dissemination index and a retrieval index, each serving its respective functions and being different as to scope and form. In accordance with this concept a dissemination index would be an instrument prepared with minimum effort and disseminated in the shortest possible time. As such it would fulfill the important task of prompt notification, and its usefulness would be substantially of temporary character. For this reason its publication by inexpensive printing methods would appear justifiable and adequate. A retrieval index, on the other hand, would be an instrument prepared with care in due course, incorporating all those features which will enhance its usefulness as a permanent tool of reference. Most likely it would take the form of a cumulative index and would obsolete dissemination indexes previously issued for material covered by it.

3. Indexing by Means of Keywords in Context

The usefulness of an index depends on the manner in which index entries have been organized. The establishment of categories by subject or other appropriate characteristics is the conventional means by which such organization is accomplished. The establishment of categories and the assignment to such categories of index entries is a matter of judgment and experience and constitutes a considerable part of the intellectual effort involved in the manual compilation of indexes. Various indexers will usually differ in their approaches to this task and will also differ in their interpretation of the material to be indexed. While there may be differences of opinion as to the effectiveness of this or that scheme, the important fact seems to be that any reasonable scheme of ordering, if understood, will save time in locating desired information.

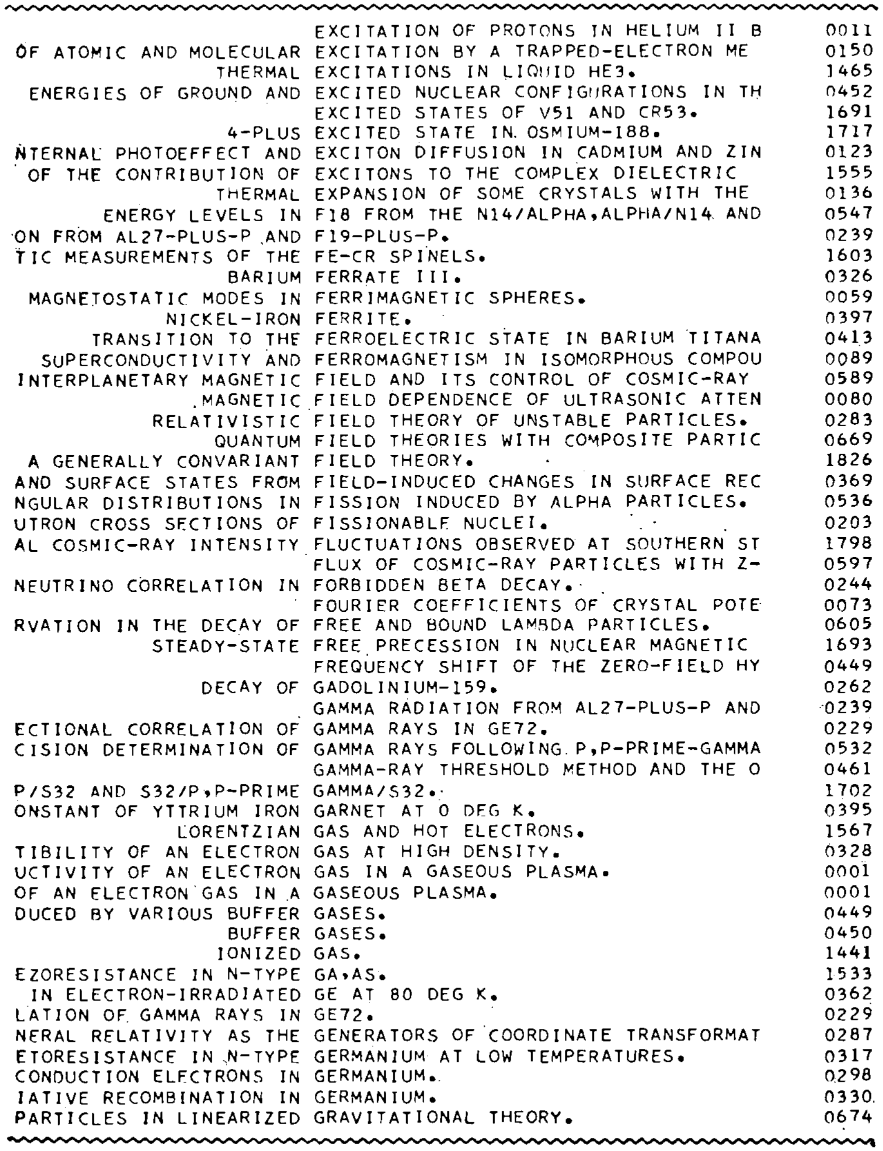
In striving for a speedy method of organizing an index, the question arises as to which of various possible schemes is adaptable to fully automatic processing. Clearly, some means of ordering is required that is based on criteria extracted from the text itself rather than assigned in accordance with human judgment.

The simplest format of a quickly assembled index might be an alphabetic listing of keywords, very much as in the index to a book. The simplicity of such an index is, however, predicated on the fact that the reader has been introduced to the subject matter treated by the book. In dealing with a variety of subjects, as would be the case in the problem under discussion, the significance of such single keywords could, in most instances, be determined only by referring to the statement from which the keyword had been chosen. This somewhat tedious procedure may be alleviated to a significant degree by listing selected keywords together with surrounding words that act as modifiers pointing up the more specific sense in which a keyword has been applied. This method of indexing words is well established in the process of compiling concordances of important works of literature of the past. The added degree of information conveyed by such keyword-in-context indexes, or "KWIC Indexes" for short, can readily be provided by automatic processing.

Keyword-in-context indexing may be carried out on various levels, depending on the purpose an index is to serve. The process may be applied to the title of an article, its abstract or its entire text. Keywords need only be defined as those which characterize a subject more than others. To derive them, rules have to be established for differentiating between what is significant and nonsignificant. Since significance is difficult to predict, it is more practical to isolate it by rejecting all obviously non-significant or "common" words, with the risk of admitting certain words of questionable status. Such words may subsequently be eliminated or tolerated as so much "noise." A list of nonsignificant words would include articles, conjunctions, prepositions, auxiliary verbs, certain adjectives and words such as "report," "analysis," "theory," and the like. It would become the task of an editor to extend this list as required. The remaining significant or "key" words would be extracted from the text together with a certain number of words that precede and follow them. By making the keywords assume a fixed position within the extracted portions and by arranging these portions in alphabetic order of the keywords, the KWIC Index is generated.

The format of a KWIC Index is illustrated in Fig. 1. The initial letters of the alphabetized keywords form a column which guides the eye when scanning for desired words. The number to the right of each line identifies the corresponding document. The sample shown in Fig. 1 was derived from

Keyword-in-Context Bibliographical Index



```
                        EXCITATION OF PROTONS IN HELIUM II B    0011
OF ATOMIC AND MOLECULAR EXCITATION BY A TRAPPED-ELECTRON ME     0150
                THERMAL EXCITATIONS IN LIQUID HE3.              1465
 ENERGIES OF GROUND AND EXCITED NUCLEAR CONFIGURATIONS IN TH    0452
                        EXCITED STATES OF V51 AND CR53.         1691
                 4-PLUS EXCITED STATE IN OSMIUM-188.            1717
NTERNAL PHOTOEFFECT AND EXCITON DIFFUSION IN CADMIUM AND ZIN    0123
 OF THE CONTRIBUTION OF EXCITONS TO THE COMPLEX DIELECTRIC      1555
                THERMAL EXPANSION OF SOME CRYSTALS WITH THE     0136
       ENERGY LEVELS IN F18 FROM THE N14/ALPHA,ALPHA/N14 AND    0547
ON FROM AL27-PLUS-P AND F19-PLUS-P.                             0239
TIC MEASUREMENTS OF THE FE-CR SPINELS.                          1603
                 BARIUM FERRATE III.                            0326
 MAGNETOSTATIC MODES IN FERRIMAGNETIC SPHERES.                  0059
            NICKEL-IRON FERRITE.                                0397
      TRANSITION TO THE FERROELECTRIC STATE IN BARIUM TITANA    0413
  SUPERCONDUCTIVITY AND FERROMAGNETISM IN ISOMORPHOUS COMPOU    0089
INTERPLANETARY MAGNETIC FIELD AND ITS CONTROL OF COSMIC-RAY     0589
               MAGNETIC FIELD DEPENDENCE OF ULTRASONIC ATTEN    0080
           RELATIVISTIC FIELD THEORY OF UNSTABLE PARTICLES.     0283
                QUANTUM FIELD THEORIES WITH COMPOSITE PARTIC    0669
 A GENERALLY CONVARIANT FIELD THEORY.                           1826
AND SURFACE STATES FROM FIELD-INDUCED CHANGES IN SURFACE REC    0369
NGULAR DISTRIBUTIONS IN FISSION INDUCED BY ALPHA PARTICLES.     0536
UTRON CROSS SECTIONS OF FISSIONABLE NUCLEI.                     0203
AL COSMIC-RAY INTENSITY FLUCTUATIONS OBSERVED AT SOUTHERN ST    1798
                        FLUX OF COSMIC-RAY PARTICLES WITH Z-    0597
NEUTRINO CORRELATION IN FORBIDDEN BETA DECAY.                   0244
                        FOURIER COEFFICIENTS OF CRYSTAL POTE    0073
RVATION IN THE DECAY OF FREE AND BOUND LAMBDA PARTICLES.        0605
           STEADY-STATE FREE PRECESSION IN NUCLEAR MAGNETIC     1693
                        FREQUENCY SHIFT OF THE ZERO-FIELD HY    0449
               DECAY OF GADOLINIUM-159.                         0262
                        GAMMA RADIATION FROM AL27-PLUS-P AND    0239
ECTIONAL CORRELATION OF GAMMA RAYS IN GE72.                     0229
CISION DETERMINATION OF GAMMA RAYS FOLLOWING P,P-PRIME-GAMMA    0532
                        GAMMA-RAY THRESHOLD METHOD AND THE O    0461
P/S32 AND S32/P,P-PRIME GAMMA/S32.                              1702
ONSTANT OF YTTRIUM IRON GARNET AT 0 DEG K.                      0395
             LORENTZIAN GAS AND HOT ELECTRONS.                  1567
TIBILITY OF AN ELECTRON GAS AT HIGH DENSITY.                    0328
UCTIVITY OF AN ELECTRON GAS IN A GASEOUS PLASMA.                0001
OF AN ELECTRON GAS IN A GASEOUS PLASMA.                         0001
DUCED BY VARIOUS BUFFER GASES.                                  0449
                 BUFFER GASES.                                  0450
                IONIZED GAS.                                    1441
EZORESISTANCE IN N-TYPE GA,AS.                                  1533
 IN ELECTRON-IRRADIATED GE AT 80 DEG K.                         0362
LATION OF GAMMA RAYS IN GE72.                                   0229
NERAL RELATIVITY AS THE GENERATORS OF COORDINATE TRANSFORMAT    0287
ETORESISTANCE IN N-TYPE GERMANIUM AT LOW TEMPERATURES.          0317
CONDUCTION ELECTRONS IN GERMANIUM.                              0298
IATIVE RECOMBINATION IN GERMANIUM.                              0330
PARTICLES IN LINEARIZED GRAVITATIONAL THEORY.                   0674
```

Fig. 1

titles of technical papers. Since a title may contain several keywords there would be index entries in as many places as there are keywords. For instance, on the sample page the concept "Gamma Rays in Ge 72," will be found under "Gamma" and under "Ge."

A maximum of sixty characters of a title are printed to serve as the index entry. This provides for an adequate number of letters on either side of the centrally located keyword for including immediately associated significant words. The process of slicing a fixed number of letters out of a title necessitates mutilations of some words on either end of the resulting fragment.

4. Organization of a Bibliographical KWIC Index

As is evident from the preceding explanation, the grouping of a given set of bibliographical items into subject categories is eliminated and is replaced by a grouping according to keywords. This arrangement overcomes all arguments as to the appropriateness of assignment of certain items to pre-established subject headings and abolishes the nondescript category of "Miscellaneous." If the index is based on titles of documents, its quality depends on how well the authors have composed the titles of their papers. It will be a matter of experience as to whether KWIC indexing needs to be extended to include abstracts or even portions of the text in order to provide the degree of resolution required under given circumstances.

One of the problems a user of a KWIC Index faces is that of synonyms and variations in word usage and spelling. It must however be assumed that the expert in his field is sufficiently familiar with such variations and is resourceful enough to overcome this problem, as he had to in the past. It is of course quite simple to insert at appropriate places of the index a "see also" cross reference to take care of the less obvious instances. This convenience does not call for additional intellectual effort on the part of the editors once the need for such a reference has been established. Thereafter the insertion of such references will be provided automatically by the machine.

The type of bibliography here proposed would necessarily consist of two parts: a listing of the bibliographical items and the KWIC Index. The items would be listed in alphabetical order of the authors' names and comprise author, title and source data. This list would thus serve as an author index.

Since each KWIC Index entry must be related to the bibliographical items it stands for, there arises a problem of identification. A simple means of identification would be the use of consecutive reference numbers assigned to the bibliographical items in sequence as listed alphabetically by author. These numbers would be given after each index entry (see Fig. 1) and would refer the user to the corresponding item in the bibliography. Such reference numbers are limited to the function just mentioned and would serve no useful purpose outside of the individual bibliography to which they have been applied.

One of the principal advantages resulting from the type indexes here proposed is the promptness, owing to their machine origin, with which they can be disseminated. It would therefore become feasible to issue KWIC Indexes at frequent intervals, perhaps monthly. While this would fulfill the demand for currency, the subscriber of such a service would, however, soon be inconvenienced by having to handle a multiplicity of individual issues. To facilitate bibliographical search of material from the time it is published until it is noted in some more refined reference manual, it would be most useful if the KWIC Indexes were furnished in cumulative form over certain periods. Since they are to be produced automatically, the effort and cost for providing this extra convenience is quite moderate.

5. Automatic Preparation of KWIC Indexes

The various steps involved in the automatic preparation of KWIC Indexes for technical literature will be described briefly and without tying them to any particular type of information processing equipment, except by way of example.

5.1. Creation of Machine-Readable Record

Automatic processing requires that information be available in machine-readable form. Although print-reading devices might eventually translate printed characters into machinable codes, there are today many instances of machine-readable records being produced as a by-product of typing and typesetting operations. These are available in the form of punched tapes or cards and can readily serve as input to present information processing equipment.

In the case of technical literature, the typesetting of many professional journals and of technical magazines is done on punched-tape controlled Monotype or Teletypesetter equipment. Flexowriters are often used for preparing technical reports in order to produce a punched tape for various subsequent retyping operations. In these instances no further manual operations are required to obtain the input for automatic processing.

Where no such records are available, they must be prepared by hand. A most convenient method entails the preparation of punched cards by manual key-punching from the printed text of the portions needed for the process. These portions are the author, title and source of a document if the KWIC Index is to be derived from titles only. Otherwise the abstract or even the text would have to be hand-punched.

Limiting the description to the use of titles only, the punching of cards would best be performed in accordance with certain rules which will facilitate machine processing not only for the creation of the KWIC Index but the creation of many other useful records for facilitating various tasks of publishers, information centers, documentalists, and librarians.

These rules would standardize the format of cards and the manner in which information is to be recorded. For instance, it might be advantageous to prepare a separate card for each author and one or several cards each for the title and the source. The arrangement would be such that a listing of these cards by automatic printing devices would produce a bibliography of good appearance. Furthermore the standardization of these card records will simplify the programming of information processing equipment for performing the routines necessary for deriving identification codes and for extracting the index entries. As was mentioned before, the selection of keywords might best be carried out by rejecting insignificant words of the kind previously described. A dictionary of such words must therefore be compiled and revised in machine-readable form so that it may be transferred to the memory of the machine for reference during processing.

5.2. Machine Processing

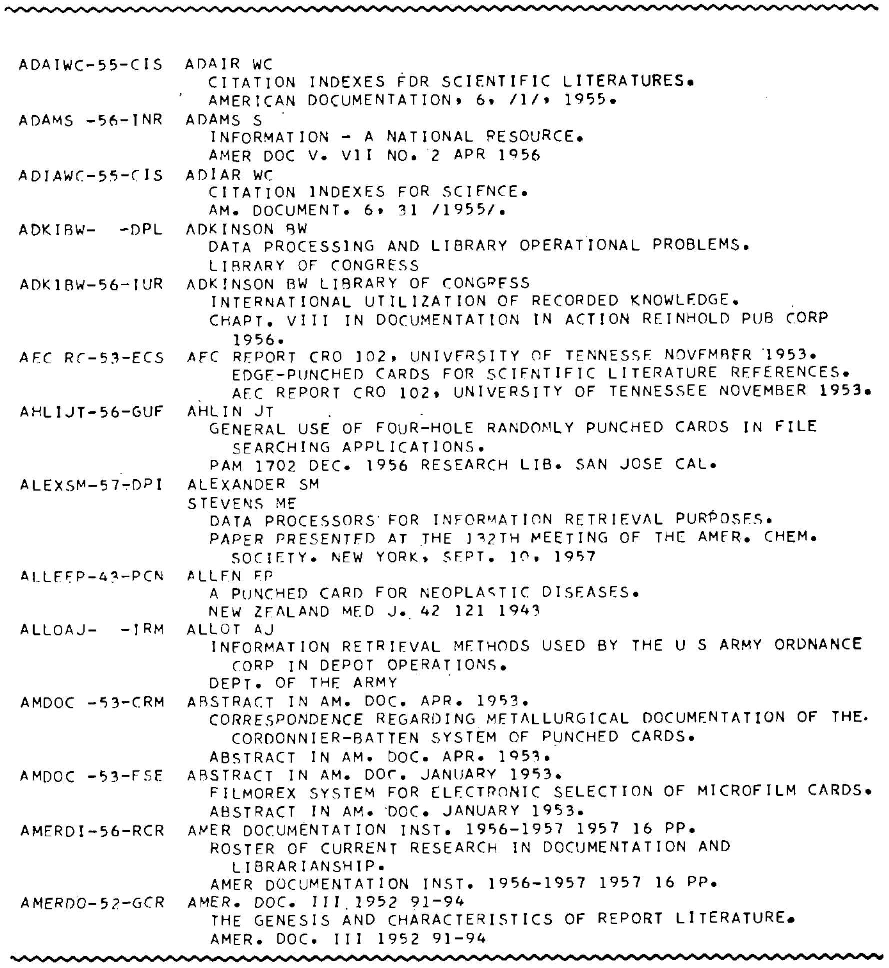
There is no intention here to go into the details of programming information-processing equipment, particularly since many different types of machines may be used to obtain similar effects. Basically, the following major functions need to be performed on each record fed into the machine.

First an identification code is derived. Each word of the title is then looked up in the dictionary of insignificant words stored in the machine. For each word not contained in the dictionary an index entry is generated by shifting the text of the title so that the word in question will start at position twenty-five of a sixty-position field. The contents of this field is then stored together with the identifying code.

After this process has been repeated for each of the documents which are to constitute the bibliography, the records are sorted in the alphabetic order of their identification code and are printed out in the form shown in Fig. 2. The index entries are then sorted in the alphabetic order of the keywords and are printed out in a form similar to that shown in Fig. 3 with their identification codes at the right. Figs. 2 and 3* are typical pages of an index.

* From *Bibliography and Auto-Index, Literature on Information Retrieval and Machine Translation,* Service Bureau Corporation, New York. (Second Edition, June 1959; First Edition, September 1958).

Bibliography



```
ADAIWC-55-CIS  ADAIR WC
                 CITATION INDEXES FOR SCIENTIFIC LITERATURES.
                 AMERICAN DOCUMENTATION, 6, /1/, 1955.
ADAMS -56-INR  ADAMS S
                 INFORMATION - A NATIONAL RESOURCE.
                 AMER DOC V. VII NO. 2 APR 1956
ADIAWC-55-CIS  ADIAR WC
                 CITATION INDEXES FOR SCIENCE.
                 AM. DOCUMENT. 6, 31 /1955/.
ADKIBW-  -DPL  ADKINSON BW
                 DATA PROCESSING AND LIBRARY OPERATIONAL PROBLEMS.
                 LIBRARY OF CONGRESS
ADKIBW-56-IUR  ADKINSON BW LIBRARY OF CONGRESS
                 INTERNATIONAL UTILIZATION OF RECORDED KNOWLEDGE.
                 CHAPT. VIII IN DOCUMENTATION IN ACTION REINHOLD PUB CORP
                   1956.
AEC RC-53-ECS  AEC REPORT CRO 102, UNIVERSITY OF TENNESSE NOVEMBER 1953.
                   EDGE-PUNCHED CARDS FOR SCIENTIFIC LITERATURE REFERENCES.
                   AEC REPORT CRO 102, UNIVERSITY OF TENNESSEE NOVEMBER 1953.
AHLIJT-56-GUF  AHLIN JT
                 GENERAL USE OF FOUR-HOLE RANDOMLY PUNCHED CARDS IN FILE
                   SEARCHING APPLICATIONS.
                 PAM 1702 DEC. 1956 RESEARCH LIB. SAN JOSE CAL.
ALEXSM-57-DPI  ALEXANDER SM
               STEVENS ME
                 DATA PROCESSORS FOR INFORMATION RETRIEVAL PURPOSES.
                 PAPER PRESENTED AT THE 132TH MEETING OF THE AMER. CHEM.
                   SOCIETY. NEW YORK, SEPT. 10, 1957
ALLEEP-43-PCN  ALLEN EP
                 A PUNCHED CARD FOR NEOPLASTIC DISEASES.
                 NEW ZEALAND MED J. 42 121 1943
ALLOAJ-  -IRM  ALLOT AJ
                 INFORMATION RETRIEVAL METHODS USED BY THE U S ARMY ORDNANCE
                   CORP IN DEPOT OPERATIONS.
                 DEPT. OF THE ARMY
AMDOC -53-CRM  ABSTRACT IN AM. DOC. APR. 1953.
                 CORRESPONDENCE REGARDING METALLURGICAL DOCUMENTATION OF THE.
                   CORDONNIER-BATTEN SYSTEM OF PUNCHED CARDS.
                 ABSTRACT IN AM. DOC. APR. 1953.
AMDOC -53-FSE  ABSTRACT IN AM. DOC. JANUARY 1953.
                 FILMOREX SYSTEM FOR ELECTRONIC SELECTION OF MICROFILM CARDS.
                 ABSTRACT IN AM. DOC. JANUARY 1953.
AMERDI-56-RCR  AMER DOCUMENTATION INST. 1956-1957 1957 16 PP.
                 ROSTER OF CURRENT RESEARCH IN DOCUMENTATION AND
                   LIBRARIANSHIP.
                 AMER DOCUMENTATION INST. 1956-1957 1957 16 PP.
AMERDO-52-GCR  AMER. DOC. III 1952 91-94
                 THE GENESIS AND CHARACTERISTICS OF REPORT LITERATURE.
                 AMER. DOC. III 1952 91-94
```

Fig. 2

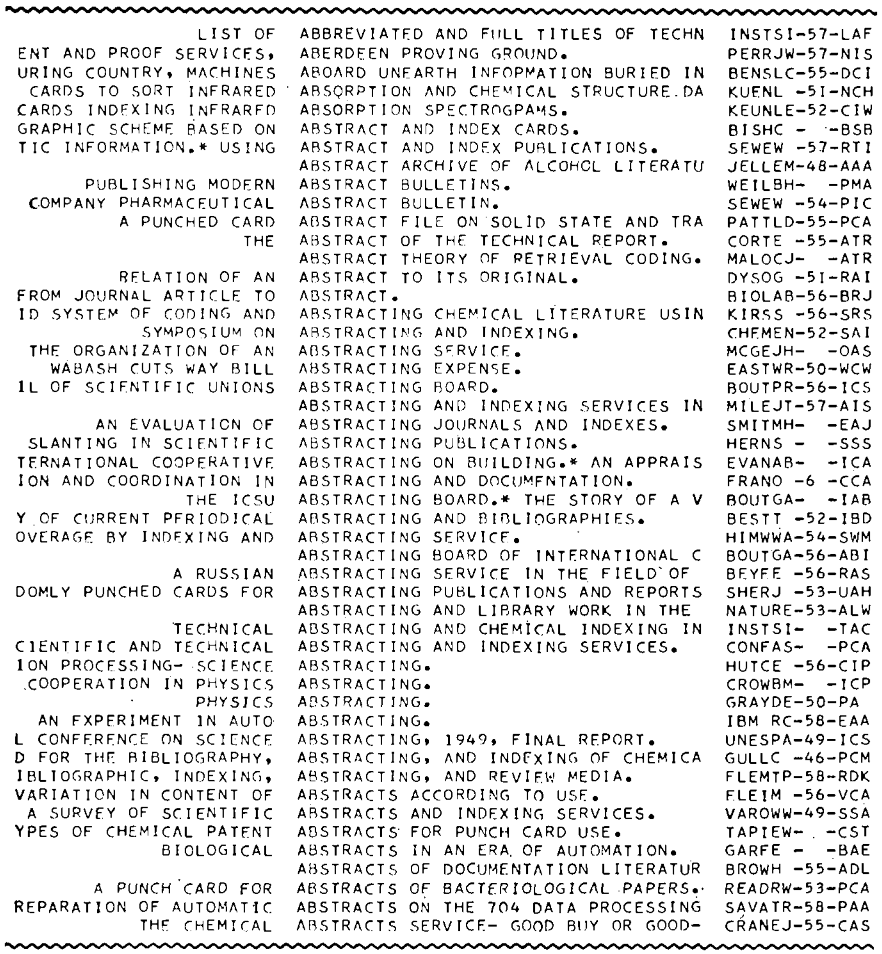
The finished prints of the bibliography and the index are mounted in two columns of 125 lines each for photographic reduction to fit 8½ × 11 size pages. The whole material is then printed and bound, and the KWIC Index is ready for mailing.

6. Conclusion

So far only a few KWIC Index services have been installed on an experimental basis. While user-acceptance has been very favorable, only experience will tell to what extent the objectives of this new device can be realized.

The following advantages are apparent at this time:

Key Words-In-Context Index



```
                   LIST OF  ABBREVIATED AND FULL TITLES OF TECHN   INSTSI-57-LAF
 ENT AND PROOF SERVICES,  ABERDEEN PROVING GROUND.                  PERRJW-57-NIS
 URING COUNTRY, MACHINES  ABOARD UNEARTH INFOPMATION BURIED IN      BENSLC-55-DCI
   CARDS TO SORT INFRARED  ABSORPTION AND CHEMICAL STRUCTURE.DA      KUENL -51-NCH
 CARDS INDEXING INFRARED  ABSORPTION SPECTROGPAMS.                  KEUNLE-52-CIW
 GRAPHIC SCHEME BASED ON  ABSTRACT AND INDEX CARDS.                 BISHC -  -BSB
 TIC INFORMATION.* USING  ABSTRACT AND INDEX PUBLICATIONS.          SEWEW -57-RTI
                          ABSTRACT ARCHIVE OF ALCOHOL LITERATU     JELLEM-48-AAA
       PUBLISHING MODERN  ABSTRACT BULLETINS.                       WEILBH-  -PMA
  COMPANY PHARMACEUTICAL  ABSTRACT BULLETIN.                        SEWEW -54-PIC
         A PUNCHED CARD  ABSTRACT FILE ON SOLID STATE AND TRA      PATTLD-55-PCA
                     THE  ABSTRACT OF THE TECHNICAL REPORT.         CORTE -55-ATR
                          ABSTRACT THEORY OF RETRIEVAL CODING.      MALOCJ-  -ATR
         RELATION OF AN  ABSTRACT TO ITS ORIGINAL.                 DYSOG -51-RAI
  FROM JOURNAL ARTICLE TO  ABSTRACT.                                BIOLAB-56-BRJ
 ID SYSTEM OF CODING AND  ABSTRACTING CHEMICAL LITERATURE USIN     KIRSS -56-SRS
             SYMPOSIUM ON  ABSTRACTING AND INDEXING.                 CHEMEN-52-SAI
   THE ORGANIZATION OF AN  ABSTRACTING SERVICE.                      MCGEJH-  -OAS
     WABASH CUTS WAY BILL  ABSTRACTING EXPENSE.                      EASTWR-50-WCW
 IL OF SCIENTIFIC UNIONS  ABSTRACTING BOARD.                        BOUTPR-56-ICS
                          ABSTRACTING AND INDEXING SERVICES IN     MILEJT-57-AIS
       AN EVALUATION OF  ABSTRACTING JOURNALS AND INDEXES.         SMITMH-  -EAJ
  SLANTING IN SCIENTIFIC  ABSTRACTING PUBLICATIONS.                 HERNS -  -SSS
 TERNATIONAL COOPERATIVE  ABSTRACTING ON BUILDING.* AN APPRAIS     EVANAB-  -ICA
 ION AND COORDINATION IN  ABSTRACTING AND DOCUMFNTATION.            FRANO -6 -CCA
                THE ICSU  ABSTRACTING BOARD.* THE STORY OF A V     BOUTGA-  -IAB
 Y OF CURRENT PERIODICAL  ABSTRACTING AND BIBLIOGRAPHIES.           BESTT -52-IBD
 OVERAGE BY INDEXING AND  ABSTRACTING SERVICE.                      HIMWWA-54-SWM
                          ABSTRACTING BOARD OF INTERNATIONAL C     BOUTGA-56-ABI
               A RUSSIAN  ABSTRACTING SERVICE IN THE FIELD OF      BEYEE -56-RAS
 DOMLY PUNCHED CARDS FOR  ABSTRACTING PUBLICATIONS AND REPORTS     SHERJ -53-UAH
                          ABSTRACTING AND LIBRARY WORK IN THE      NATURE-53-ALW
               TECHNICAL  ABSTRACTING AND CHEMICAL INDEXING IN     INSTSI-  -TAC
 CIENTIFIC AND TECHNICAL  ABSTRACTING AND INDEXING SERVICES.        CONFAS-  -PCA
 ION PROCESSING- SCIENCE  ABSTRACTING.                              HUTCE -56-CIP
  COOPERATION IN PHYSICS  ABSTRACTING.                              CROWBM-  -ICP
                  PHYSICS  ABSTRACTING.                              GRAYDE-50-PA
   AN EXPERIMENT IN AUTO  ABSTRACTING.                              IBM RC-58-EAA
 L CONFERENCE ON SCIENCE  ABSTRACTING, 1949, FINAL REPORT.          UNESPA-49-ICS
 D FOR THE BIBLIOGRAPHY,  ABSTRACTING, AND INDEXING OF CHEMICA     GULLC -46-PCM
 IBLIOGRAPHIC, INDEXING,  ABSTRACTING, AND REVIEW MEDIA.            FLEMTP-58-RDK
 VARIATION IN CONTENT OF  ABSTRACTS ACCORDING TO USE.               FLEIM -56-VCA
  A SURVEY OF SCIENTIFIC  ABSTRACTS AND INDEXING SERVICES.          VAROWW-49-SSA
 YPES OF CHEMICAL PATENT  ABSTRACTS FOR PUNCH CARD USE.             TAPIEW-  -CST
               BIOLOGICAL  ABSTRACTS IN AN ERA OF AUTOMATION.        GARFE -  -BAE
                          ABSTRACTS OF DOCUMENTATION LITERATUR     BROWH -55-ADL
       A PUNCH CARD FOR  ABSTRACTS OF BACTERIOLOGICAL PAPERS.      READRW-53-PCA
 REPARATION OF AUTOMATIC  ABSTRACTS ON THE 704 DATA PROCESSING     SAVATR-58-PAA
            THE CHEMICAL  ABSTRACTS SERVICE- GOOD BUY OR GOOD-     CRANEJ-55-CAS
```

Fig. 3

1. Because of the mechanical method of preparation, more information may be displayed than would have been practical by conventional means.
2. Keywords-in-context permit the cross correlation of subjects to an extent not realizable by conventional procedures.
3. KWIC indexes provide an invaluable basis for the compilation of reference material by professional catalogers and indexers.

It has to be kept in mind that machine products of the kind discussed here can never reach the level of perfection that humans are capable of and that there will always be residual effort left for humans. It is hoped that in the case of the KWIC Index this effort is acceptable to the user.

Luhn, H.P.

READINGS IN

AUTOMATIC LANGUAGE PROCESSING

EDITED BY

DAVID G. HAYS

The RAND Corporation

NEW YORK

AMERICAN ELSEVIER PUBLISHING COMPANY INC.

1966