

KEYWORD-IN-CONTEXT INDEX FOR TECHNICAL LITERATURE

(KWIC INDEX)

by

H. P. Luhn

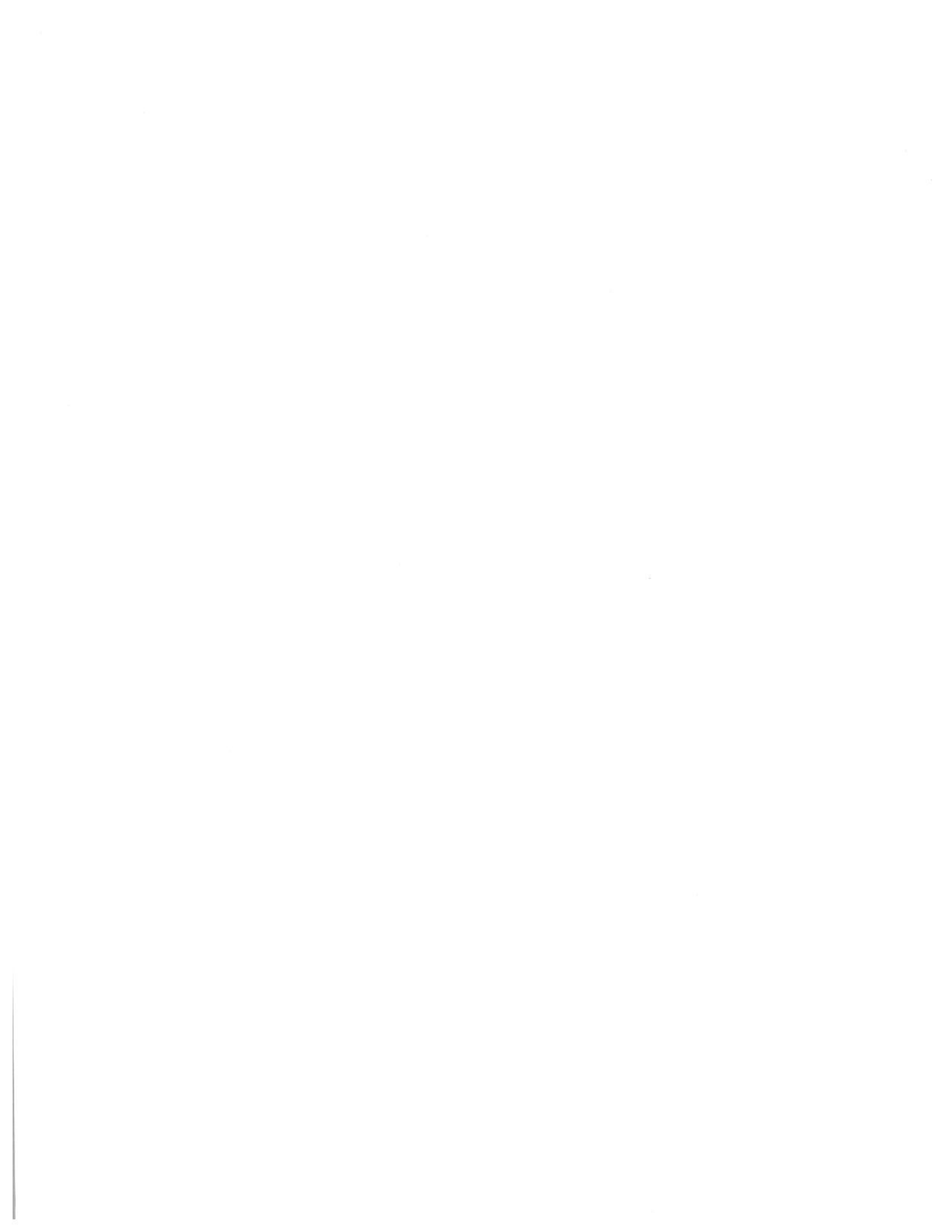
International Business Machines Corporation
Advanced Systems Development Division
Yorktown Heights, New York

ABSTRACT: A distinction is made between bibliographical indexes for new and past literature based on the willingness of the user to trade perfection for currency. Indexes giving keywords in their context are proposed as suitable for disseminating new information. These can be entirely machine-generated and hence kept up-to-date with the current literature. A compatible coding scheme to identify the indexed documents is also proposed. In it elements are automatically extracted from the usual identifiers of the document so that the coded identifier yields a maximum of information while remaining susceptible to normal methods of ordering.

Printed in U. S. A.

TABLE OF CONTENTS

	Page
Introduction	1
Dissemination vs. Retrieval	2
Indexing by Means of Keyword-in-Context	3
Organization of a Bibliographical KWIC Index	6
A Derived Code for the Identification of Bibliographical Items	8
Automatic Preparation of KWIC Indexes	11
Creation of Machine Readable Records	11
Machine Processing	12
Conclusion	16



KEYWORD-IN-CONTEXT INDEX FOR TECHNICAL LITERATURE

(KWIC Index)

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.

DISSEMINATION vs. 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, i. e., 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.

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 non-significant. 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 non-significant 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 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 60 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.

KEYWORD-IN-CONTEXT BIBLIOGRAPHICAL INDEX

COULOMB EXCHANGE ENERGY FROM SHELL-MODEL WAV	1719
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
INTERNAL 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
ANGULAR DISTRIBUTIONS IN FISSION INDUCED BY ALPHA PARTICLES,	0536
UTRON CROSS SECTIONS OF FISSIIONABLE NUCLEI.	0203
AL COSMIC-RAY INTENSITY FLUCTUATIONS OBSERVED AT SOUTHERN ST	1798
NEUTRINO CORRELATION IN FLUX OF COSMIC-RAY PARTICLES WITH Z-	0597
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
DECAY OF FREQUENCY SHIFT OF THE ZERO-FIELD HY	0449
GADOLINIUM-159.	0262
SECTIONAL CORRELATION OF GAMMA RADIATION FROM AL27-PLUS-P AND	0239
CISION DETERMINATION OF GAMMA RAYS IN GE72.	0229
GAMMA RAYS FOLLOWING P,P-PRIME-GAMMA	0532
P/S32 AND S32/P,P-PRIME GAMMA-RAY THRESHOLD METHOD AND THE O	0461
ONSTANT OF YTTRIUM IRON GAMMA/S32.	1702
LORENTZIAN GARHET AT 0 DEG K.	0395
TIBILITY OF AN ELECTRON GAS AND HOT ELECTRONS.	1567
UCTIVITY OF AN ELECTRON GAS AT HIGH DENSITY.	0328
OF AN ELECTRON GAS IN A GAS IN A GASEOUS PLASMA.	0001
DUCTED BY VARIOUS BUFFER GASEOUS PLASMA.	0001
BUFFER GASES.	0449
IONIZED GASES.	0450
EZORESISTANCE IN N-TYPE GA,AS.	1441
IN ELECTRON-IRRADIATED GE AT 80 DEG K.	1533
LATION OF GAMMA RAYS IN GE72.	0362
NERAL RELATIVITY AS THE GENERATORS OF COORDINATE TRANSFORMAT	0229
ETORESISTANCE IN N-TYPE GERMANIUM AT LOW TEMPERATURES.	0287
CONDUCTION ELECTRONS IN GERMANIUM.	0317
IATIVE RECOMBINATION IN GERMANIUM.	0298
PARTICLES IN LINEARIZED GERMANIUM.	0330
ENERGIES OF GRAVITATIONAL THEORY.	0674
KINEMATICS OF GROUND AND EXCITED NUCLEAR CONFIGURA	0452
RIC CONSTANTS OF ALKALI HALIDE CRYSTALS.	1649
TWO HALL EFFECT, MAGNETORESISTANCE, AND	1488
HALL EFFECTS OF IRON-COBALT ALLOYS.	0090
HALL MOBILITY OF CARRIERS IN IMPURE	0381
A DILUTE BOSE SYSTEM OF HARD SPHERES. I. EQUILIBRIUM PROPERT	1516
OLUME ANOMALY OF LIQUID HE3 ARISING FROM ITS NUCLEAR SPIN 'SY	0044
L EXCITATIONS IN LIQUID HE3.	1419
OF 95-MEV PROTONS WITH HE4.	1483
SPECIFIC HEAT OF LI,F AND KI AT LOW TEMPERATU	1465
TION OF DONOR STATES IN HEAT-TREATED SILICON.	1658
UCLEAR ENERGY LEVELS IN HEAVY ELEMENTS.	0049
XCITATION OF PROTONS IN HELIUM II BY COLD NEUTRONS.	1546
MAGNETIC MOMENT OF HELIUM IN ITS 3S1 METASTABLE STATE.	0518
LITY OF LI-PLUS IONS IN HELIUM.	0011
OF SN, IN, TA, TL, AND HG.	1627
ISOMERS IN TB158 AND HO163.	0445
LORENTZIAN GAS AND HOT ELECTRONS.	0031
ICROWAVE PROPAGATION IN HOT MAGNETO-PLASMAS.	0489
OF THE ELECTRON ON THE HYDROGEN ENERGY LEVELS.	1567
DISSOCIATION OF THE HYDROGEN MOLECULE ION BY ELECTRON IM	1460
SS OF SLOW ELECTRONS IN HYDROGEN.	1637
SHIFT OF THE ZERO-FIELD HYDROMAGNETIC EQUATIONS FOR TWO ISOT	0155
NARROW HYPERFINE SPLITTING OF CS133 PRODUCE	0159
MASSSES OF CHARGED SIGMA HYPERFINE-ABSORPTION LINES OF CS133	1441
HYPERFINE-STRUCTURE SEPARATIONS AND	0449
HYPERONS AND THE NEGATIVE K MESON.	0450
	0186
	0622

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.

The creation of cumulative KWIC Indexes poses a new problem in that an updated issue calls for renumbering of previously listed items and of the related index entries. In order to eliminate this nuisance and possible confusion, it is proposed to use a novel type of identification code which remains unchanged and yet permits the insertion of new items in a previous list while maintaining sequential order. This identification code will be described in some detail.

A DERIVED CODE
FOR THE IDENTIFICATION OF BIBLIOGRAPHICAL ITEMS

This code is derived from factual data inherent to a document as evidenced by the publisher's printed identification, comprising the following elements:

1. The name of the author (or senior author) or originating agency.
2. The year of publication.
3. The title of the document.

The objective is to derive unique notations with a reasonably low probability of duplication. A further objective is to produce a notation which conveys intelligible information which will enhance its usefulness.

The code comprises eleven character positions. The first six are derived from the name of the author or originating agency, the next two consist of the ten's and unit digit of the year of publication, and the last three are derived from the title.

The rules for deriving the first six letters of the code do not distinguish between names of authors and originating agencies, but apply to whatever words or letters follow each other in spelling such names. It is, however, required that such names be separated from subsequent other information by a two-space interval and that an author be identified by recording his surname first. If an author's given names are represented by initials only, these initials should either be written without space between them, or each initial should be followed by a period. Punctuation marks and special characters are not made part of the code. The following 16 words are disregarded except when followed by a period or at least two spaces:

a, an, and, as, at, by, for, from
if, in, of, on, or, the, to, with

This latter provision is to prevent the elimination of initials because of their possible similarity with the words just listed.

The formation of the code notation is carried out by entering the first four letters of the adjusted name into the first four positions of the code, irrespective of whether these letters are adjoining or separated. If the remaining portion of the name contains no additional words or letters separated by spaces, the 5th and 6th letter of the name, to the extent that they are present, are placed in position 5 and 6 of the code. If one additional word remains, its two starting letters are placed in position 5 and 6. If two or more additional words remain, the starting letters of the then first two remaining words are placed in position 5 and 6 respectively.

In applying these rules to an author's name, the first four letters of his surname will become the first four letters of the code and the initials of the author's first and second given name will become the next two letters.

The rules for deriving from the title the last 3 digits, position 9, 10 and 11 of the code, require that, as before, punctuation marks and special characters be disregarded and likewise any of the 16 words previously listed. The three code letters for the title are derived from the initial letters of the first three words of the adjusted title to the extent that they are present. These letters are placed in positions 9, 10 and 11 of the notation in the order in which the corresponding words appear in the title. Subsequent words, if any, are disregarded.

In accordance with the above rules the identification code for this very document would read: "LUHNHP59KCI". Other typical examples of bibliographical identification codes are given in Figure 2. The letters or numbers extracted to form the code have been underlined.

The above code format was chosen over other possible variations for the reason that when bibliographical entries are ordered in alphabetical sequence in accordance with this code, the utility of the resulting listing as an author index is not seriously impaired since the variations between this order and that demanded by the fully spelled out names are slight.

The usefulness of a derived identification code for bibliographical items might extend beyond the present application. Such a code might very well solve some of the identification problems

encountered in documentation in that identical codes may be derived independently for the same documents. If desirable, a self-checking feature may be incorporated in the code by adding an extra character with the aid of which a simple arithmetic test may be performed to ascertain correctness of spelling. The code may furthermore be amplified by the addition of source date, derived by methods similar to those used for authors as explained above.

DERIVATION BY MACHINE OF 11-CHARACTER INDEX CODES
FOR THE IDENTIFICATION OF BIBLIOGRAPHICAL ITEMS.

CCGOML-52-WHT	<u>C</u> . <u>C</u> . <u>G</u> OODRICH MEMORIAL <u>L</u> IBRARY <u>W</u> HY AND <u>H</u> OW THE <u>T</u> ECHNICAL LIBRARY SHOULD BE SET UP AND UTILIZED IN CREATIVE ENGINEERING. MACHINE DESIGN SEPT 1952 PP. 111
HOLMJE-57-MDD	<u>H</u> OLMSTROM <u>J</u> E <u>M</u> ULTILINGUAL <u>D</u> ICIONARIES AND <u>D</u> OCUMENTATION NACHRICHTEN DOKUMENTATION MAR. 1957
INSTAS-----SST	<u>I</u> NSTITUTE OF THE <u>A</u> ERONAUTICAL <u>S</u> CIENCES <u>S</u> YMPOSIUM ON STANDARDIZATION IN <u>T</u> ECHNICAL INFORMATION SERVICES FOR GOVERNMENT US RESEARCH AND DEVELOPMENT BOARD
JOHNHU-55-MIP	<u>J</u> OHNS HOPKINS UNIVERSITY <u>M</u> EICAL <u>I</u> NDEXING PROJECT, FINAL REPORT. WELCH MEDICAL LIBRARY, JOHNS HOPKINS UNIVERSITY% MEDICAL INDEXING PROJECT, FINAL REPORT, 1955
KENTA -57-MSM	<u>K</u> ENT <u>A</u> <u>M</u> ACHINE <u>S</u> EARCHING OF <u>M</u> ETALLURGICAL LITERATURE. METAL PROGRESS, FEB. 1957
KINGGW-55-NAI	<u>K</u> ING <u>G</u> W <u>A</u> <u>N</u> EW <u>A</u> PPROACH TO <u>I</u> NFORMATION STORAGE. CONTROL ENGINEERING AUGUST 1955
KOELGJ-58-PFM	<u>K</u> OELEWIJN <u>G</u> J <u>T</u> HE <u>P</u> OSSIBILITIES OF <u>F</u> AR- <u>R</u> EACHING <u>M</u> ECHANIZATION OF <u>N</u> OVELTY SEARCH OF THE PATENT LITERATURE. PREPRINTS OF PAPERS FOR THE INTERNATIONAL CONFERENCE ON SCIENTIFIC INFORMATION WASH. DC 1958
MACCGE-54-CFS	<u>M</u> AC <u>C</u> ASLAND <u>G</u> E <u>A</u> <u>C</u> ONCISE <u>F</u> ORM FOR <u>S</u> CIENTIFIC LITERATURE CITATIONS. SCIENCE 120, JULY 1954
MIDWRI-57-EBM	<u>M</u> IDWEST RESEARCH <u>I</u> NSTITUTE, KANSAS CITY, MO. <u>E</u> LECTRONIC BRAIN <u>M</u> ULLS <u>N</u> EW <u>C</u> HEMICAL USES. CHEMICAL WEEK NOV. 23, 1957
NATLBS-57-SPE	<u>N</u> ATL. <u>B</u> UR. OF <u>S</u> TANDARDS WASHINGTON DC <u>S</u> YNTAX <u>P</u> ATTERNS IN <u>E</u> NGLISH STUDIED BY ELECTRONIC COMPUTER. COMPUTERS AND AUTOMATION JULY 1957

Note: The letters or numbers extracted by the machine to form
the code have been underlined.

Fig. 2

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.

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.

Machine Processing

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

First the 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 25 of a 60 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. 3. 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. 4 with their identification codes at the right. Figs. 3 and 4*, are typical pages of an index. It should be noted that

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

BIBLIOGRAPHY

- ABA OR-54-SPL A.B.A. OPERATIONS RESEARCH COMMITTEE
SUBCOMMITTEE ON PUNCHED CARDS AND LEGAL LITERATURE
SEARCHING.
NATURE AND SCOPE OF PROGRAM JUNE 1954
- ADAINC-55-CIS ADAIR WC
CITATION INDEXES FOR SCIENTIFIC LITERATURES.
AMERICAN DOCUMENTATION, 6, 1/1, 1955.
- ADAMS -56-INR ADAMS S
INFORMATION - A NATIONAL RESOURCE.
AMER DOC V. VII NO. 2 APR 1956
- ADIANC-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 TENNESSEE 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..
- ALEXSN-57-DPI ALEXANDER SN
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
- AMERDO-52-PEM AMER. DOC. III 1952 58-60
PRACTICAL EXPERIENCE IN MICROFACSIMILE PUBLICATION.
AMER. DOC. III 1952 58-60
- AMERDO-54-PRM AMER. DOC. JANUARY 1954.
PRELIMINARY REPORT ON THE MECHANICAL ANALYSIS OF INFORMATION
BY USE OF THE 101 STATISTICAL PUNCHED-CARD MACHINE.
AMER. DOC. JANUARY 1954.
- AMERDO-57-ICS AMERICAN DOCUMENTATION 8 /1/ 76-77 JAN., 1957
INTERNATIONAL CONFERENCE ON SCIENTIFIC INFORMATION.
AMERICAN DOCUMENTATION 8 /1/ 76-77 JAN., 1957
- AMERDO-57-LSL AMER. DOC. OCTOBER 1957.
LOGICAL STRUCTURES IN LANGUAGE.
AMER. DOC. OCTOBER 1957.
- AMERDO-57-TLN AMER. DOCUMENTATION V. VIII, NO. 1 JAN. 1957
THE TECHNICAL LIBRARY IN THE NEW RESEARCH LABORATORY OF THE
NATIONAL CARBON COMPANY.
AMER. DOCUMENTATION V. VIII, NO. 1 JAN. 1957
- AMERFO-48-FSU AMER. FOUNDRYMAN, P. 29 OCTOBER 1948.
A.F.S. USES PUNCHED CARD SYSTEM TO INDEX FOUNDRY LITERATURE.
AMER. FOUNDRYMAN, P. 29 OCTOBER 1948.
- AMERIA-53-ISF AMER. INST. OF ARCHITECTS, 1953 63P. WASHINGTON D.C.
A.I.A. STANDARD FILING SYSTEMS.
AMER. INST. OF ARCHITECTS, 1953 63P. WASHINGTON D.C.
- AMERSC-54-CTB AMER. SCI. 42, P. 248 APRIL 1954.
CAN TRANSLATION BE MECHANIZED?
AMER. SCI. 42, P. 248 APRIL 1954.
- AMERSM-57-MIC AMERICAN SOCIETY FOR METALS, 1957
THE METALS INFORMATION CENTER OF TOMORROW/16-MM COLOR SOUND
MOVIE/
AMERICAN SOCIETY FOR METALS, 1957
- AMES -48-UPL AMES
THE USE OF PUNCHED CARDS FOR INDEXING AND CLASSIFYING
BIO-CHEMICAL LITERATURE.
SPEC. LIB. 39, P. 233 1948.
- AMSTAB-49-UPL AMSTER AB ET AL
USE OF PUNCHED CARDS.
J. CHEM. EDUC. 26.* 304-6 1949
- ANDRDD- -A00 ANDREWS DD
NEWMAN SM
ACTIVITIES AND OBJECTIVES OF THE OFFICE OF RESEARCH AND
DEVELOPMENT IN THE U.S. PATENT OFFICE.
U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
- ANDRDD- -CIP ANDREWS DD
COOPERATIVE INFORMATION PROCESSING - PATENTS.
U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
- ANDRDD- -ILA ANDREWS DD
INTERRELATED LOGIC ACCUMULATING SCANNER.
U.S. PATENT OFF.
- ANDRDD- -LPO ANDREWS DD
THE LAW AND PATENT OFFICE RESEARCH.
U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
- ANDRDD- -MFG ANDREWS DD
MAJOR FACTORS GOVERNING THE SELECTION OF MACHINE DATA
HANDLING TECHNIQUE.
U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
- ANDRDD- -PEP ANDREWS DD
PATENT EXAMINING PAST, PRESENT AND FUTURE.
U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
- ANDRDD- -PRU ANDREWS DD
PROGRESS REPORT ON U.S. PATENT OFFICE MECHANIZED SEARCHING.
U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
- ANDRJL- -DEP ANDRIOT JL
DOCUMENTS EXPEDITING PROJECT.
LIBRARY JOURNAL, LXXVII, 693-695.
- APPLPL-52-SHL APPLIED PHYSICS LAB. JOHNS HOPKINS UNIV. NOV. 1952
SUBJECT HEADING LIST FOR GUIDED MISSILES.
APPLIED PHYSICS LAB. JOHNS HOPKINS UNIV. NOV. 1952
- ARCECJ-56-MPT ARCEAUX C /ETHYL CORP. BATON ROUGE/
MOVIES PRESENT TECHNICAL PAPERS.
CHEM. & ENG. NEWS 34, 627-9 FEB. 1956
- ARMEFM-56-PBM THE ARMED FORCES MEDICAL LIBRARY
PROPOSED BIBLIOGRAPHY OF MEDICAL CONGRESS PUBLICATIONS.
UNESCO BULL. FOR LIBRARIES, 10, 22, HAN JAN. 1956
- ARREST-57-MAS ARMED SERVICES TECHNICAL INFORMATION AGENCY WASHINGTON FEB.
1957.
MULTIPLE ASPECT SEARCHING FOR INFORMATION RETRIEVAL.
ARMED SERVICES TECHNICAL INFORMATION AGENCY WASHINGTON FEB.
1957.
- ARMYAF-55-N5V ARMY-AIR FORCE REGISTER, 77, 8, 15 OCT. 1955
NEW 50-LANGUAGE VARI-TYPER DEVELOPED.
ARMY-AIR FORCE REGISTER, 77, 8, 15 OCT. 1955
- ARNHAA-47-NCL ARNHAM AA
A NEW CLASSIFICATION.* LIMITED DISTRIBUTION.
TECHNICAL DATA DIGEST XII SEPT 1, 1947 2-3
- ARNHF -57-PCA ARNHOLDT F
PUNCHED CARDS AS AN AID TO INDEXING UROLOGICAL REPORTS.
NACHRICHTEN DOKUMENTATION 8/2/ 82-85 JUNE 1957
- ASBUNC-57-MVT ASBURY WC
REEVES ED
KNOX WT
MANAGEMENTS VIEWS ON TECHNICAL - INFORMATION ACTIVITIES.
PRESENTED BEFORE THE AMERICAN CHEMICAL SOCIETY SEPT 10, 1957
- ASHTHD-52-PIE ASHTHORPE HD
THE PUNCHED-CARD INDEXING EXPERIMENT AT THE LIBRARY OF THE
ATOMIC ENERGY RESEARCH ESTABLISHMENT, HARWELL.
ASLIB PROCEEDINGS V. 4, NO. 2 MAY 1952
- ASLIB -57-PIE DORKING, ENGLAND, MAY 13-17, 1957 LONDON.* ASLIB, 1957
PROCESSINGS OF THE INTERNATIONAL STUDY CONFERENCE ON
CLASSIFICATION FOR INFORMATION RETRIEVAL.
DORKING, ENGLAND, MAY 13-17, 1957 LONDON.* ASLIB, 1957
- ASLIPR-57-IMA ASLIB PROCEEDINGS 9 /10/ 301-309., DISCUSSION 309-314.
OCT. 1957
INFORMATION METHODS, ADAPTATION TO GROWTH.
ASLIB PROCEEDINGS 9 /10/ 301-309., DISCUSSION 309-314,
OCT. 1957
- ASTRMM-58-RLM ASTRAHAN MM
THE ROLE OF LARGE MEMORIES IN SCIENTIFIC COMMUNICATIONS.
IBM JOURNAL, OCTOBER, 1958
- ATANJV-36-APE ATANASOFF JV
BRANDT AE
APPLICATION OF PUNCHED-CARD EQUIPMENT TO THE ANALYSIS OF
COMPLEX SPECTRA.
J. OPT. SOC. AMER. V. 26, FEB. 1936, PP. 83-88
- AUTOCO-56-PIE AUTOMATIC CONTROL V. 5, JULY 1956 P. 4
PHOTOSCOPIC INFORMATION STORAGE UNIT.
AUTOMATIC CONTROL V. 5, JULY 1956 P. 4
- AUTOCO-58-AOP AUTOMATIC CONTROL, V. 7, NO. 1, JAN. 1958 6-10
AUTOMATIC OPERATIONS IN POSTAL PROCESSING.* CONTROL ELEMENTS
SPEED SORTING FUNCTION.
AUTOMATIC CONTROL, V. 7, NO. 1, JAN. 1958 6-10
- AVAKE -57-AAM AVAKIIE E
AMFIS--THE AUTOMATIC MICROFILM INFORMATION SYSTEM.
SPEC.LIB. /APR.1957/
- AVIAJ -52-10S AVIAS J
INTERNATIONAL ORGANIZATION OF SCIENTIFIC DOCUMENTATION BASE
ON LEGISLATION.
SCIENCE, 115, 250 /1952/.
- AYERJA-48-LTL AYER JA
LANGUAGE, TRUTH AND LOGIC.
VICTOR GOLLANCZ, LONDON, 1948, PP. 33-44, 87-102.
- BABAS -56-SRD BABA S
SOME RECENT DEVELOPMENTS IN JAPANESE DOCUMENTATION.
UNESCO REF. NS 94/LX20 1956 13 P.
- BABEP -56-OMT BABEL P
THE OUTLOOK FOR MECHANICAL TRANSLATION.
BABEL, P. 99 OCTOBER 1956.
- BACHH -55-NIA BACH H
THE NEED FOR AN INSTITUTE FOR ADVANCED LIBRARY RESEARCH.
WILSON LIBRARY BULLETIN, 29, 709-711 MAY 1955
- BACHRW-52-SVM BACHELDER RW
THE SCOPE AND VALUE OF THE MICROCARD.
SPEC. LIB. XLIII 1952 157-61
- BAGEP -51-EDM BAGELEY P
ELECTRONIC DIGITAL MACHINES FOR HIGH SPEED INFORMATION
ON SEARCHING.
MASSACHUSETTS INSTITUTE OF TECHNOLOGY /NOV. 1951/ /MASTER-5
THESIS, AUG. 1951
- BAGLPR-51-ANE BAGLEY PR
PERRY JW
APPLICATION OF NEWER ELECTRONIC TECHNIQUES TO INFORMATION
SEARCHING.
PAPER PRESENTED BEFORE DIVISION OF CHEMICAL LITERATURE OF
THE AMERICAN CHEMICAL SOCIETY AT THE DIAMOND JUBILEE, N.Y.
CITY, SEPT. 1951.
- BAGLPR-52-ANE BAGLEY PR
APPLICABILITY OF NEWER ELECTRONIC TECHNIQUES TO INFORMATION
SEARCHING.
MACHINE TECHNIQUES FOR INFORMATION SELECTION JUNE 1952
- BAHRMH-55-NAA BAHMER RH
THE NATIONAL ARCHIVES AFTER 20 YEARS.
AMERICAN ARCHIVIST, 18, 195-205, JULY 1955
- BAILBS-55-BAC BAILEY BROS. & SWINFEN LTD. LONDON GILES HIGH ST. 1955
BAILEY-S. AFRICAN-ORIENTAL CATALOGUE OF DICTIONARIES,
GRAMMARS AND PHRASEBOOKS.
BAILEY BROS. & SWINFEN LTD. LONDON GILES HIGH ST. 1955
- BAILCF-46-PCI BAILEY CF
CASEY RS
COX GJ
PUNCHED CARDS FOR INDEXING SCIENTIFIC DATA.
SCIENCE 104.* 181, 1946
- BAILM -53-MSU BAILEY M
MECHANIZED SEARCHING IN THE U.S. PATENT OFFICE.
JOUR. PATENT OFFICE SOCIETY, P. 566 AUG. 1953

the identification codes shown in these samples have been derived by rules differing slightly from those described in the present paper.

The finished prints of the bibliography and the index are mounted in two columns of 125 lines each for photographic reduction to fit 8 1/2 X 11 size pages. The whole material is then printed and bound and the KWIC Index is ready for mailing.

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:

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.