

UNCLASSIFIED

AD NUMBER
AD443912
NEW LIMITATION CHANGE
TO Approved for public release, distribution unlimited
FROM Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; 1964. Other requests shall be referred to US Naval Academy, Annapolis, MD.
AUTHORITY
USNA/Pitts Univ ltr, 15 Apr 1969

THIS PAGE IS UNCLASSIFIED

UNCLASSIFIED

AD. 4 4 3 9 1 2

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION ALEXANDRIA, VIRGINIA

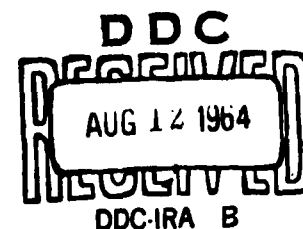


UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

4 4 3 9 1 2

John M. Sedano
Department of Industrial Engineering



Keyword-In-Context (KWIC) Indexing:
Background, Statistical Evaluation, Pros and Cons, and Applications

Major Advisor: Dr. A. G. Holzman

Encouraged by among other things the "publish or perish" milieu of today, scientists, educators and researchers are a prolific lot, publishing technical literature in explosive proportions. The volume is such that machines have been called on to aid and expedite the processing and retrieval of information.

This thesis discusses keyword-in-context (KWIC) indexing, a new method of machine indexing literature. In presenting the subject, the author assumes that the reader is unfamiliar with the new discipline of Information Science, precipitated by the information explosion.

The background of machine indexing is developed deductively. Next, the author describes KWIC indexing, evaluates statistically the hypothesis upon which it is based (that titles are descriptive of the article or document they represent), lists its pros, cons and applications. And finally, the author concludes that KWIC indexing is indeed an efficacious new tool.

KEYWORD-IN-CONTEXT (KWIC) INDEXING:

BACKGROUND, STATISTICAL EVALUATION, PROS AND CONS, AND APPLICATIONS .

by

John Michael Sedano .

B.S., U.S. Naval Academy, 1957

Submitted to the Graduate Faculty of the Schools of
Engineering and Mines in partial fulfillment of
the requirements for the degree of
Master of Science
in
Industrial Engineering

University of Pittsburgh

1964

FOREWORD

Six months ago the author was first introduced to a keyword-in-context (KWIC) index. That the KWIC index was based upon the titles of documents rather than document contents seemed unusual, that it was produced by a computer, not a human, in minutes at a fractional cost seemed incredible and that there were individuals who doubted and even belittled the efficacy of this wonderous technique seemed perplexing. As time passed the author became more intrigued with KWIC indexing and finally decided to make it the subject of this his Masters thesis. Format-wise, the background of machine indexing in general is developed deductively. Next the author describes KWIC indexing, evaluates the hypothesis upon which it is based, lists its pros, cons and applications, and lastly draws conclusions. Furthermore, this complete investigation into KWIC indexing is basic in approach; it is written for the reader who is unfamiliar with the newly emerged discipline, Information Science.

ACKNOWLEDGEMENT

The author wishes to express his gratitude to the members of the reviewing committee - Dr. A.G. Holzman, Dr. A. Kent and Professor H.W. Wolfarth - for their comments and suggestions; to Mr. C.K. Bauer, Manager, Scientific and Technical Information Department, Lockheed Aircraft Corporation, Marietta, Georgia for sending the author exhibits for inclusion in this thesis; and especially to Mr. D.H. Kraft, Industry Representative, Information Retrieval, IBM Corporation, Chicago, Illinois for generously providing ideas, information and materials used throughout the thesis.

TABLE OF CONTENTS

	Page
FOREWORD	11
ACKNOWLEDGEMENT	111
LIST OF FIGURES	vi
I. BACKGROUND	1
A. Information Explosion	1
B. Advent of Mechanized Information Retrieval	4
C. Unit Operations	7
D. Analysis	9
E. Indexing	13
F. Machine Indexing	16
1. General	16
2. Approaches to Machine Indexing	17
II. KEYWORD-IN-CONTEXT (KWIC) INDEX	20
III. STUDY OF ADEQUACY OF TITLES AS SEARCH TOOLS	28
A. General	28
B. Procedure	29
C. Engineering Index Sample	35
1. General	35
2. Estimation of Sample Size	35
3. Classification	37
4. Confidence Limits	38
D. Applied Science and Technology Index Sample	40
1. General	40
2. Classification	41
3. Confidence Limits	41
E. Metal Literature Sample	41
1. General	41
2. Classification	42
3. Confidence Limits	42

TABLE OF CONTENTS (CONT'D.)

	Page
F. Public Affairs and Information Service Sample	43
1. General	43
2. Classification	44
3. Confidence Limits	44
G. Business Periodicals Index	44
1. General	44
2. Classification	45
3. Confidence Limits	45
H. International Index Sample	46
1. General	46
2. Classification	46
3. Confidence Limits	46
I. Interpretation of Data	47
IV. PROS AND CONS OF KWIC INDEXING	52
A. General	52
B. Advantages	52
C. Disadvantages	55
V. APPLICATIONS OF KWIC INDEXING	60
VI. CONCLUSIONS	62
VII. FUTURE	65
APPENDIX A	67
BIBLIOGRAPHY	73

LIST OF FIGURES

	Page
Figure 1 - Degree of Document Processing	10
Figure 2 - Preliminary Steps Common to Generating Abstracts, Extracts, Classifications and Indexes	12
Figure 3 - Degree of Format Structure	18
Figure 4 - Development of Thesis: From the General to the Specific	19
Figure 5 - KWIC Index Example Using a Single Title	20
Figure 6 - KWIC Index Using "Wrap-Around"	22
Figure 7 - Typical Page From a KWIC Index Proper	23
Figure 8 - Typical Bibliography Page	24
Figure 9 - KWIC Author Index Page	25
Figure 10 - Generation of KWIC Index	27
Figure 11 - Logic Followed in Sorting Titles by Type	32
Figure 12 - KWIC Index to Illustrate a Method of Improving Readability	57
Figure 13 - The 103 Character, Single Line KWIC Index Format	59
Figure 14 - Accession List	69
Figure 15 - Report Number Index	70
Figure 16 - KWIC Index (For Books and Reports)	71
Figure 17 - Personal Author Index	72

I. BACKGROUND

A. Information Explosion

The importance of the availability of information to the scientific community has long been recognized. As early as 1852, Joseph Henry, secretary of the Smithsonian Institution reported to Congress:

"...It is estimated that about twenty thousand volumes, including pamphlets, purporting to be additions to the sum of human knowledge, are published annually; and unless this mass be properly arranged, and the means furnished by which its contents may be ascertained, literature and science will be overwhelmed by their own unwieldy bulk. The pile will begin to totter under its own weight." (1)*

A half-century later, Henry Adams wrote of the law of acceleration of human knowledge, that if it were "prolonged one generation longer it would require a new social mind...subject to new laws."^{(2),(3)}

World War II precipitated a massive increase in scientific research and a concomitant increase in scientific literature of all sorts. Noting the momentum of wartime research and its implications, Dr. Vannevar Bush, the great American scientist prophetically urged his colleagues at war's end to direct their efforts toward the problem of sophisticating the storage, processing and retrieval of information. In his now famous article entitled "As We May Think" he stated:

"...He (man) has built a civilization so complex that he needs to mechanize his records more fully if he is to push his experiment to its logical conclusion and not merely become bogged down part way there by over taxing his limited memory. His excursions may be more enjoyable if he can reacquire the privilege of forgetting the manifold things he does not need to have immediately at hand, with some assurance that he can find them again if they prove important." (4)

* Parenthetical references placed superior to the line of text refer to the bibliography.

But the growth of scientific output has continued unabated such that in more recent years an ever increasing chorus of voices crossing all disciplines may be heard expressing alarm at the accretion rate of records in fields of science. Indeed many scholars have elaborated the theory of the exponential growth of scientific knowledge and literature. The following quotes illustrate the existing concern.

"The world's scientific community is presently generating a flood of technical literature, and much of it is not getting into the hands of people who could use it." (5)

"As the mass of information produced by scientists and engineers increases (at rates estimated to be exponential), the inadequacies of available methods of handling information are revealed: dissemination is delayed because publication media are backlogged: retrieval is delayed because the time needed to search for information is a function of the amount of information to be searched." (6)

Senator Hubert H. Humphrey in a televised introduction to a panel discussion presented in connection with the 1962 Convention of Special Libraries Association stated that the world is in a midst of a "Revolution in Information."

"The ironic fact is that the sheer mass of information almost defeats our efforts to find what we are looking for. The result is that a vast amount of information is unfortunately, not really at our command. It may be known, but inaccessible; or it may be unknown or perhaps lost or forgotten." (7)

Francis Bello earlier expressed a similar sentiment and suggested the obvious means for wrestling with the problem.

"Mankind is learning things so fast that its a problem how to store information so it can be found when needed.

Not finding it costs the U.S. over \$1 billion a year.
Now machines are being called on for help." (8)

Currently there exists no quantitative means for accurately measuring the output of technical literature. However this does not preclude the possibility of obtaining some partial definition of it. It may be said that there are now somewhere between 50,000 and 100,000 technical journals being published in more than sixty languages. New journals keep appearing at the rate of at least two a day. This year's journals will carry between one million and two million significant technical articles with a yearly increase of five to eight per cent. (9), (10) And what is more astounding is the fact that the world's supply of recorded knowledge more than doubles (11) every 15 years. To this must be added an increase in content so vast it can not be measured.

In view of the foregoing a literature crisis might be defined as: an explosion or revolution producing at a geometrical or exponential rate piles, masses, mountains or floods of information to bogg down, inundate, bury or even drown poor humanity. Such a definition is exaggeratory only to the extent of its redundant terms, beyond this it is accurate. Even the famed mathematician and philosopher, Yehoshua Bar-Hillel, who rejects the urgency and the existence of a crisis in science, recognizes (12) the possible crisis in the fields of patents, law and library acquisitions. Suffice it to say that an information crisis does exist. Beyond the ominous declarations of competent scientists, the incontrovertible fact remains that millions of dollars are being spent annually by the National Science Foundation, the three military services, with particular emphasis by the U.S. Air Force, the Council of Library Resources, commercial organizations and

a handful of interested academicians in developing systems that will analyze, store, and, on demand, accurately retrieve what has been published.
(13), (14)

B. Advent of Mechanized Information Retrieval

The rapid escalation in recorded knowledge has had the following irreversible effects upon man as a user of information.

(A) Limited by his finite capacity for absorption and retention, man is forced to specialize.

(B) The scope of his specialty is shrinking as man finds it increasingly more difficult to keep informed and current.

(C) Man desires more specific and detailed responses to his information needs.

(D) Attendant to the foregoing is the matter of timeliness. Man wants to find the information he is searching for more rapidly.

Man is not alone, repositories of recorded knowledge are also being effected.

(A) Libraries sample and offer less material because of the enormity of recorded knowledge.

(B) The traditional library has given way to specialized libraries and specialized information centers in response to man's more detailed searches.

(C) Libraries and specialized information centers are applying modern machines to expedite the processing, searching and delivery of information.

In essence the inordinate size of recorded knowledge has precipitated a new discipline. Some call it "librarianship", others call it "documentation science" or "information science". Whatever the label, its reference is clear; the effective management of man's heritage of information.⁽¹⁵⁾

It transcends traditional librarianship by featuring the skills combination of the librarian, the documentalist, the publisher, the writer, the editor, and very important - the scientist and engineer. That⁽¹⁶⁾ puts everyone into the game. The interest of this multidisciplinary corps is in the handling of specialized knowledge. Its exponents seek to make information available - quickly, conveniently, accurately - to the man who seeks to add to knowledge or to draw upon it. The embodiment of this aspiration logically leads to an information storage and retrieval system. And for clarity's sake the definition of such a system seems appropriate.

"A storage and retrieval system in its simplest terms is an organized method for putting items away in a manner which permits or facilitates their recall or retrieval from storage. This definition, although essentially circular, is intended to establish that a storage and retrieval system must be considered as a single system and not as a storage system plus a retrieval system."⁽¹⁷⁾ In other words, the method of storage determines to a considerable extent the possibilities of retrieval. Conversely, the requirements of retrieval establish the range of methods of storage. Continuing, "mechanized" information retrieval merely implies the utilization of machines (from the simple to the complex) to augment a storage and retrieval system.

Experience in mechanized information retrieval indicates that, in general, it must handle these types of material:

- (a) large files or collections of documents;
- (b) complex subject matter, usually containing many subjects of potential interest;
- (c) scattered information, usually requiring selection from among many documents and subsequent correlation.

Having explained that mechanized information came into being in order to ameliorate the conditions precipitated by the "Information Explosion" and having stated the types of material a mechanized information retrieval system handles, it is appropriate to discuss the component operations of such a system. Further, a discussion of the component operations of an information retrieval system continues the deductive manner (first stated in the introduction) in which this thesis is being developed. Parenthetically this "modus operandi" - proceeding from a general background of information retrieval to the specifics of machining indexing - was chosen on the assumption that a complete understanding of the evaluation of machine indexing is contingent upon the reader's having a clear idea how machine indexing, per se, fits into the total information retrieval picture. With this in mind, the component operations or unit operations of an information retrieval system follow in logical sequence.

C. Unit Operations

Any information retrieval system, whether based on traditional library methods or on the most modern machines, involves a series of steps which Allen Kent, Director of the Knowledge Availability Systems Center at the University of Pittsburgh, terms "unit operations." They are: acquisition, analysis, terminology control, recording of results of analysis, storage of source documents, question analysis, searching, and delivery of search results. Except for a description of each, only the second unit operation - analysis - will be discussed at any length.

It is worth noting that these unit operations can not be performed to precise specifications, that is, it is difficult to delineate precisely where one begins and the other ends. Additionally, some may be performed simultaneously or at different times. The following columnar list gives (19) Mr. Kent's unique and original description of each.

<u>Unit Operation</u>	<u>Functions Performed</u>
<u>Input</u>	
1. Acquisition	Locating, selecting, ordering and receiving of source documents for a collection.
2. Analysis	Perusal of source documents and selection of points of view (or analytics) that are considered to be of sufficient probable importance to warrant the effort of rendering them searchable in the system.
a. Abstracting	Concentrating the essential qualities of a larger thing - a summary of a publication accompanied by an adequate bibliographic description, to enable the publication to be traced.

- b. Indexing Selection of words or ideas from a graphic record on the basis of well-defined rules, in order to facilitate the identification or selection of desired records after they have been stored.
- c. Classifying Arranging or placing in a class on the basis of similarities or differences.
- d. Extracting Selecting for quotation portions or source material.
- 3. Terminology Control Establishing arbitrary relationships among analytics, e.g., based on similarities among analytics as revealed in dictionary definitions.
- 4. Recording of Results of Analysis on Searchable Medium Use of card, tape, film or other medium on which the analytics are transcribed.
- 5. Storage of Source Documents Physical placement of record in some location, either in original form, or transcribed or copied onto a new medium.

Output

- 1. Question Analysis and Development of Search Strategy Expression of question or problems, selection of analytics in terms of a particular search mechanism, and arrangement into a configuration that represents a probable link between the question as expressed and the records on file as analyzed.
- 2. Conducting of Search Manipulation or operation of the search mechanism in order to identify records in the file.
- 3. Delivery of Results of Search Physical removal or copying of record from file in order to provide it in response to question.

As may be readily inferred from the foregoing list, the term "input" implies the processing of documents into the system by system

operator personnel. Contrariwise, the term output implies the retrieval of data from the system by either system personnel or the user. Also, for simplicity's sake Mr. Kent limits the results of analysis to abstracts, extracts, classifications and indexes. It is recognized that this is not an exhaustive grouping. Results of analysis among other things includes annotations, first paragraphs, last paragraphs, tables of contents, etc.

D. Analysis

Analysis is the most time-consuming, laborious, ill-defined, complex, and undoubtedly the most important input unit operation. Its complexity stems from the fact that it is basically predictive. Naturally, because the analyst - digesting, reducing and making readily searchable the masses of documents or records being put into a system - is essentially attempting to predict the desires of a searcher who is separated from him in time and space. In effect the analyst is a middleman trying to serve two masters: the author by accurately representing his document, and the searcher or user by answering his information needs. It is axiomatic that no analyst can predict every possible point of view or use that could
(20)
be demanded of any document.

Analysis may be formally defined as the detailed examination of anything complex in order to understand its nature or determine its essential features. The specific purposes of this unit operation are:

- (a) to provide a substitute for the original source material - abstract, extract or annotation;

(b) to permit determination as to whether it is useful to obtain and read the source material - abstract, extract or annotation;

(c) to permit identification of source materials that are probably of interest in connection with one or another quest for information - index entries or classification subject headings. (21)

These purposes establish the amount of detail, or depth of analysis as well as the form in which the result is provided, i.e., abstract, index, etc. The results of the analysis function may be represented graphically on a continuum which depicts the degree of processing to which the original document is processed. See Figure 1.

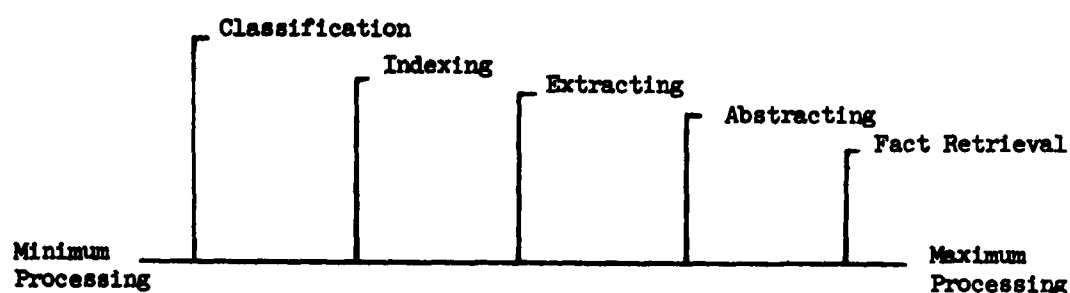


Figure 1: Degree of Document Processing

The four processes which comprise the analysis function are similar in certain respects. First, each of the processes, namely, abstracting, extracting, indexing, and classifying, could be performed independently. Second, all involve the following preliminary steps:

- (a) reading and understanding the document or record being processed;
- (b) deciding what aspects of subject content are important;
- (c) selection of terminology or other symbolisms to designate
(22)
the important aspects of subject content.

As shown in Figure 2, these three basic preliminary operations are followed by different steps, in particular the composing of sentences in writing literary style abstracts, the copying of desired sentences from the source document for extracts, the establishment of index entries in the case of indexes and the assignment of class designations with classification systems.

Third, each process results in loss of information through the elimination of certain individual facts and the corresponding reduction in amount of recorded detail. Depending upon policy, the relative loss of information for each of these four types of output may vary within wide limits. For example, if current policy emphasizes research and development, scientific information will undoubtedly be retained in great detail while other information or facts may be arbitrarily lost. As a general rule, the information loss is minimal with literary style abstracts, particularly when these are written to be informative in character. Next in loss is extracts which lack continuity from sentence to sentence. The greatest loss in information occurs in assigning class designations which, as a rule, do little more than indicate that the subject contents of a given document pertain to one or more generic classification subdivisions. The totality of index entries pertaining to the contents of a document or graphic record usually is more informative than a classification heading. Additionally, although indexing will be discussed in more detail subsequently, it is worth noting

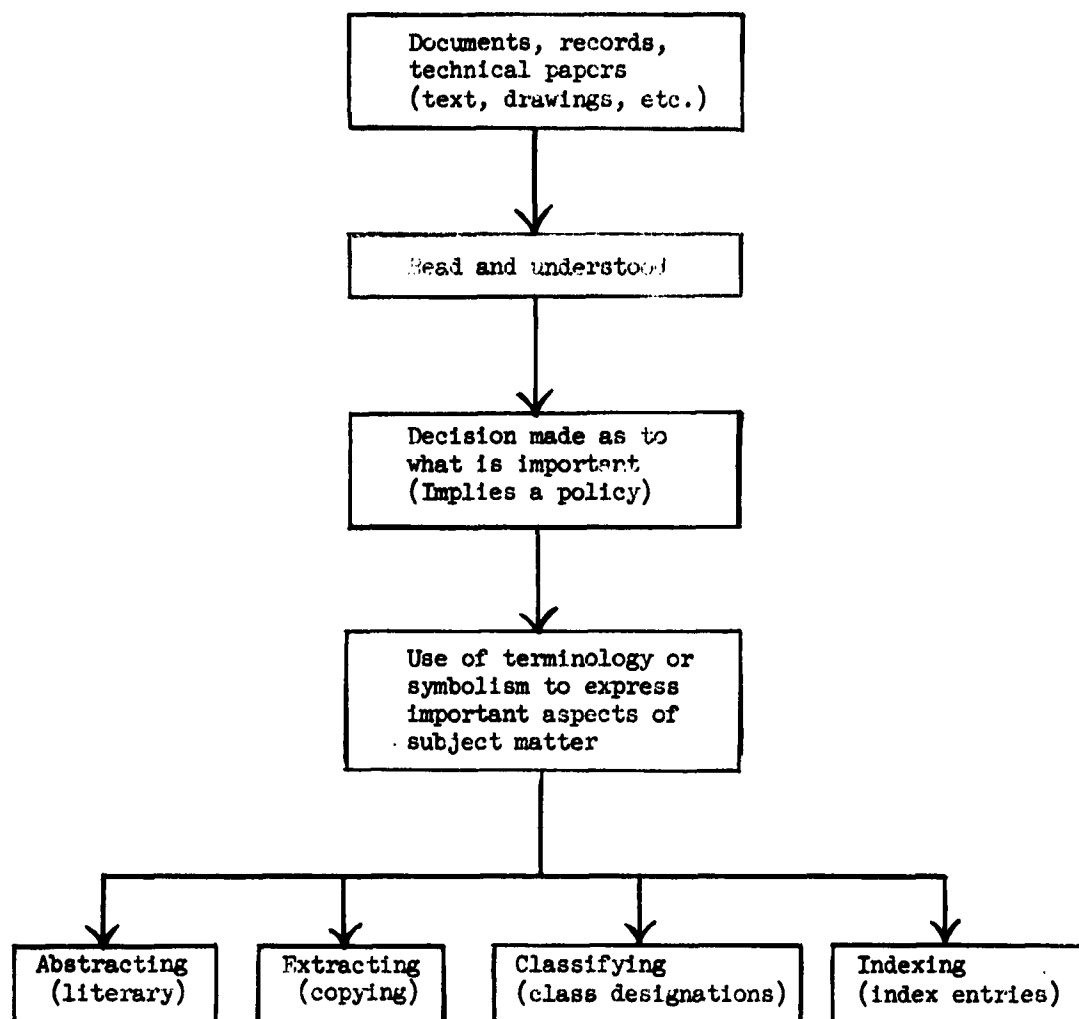


Figure 2: Preliminary Steps Common to Generating Abstracts, Extracts, Classifications and Indexes

now that taken separately each individual index entry, as encountered in alphabetized or similarly arranged lists, usually pertains to a single narrow feature of the subject contents of the original document. (23)

A detailed discussion of the analysis function is beyond the scope of this thesis. The preceding remarks concerning analysis are intended to provide a bridge between the unit operations and indexing.

E. Indexing

The importance of indexing is succinctly expressed by authoritative Calvin Mooers in the following comment:

"The indexing language of an information retrieval system is the intellectual mechanism that makes the system operate. Next to the actual information stored in the collection, it is probably the most important part of the system. The indexing language is the means of mediating between the minds of the customers and the information stored in the collection. It is the bridge, the connection between the users and the information." (24)

Moreover, Allen Kent confirms the bridge idea and expresses the myriad nature of indexing methodology. Here is Mr. Kent on the subject:

"If analysis may be thought of as the cry to open the gates to the flood of human imagination, then indexing may be considered the force that burst the dam and turned the stream of information handling methods into an unconstrained torrent. It is almost as though everyone concerned with specialized information activities has invented an indexing method of his own, and there seems to be more enthusiasm connected with this unit operation fragment than with any other." (25)

This "bridge" - indexing - may be defined as the selection of a set of terms or descriptors from an available vocabulary to describe an

item. Putting it in other words, the process of indexing involves the selection of words or ideas from a graphic record on the basis of well-defined rules; the purpose of indexing is to facilitate the identification or selection of desired documents after they have been sorted and stored. (26)

A number of factors must be taken into consideration in determining index requirements for a particular collection of documents. These include the value of information contained in the collection, its size and use, the required speed and completeness of retrieval, and last but not least, the type or types of questions asked of the collection. (27)

The type of index is directly related to types of search questions. The following table illustrates how one research company matches index requirements with types of search questions. (28)

<u>Types of Requests</u>	<u>Types of Indexes</u>
Report by code	Collection arranged by code - no index required
Report by author	Author index
Report by title	Index to titles, shallow subject index, permuted title index.
General background information	Shallow subject index
Information in document on:	
Specific subject	Alphabetic subject index
Generic subject	Classified index
Coordinate subject, i.e. effect of part of chemical such as functional group on its properties	Coordinate index

Essentially, the general operations in indexing and searching may be reduced to this: A document is read for indexable information; concepts are selected that represent important aspects of this document, aspects that will provide future handles for getting at the document. This indexable information is then translated into standard index language. In searching the index the operation is reversed. The question is first translated into the language of the index. Thereupon the (29) desired document is chosen. This description is elementary, almost infantile, but simple as the operation may seem it is not without serious complications. These complications arise because indexing is predictive (30) and represents a loss in information as stated earlier. Moreover, according to Frederick Jonker it is diffuseness, not just volume, which compounds the indexing and, for that matter, the entire information problem. Dr. Jonker asserts:

"Diffuseness of information implies that items of information may be of interest in a great many problems, uses, and fields of pure and applied science. It implies that these items should be indexed from each viewpoint of potential importance. The diffuseness of an item of information therefore is measured by the number of potential indexing viewpoints or criteria."

"By far the greatest proportion of our present store of knowledge which we need to consult constantly and rapidly is of a highly diffuse nature. This high degree of diffuseness is not something accidental. It is inherent in the very nature of our technological civilization, in the fact that every achievement or finding in any particular field may be of enormous potential interest in almost any other field. It is this crossfertilization which is responsible for the rapid acceleration of progress, which, measured by the time scale of history, almost assumes the proportions of an explosion. Here indexing theory touches the mechanism of progress." (31)

Studies devoted to evaluating the capabilities and limitations of subject indexes have led to the conclusion that they are inadequate in the form in which they have been previously produced to respond to many information requirements, especially those whose scope may specify one or more generic concepts. These inadequacies in subject indexes, mounting concern at the time and cost of index preparation, plus an urgent need to increase the level of efficiency in scientific research and development, have provided the impetus for applying and developing various automatic and semiautomatic devices to accomplish the identification of documents, reports and other technical literature that are of keen interest to a given R and D problem. (32) The net result is that in the past 20 years of increasing information retrieval research many breakthroughs have been achieved, one of which is the successful application of machines to the indexing syndrome.

F. Machine Indexing

1. General

Before discussing the various approaches to automatic or machine indexing a few general comments seem appropriate. First, one of the primary objectives of research in machine indexing is minimizing the human physical and intellectual effort at the "input" end, that is, in processing documents or records into the system. The aim is to derive the appropriate index entries directly from the unedited text of a document. The intermediate human efforts of creating or designating index entries to stand

for a document in storage and coding this data prior to feeding it into a computer are eliminated.

Second, automatic indexing should not be confused, as it often is in the literature, with mechanical storage or with mechanical searching systems. The latter merely store away and search for index entries that were created and coded by human effort. Understand, of course, that a machine index may be used in conjunction with mechanical storage and retrieval systems.

And third, the automatic processing of information requires that the full text of a document be available in some machine-readable form such as punched cards or paper or magnetic tape. Also, the earlier in the production process the text is available in machinable form, the greater the saving. (33) In the case of the permuted index, which will be discussed in great detail subsequently, this aspect is not significant because only the bibliographic data pertaining to a document need be in machine readable form.

2. Approaches to Machine Indexing

There are three basic techniques for machine indexing. They are: the statistical, the linguistic (syntactic) and the semantic.

The statistical approach utilizes the frequency of occurrence of words and their relative distance from one another within the document as the rationale for determining which words from the text are suitable index entries.

The linguistic approach depends upon a modern linguistic analysis of the language structure such as grammar, word order and other distinctive

forms a word may assume. It is based on the assumption that "the ability of the machine adequately to determine information relevance depends strongly on its ability to recognize and manipulate the syntactic structure of the text."^{(34),(35),(36)}

The semantic approach embraces the nature of meaning of words. Strictly speaking, words have "meaning" only as they are used in particular statements. They can be studied individually, but as they are recorded in a dictionary, which offers their linguistic data, they have only typical or possible meaning. Which sense was intended is usually clear from the sentence in which it is used - that is, from the "context".⁽³⁷⁾ "The semantic approach, according to Climensson and his colleagues, evolved from the traditional interest in semantic classification, descriptive terms, cross-indexing of related subject fields, etc."^{(38),(39)}

In practice, most approaches are a combination of the three. It may help to think of these techniques as a continuum of structural information, see Figure 3, ranging from the individual word as a complete and highly structured unit on the one hand to pure meaning on the other hand.⁽⁴⁰⁾

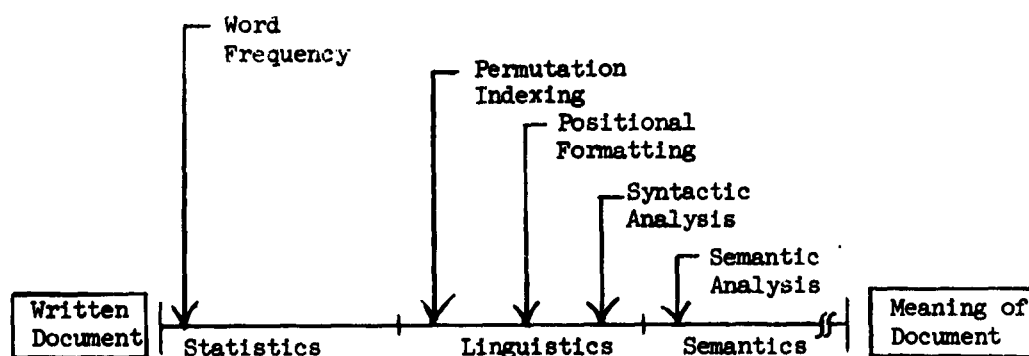


Figure 3: Degree of Format Structure

To date, these approaches have achieved varying degrees of success in producing machine-made indexes from textual material. More sophisticated techniques for automatic selection of indexable matter from unedited text and automatic processing into suitable index entries are currently being developed.

Still there is another approach to machine indexing, one which obviates the complexities of manipulating unedited text since it deals exclusively with the titles of documents. There are, in fact, several methods based on the use of the title only. The most widely used of these is the "Keyword-in-Context" or KWIC index. It is this mode of indexing which will comprise the body of this thesis.

Before discussing KWIC indexing in detail, however, a glance at Figure 4 will graphically reveal how KWIC indexing fits into an information retrieval system and, no less important, the deductive development of this thesis.

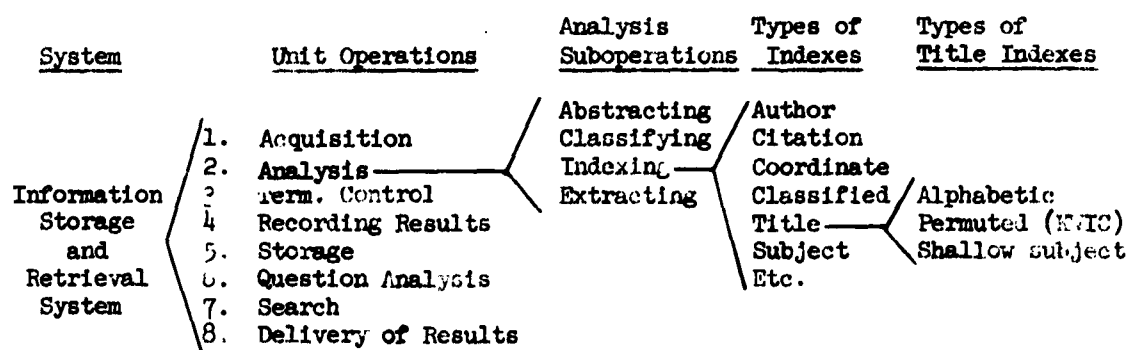


Figure 4: Development of Thesis: From the General to the Specific

II. KEYWORD-IN-CONTEXT (KWIC) INDEX

KWIC is a machine-generated printed index of documents which uses as index entries solely the keywords in the document titles. A KWIC index is frequently referred to as a permuted title index, since the words in the title to a document are rotated or permuted, one at a time. Figure 5 illustrates how the title "Statistical Analysis in Advanced Information Retrieval", might appear in a KWIC or permuted title index.

	<u>Keyword Column</u>
	↓
Statistical Analysis in Advanced Information Retrieval.#	
Statistical Analysis in Advanced Information Re	
cal Analysis in Advanced Information Retrieval.#	
in Advanced Information Retrieval.#	
Statistical Analysis in Advanced In	

Figure 5: KWIC Index Example Using a Single Title

Note that all words in the title with the exception of "in" have been rotated under the keyword column and that each indexing word or keyword appears in alphabetical order.

Normally this title would appear in a KWIC index five times, once for each keyword. On the subject of keywords, H. P. Luhn, the originator of KWIC indexing has this to say:

"Keywords need only be defined as those which characterize a subject more than others. To derive them, simple rules have to be established for differentiating between what is significant and what is 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 (called a stoplist) 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." (41)

A typical stoplist might include, but certainly not be restricted to, the following:

a	been	if	our	these
ago	but	in	put	those
all	by	into	see	thru
also	do	it	she	to
an	for	me	so	too
and	from	my	some	upon
any	he	of	than	we
as	her	off	the	with
at	him	on	then	yet
be	his	or	there	you (42)

The phrase "in-context" suggests that after the keywords are alphabetically ordered, they appear at one location in relation to the other words in the title. Further, the resulting index listing contains the words which surround the keyword, so that the user can read the keyword in its context.

Usually, a KWIC index consists of three parts or sections: Keyword Index, Bibliography and Author Index. Figures 7, 8 and 9 are examples of each. (43) In Figure 7 note that the keywords have been arranged alphabetically to the right of the blank column. A total of 60 characters is printed (this may vary with individual preference), 24 to the left of the keyword and 35 to the right of it. Because the space allotted to titles in a KWIC index proper is fixed, many entries will show only part of a

title. A technique called "wrap around" (also called recirculation or snap-back) is employed to make maximum use of available space. For comparison, observe that Figure 5 does not employ wrap-around, but Figure 6 does. In Figure 5 considerable empty space may appear either to the right or left of an entry - particularly if the keyword is the first or last word in the title. This space may later be filled with more of the context of the keyword. Just how many other words can be shown depends upon the length of the particular keyword and the total number of spaces allotted to the title. (Sixty spaces are used for the titles shown in Figure 7; also, note the use of wrap-around.)

Keyword Column
↓

Statistical Analysis in Advanced Information Retrieval.#	
trieval.#	Statistical Analysis in Advanced Information Re
cal Analysis in Advanced Information Retrieval.#	Statisti
in Advanced Information Retrieval.#	Statistical Analysis
formation Retrieval.#	Statistical Analysis in Advanced In

Figure 6: KWIC Index Using "Wrap-Around"

To the right of each entry in the KWIC index proper appears a reference code. This directs the user to the proper bibliographical entry, where he will find complete information regarding the title, author and source (see Figure 7). Thus the usual KWIC index requires a double look-up to actually find or order a document - once in the KWIC (or author) index and once in the bibliographic index.

PROVIDING INFORMATION RETRIEVAL
MECHANIZED INFORMATION RETRIEVAL
FOR AN INFORMATION RETRIEVAL
SELECTION AND FACSIMILE RETRIEVAL
E NORDIC INFORMATION RETRIEVAL
A BUREAU INFORMATION RETRIEVAL
SCIENTIFIC INFORMATION RETRIEVAL
INFORMATION STORAGE AND RETRIEVAL
INFORMATION STORAGE AND RETRIEVAL
ING RECORDS STORAGE AND RETRIEVAL
ATIONS FOR INFORMATION RETRIEVAL
ICTORY AND STORAGE AND RETRIEVAL
SCIENTIFIC INFORMATION RETRIEVAL
BASIS FOR INFORMATION RETRIEVAL
LANGUAGE ORIENTED RETRIEVAL
LISHED BY A STORAGE AND RETRIEVAL
VAL, ACTIVITY RETRIEVAL
MECHANICAL INFORMATION RETRIEVAL
OURN OF INFORMATION RETRIEVAL
MECHANICS OF STORAGE AND RETRIEVAL
DEVELOPING INFORMATION RETRIEVAL
LEMENTS FOR INFORMATION RETRIEVAL
TURE OF INFORMATION RETRIEVAL
SCIENTIFIC INFORMATION RETRIEVAL
THEIR'S INDEXING AND RETRIEVAL
ZATEGORING INFORMATION RETRIEVAL
INFORMATION STORAGE AND RETRIEVAL
TIVE DISSEMINATION AND RETRIEVAL
F DOCUMENT STORAGE AND RETRIEVAL
IN RETRIEVAL
FOR LARGE INFORMATION RETRIEVAL
WITHIN TWO INFORMATION RETRIEVAL
MECHANIZED INFORMATION RETRIEVAL
INFORMATION RETRIEVAL
MULTIPLE CARD ABSTRACT RETRIEVAL
EFFICIENCY OF STORAGE AND RETRIEVAL
ACTIVITY INFORMATION RETRIEVAL
ACTERS IN THE DESIGN OF RETRIEVAL
DESK-SIZED INFORMATION RETRIEVAL
LY-LEVEL DOCUMENT INFORMATION RETRIEVAL
DOCUMENT STORAGE AND RETRIEVAL
REPORT ON RETRIEVAL
MICROFILM STORAGE AND RETRIEVAL
CULATING INFORMATION RETRIEVAL
INFORMATION STORAGE AND RETRIEVAL
DOCUMENT IMAGE RETRIEVAL
PERFORMANCE OF INFORMATION RETRIEVAL
ANOMORPHIC RETRIEVAL
EFFICIENT INFORMATION RETRIEVAL
NOTICED INFORMATION RETRIEVAL
OF JOURNALING INFORMATION RETRIEVAL
BUILDING AN INFORMATION RETRIEVAL
AUTOMATION OF DOCUMENT RETRIEVAL
INFORMATION FILING AND RETRIEVAL
OF A SIMPLE INFORMATION RETRIEVAL
RANGE PUNCH CARD DATA RETRIEVAL
AN-GSU-SH INFORMATION RETRIEVAL
INFORMATION STORAGE AND RETRIEVAL
FOR DOCUMENT STORAGE AND RETRIEVAL
FOR MANAGING INFORMATION RETRIEVAL
ORRELATIVE INDEXING AND RETRIEVAL
NUAL FOR A PUNCHED CARD RETRIEVAL
TRUCTION OF INFORMATION RETRIEVAL
INATION IN INFORMATION RETRIEVAL
A FLOW STORAGE AND RETRIEVAL
ME COMMUNITY INFORMATION RETRIEVAL
EFFICIENCY OF INFORMATION RETRIEVAL
UNIT, THE INFORMATION RETRIEVAL
- JEM USE INFORMATION RETRIEVAL
A PHOTO-MECHANIC INFORMATION RETRIEVAL
TION MANUAL INFORMATION RETRIEVAL
MECHANIZED INFORMATION RETRIEVAL
F PRACTICAL INFORMATION RETRIEVAL
AND/OR AN INFORMATION RETRIEVAL
AS KNOWLEDGE INFORMATION RETRIEVAL
TECHNICAL STORAGE AND RETRIEVAL
MONSTRATING INFORMATION RETRIEVAL
TACHMENTS INFORMATION RETRIEVAL
ITIONS FOR A STORAGE AND RETRIEVAL
LOCATION OF STORAGE AND RETRIEVAL
A FLOW STORAGE AND RETRIEVAL
INFORMATION RETRIEVAL
TO CARDS AND RETRIEVAL
P- INFORMATION RETRIEVAL
ASTEP PHYS TECHNIQUE RETRIEVAL
OR, MACHINE RETRIEVAL
INFORMATION RETRIEVAL
INFORMATION STORAGE AND RETRIEVAL
R MEGA ITEM INFORMATION RETRIEVAL
ING INFORMATION RETRIEVAL
CONCEPT OF INFORMATION RETRIEVAL
NENLOGY IN INFORMATION RETRIEVAL
INFORMATION STORAGE AND RETRIEVAL
CONCEPTS OF INFORMATION RETRIEVAL
INFORMATION STORAGE AND RETRIEVAL
ENGINEERING INFORMATION RETRIEVAL
IN EQUIPMENT RETRIEVAL
SYSTEMS OF INFORMATION RETRIEVAL
INFORMATION RETRIEVAL
USTICS AND INFORMATION RETRIEVAL
E DESK INFORMATION RETRIEVAL
SPRING INFORMATION RETRIEVAL
F SYSTEMS INFORMATION RETRIEVAL
EQUIPMENT INFORMATION RETRIEVAL
TY YEARS INFORMATION RETRIEVAL
INATION AND INFORMATION RETRIEVAL
CIVILITIES INFORMATION RETRIEVAL
YSTACK INFORMATION RETRIEVAL
INFORMATION RETRIEVAL
INFORMATION RETRIEVAL
DUCTION TO INFORMATION RETRIEVAL
COMPUTERS TO INFORMATION RETRIEVAL
YMBOLS IN INFORMATION RETRIEVAL
TOWARDS INFORMATION RETRIEVAL
F COUNTRIES INFORMATION RETRIEVAL
ATIONS FOR INFORMATION RETRIEVAL
ESSENTIALS FOR DOCUMENT RETRIEVAL

[illegible][illegible]

VERBAL SYSTEM/
 PRINCIPLES OF THE THEORY
 ANALYSIS OF BELIEF
 MECHANISM AND TIME
 ECONOMY A
 ON RELEVANCE, PROBABILIST
 MATHEMATICAL THEORY
 SPECULATIONS
 A CLASS OF SEA
 A CASE HISTORY IN AUTOMATA
 SHOWS
 THE USE OF MACHINE
 PRIME NUMBER
 SUBJECT A
 TIME, COST AND VALUE
 ZETTER-TO-EDITOR/ PRIME
 CLASSIFICATION WITH A PE
 RELATIVE MERITS OF GENER
 MILLS - A RELATIVELY L
 ZETTER-TO-EDITOR/ PRIME
 THE INTERRELATIONSHIP OF
 CONFERENCE ON MULTIPLE A
 ZETTER-TO-EDITOR/ PRIME
 A MACHINE LANGUAGE FOR D
 MANIPULATION
 PATTERN, STRUCTURE, AND
 REFERENCE MANUAL INDEX O
 A SURVEY OF LANGUAGES AN
 CURRENT DEVELOP
 COMPUTER, REA
 A COMPUTER WILL BE USED
 PACIFIC OCEAN ORGANIZATI
 ORGANIZATION OF CHEMICAL
 SOME LINGUISTIC
 N
 THE ALPH
 STORING NUMERICAL VALUES
 ASSOCIAT
 THE UTILIZATION OF SYST
 SYSTEMS OF
 MADE IN
 A POLYDIMENSIONAL
 CHEMICAL
 INTERRELATIONSHIPS AS
 THE LIBRARY OF CONGRESS
 THE RESEARCHER AND
 TOWARDS SEARCHING
 THE ASSOCIATIO
 A DISCUSSION OF
 AS
 EFFICIENCY - A PARAMOUNT
 A TECHNICAL INFORMATION
 A GENERALIZED COMPUTER
 CA RESULTS INDEX BACKED
 THE EFFICIENCY OF CERTAI
 ON KEEPING UP WITH KEEP
 THE THESAURUS
 THE USE OF ELECTRONIC CO
 AN APPLICATION OF THE P
 REPORT ON STUDY OF STATU
 PROGRESS REPORT
 PROGRESS IN THE USE OF T
 LOGUM
 THE IDENTIFICATION OF CO
 ON THE PROBLEM OF TESTIN
 PIONEERING IN MACHINE L
 BIBLIOGRAPHY ON THE MECH
 THE C
 THE FUNCTION OF
 ON THE UNION OF CLASSE
 USES OF THE C
 A STUDY OF THE FUNDAMENT
 NUTRITION. * A SCHEME FOR
 APPLICATIONS OF COMPUTER
 PROBABILISTIC INDEXING. *
 SOME
 PERSPECTIVES IN
 COMBINATORIAL
 SURVEY OF PR
 MULTIPLE ASPECT SE
 COMPUTER SEARCHING OF NO
 / SELECT A COMPUTER-INTEN
 RATION
 ANNOTATED-BIBLIOGRAPHY
 UNIV. DATA AUTOMATION S
 /MUSIC/
 A PHOTO- MAGNETIC SYSTEM
 / BIBLIOGRAPHY/
 / IBM-705/ ZERENCH/ AN E
 THE NEED FOR A FACETIC C
 PRESENTLY AVAILAB
 THE NEED FOR A THESAURUS
 CLASSIFICATION AT TORRIN
 / IBM 702/2/

[illegible]

[illegible]

Figure 8: Typical Bibliography Page

BUCKAC-58-SIS BUCK RE
STUDIES IN INFORMATION STORAGE AND RETRIEVAL. I. ON THE
OF CODE INDICES IN CODING.
UNIVERSITY OF WISCONSIN, MATHEMATICS RESEARCH CENTER,
MADISON, WIS., MCR TECHNICAL SUMMARY REPORT NO. 55,
AL-207325, 1954, 21P., OALS. AMERICAN DOCUMENTATION, 1971, 12,
NO. 3, 165-71.

BULLWE-52-PVF BULL WE
PROBLEMS OF VOCABULARY FREQUENCY AND DISTRIBUTION.
BIOGEOGRAPHIC PAPER, JUNE 1952, P. 1-10.

BULLWE-55-SPW BULL WE AFRICA U. TECHNIQUE U.
SOME PROBLEMS OF THE "MORL."
MACHINE TRANSLATION OF LANGUAGES, PUBLISHED JOINTLY BY THE
TECHNOLOGY PRESS OF M.I.T. AND WILEY, NEW YORK, 1954,
86-103.

BUNNLE-60-SPM BUNNOR LE
STUDY OF AND PROPOSAL FOR A MECHANIZED INFORMATION RETRIEVAL
SYSTEM FOR THE MISSILES AND SPACE SYSTEMS ENGINEERING
LIBRARY.
DOUGLAS AIRCRAFT CO., SANTA MONICA DIVISION, SANTA MONICA,
CALIF., REPORT NO. SM-37418, MAY, 1960, 62P.

BURGE-49-WS BURKHARD JE
THE WATERLOG OF SCIENCE.
REVIEW OF DOCUMENTATION, 1949, 16, 94-7.

BURGU-50-IRG BURGESS J.
INFORMATION RETRIEVAL GENERATION OF RANDOM CODE.
IBM 650.
IBM CORP., EASTERN REGIONAL OFFICE, 425 PARK AVE., NEW YORK
22, NY, AN ADDENDUM TO THE SCIENTIFIC APPLICATIONS INDEX,
(NO).

BURJW-62-CPI BURGESS J. BUSHNELL RC
CALCULATION OF THE PHI FUNCTION, PART OF THE NON-PULIAN URN
STATISTICAL DISTRIBUTION FORMULAE, USED TO CALCULATE
SPURIOUS DROP-OUT IN INFORMATION RETRIEVAL APPLICATIONS.
IBM CORP., DATA PROCESSING DIVISION, 340 S. BROADWAY,
AKRON 8, OHIO, 1962.

BURKAW-54-ALM BURKS AM WARREN DW WRIGHT JR
AN ANALYSIS OF A LOGICAL MACHINE USING PARENTHESIS-FREE
NOTATION.
RTAC, 1954, 8, NO. 46, 53-7.

BURKCA-48-UHS BURKHARD CA
THE USE OF HAND-SORTED PUNCHED CARDS IN SMALL FILES.
PAPER PRESENTED AT THE AMERICAN CHEMICAL SOCIETY MEETING,
SPRING, 1948.

BURKCA-53-SFT BURKHARD CA
STORING AND FINDING TECHNICAL INFORMATION WITH PUNCHED
CARDS.
PAPER PRESENTED AT THE AMERICAN CHEMICAL SOCIETY MEETING,
FALL, 1953.

BURKCA-58-UPC BURKHARD CA
THE USE OF PUNCHED- CARD TECHNIQUES IN PREPARING REPORTS,
PAPERS, AND BOOKS.
CHAPTER 4, IN PUNCHED CARDS, THEIR APPLICATIONS TO SCIENCE
AND INDUSTRY, 2D EDITION, REINHOLD, NEW YORK, 1958.

BURKJ-57-MMR BURKIG J.
MAGNACARD - MAGNETIC RECORDING STUDIES.
INSTITUTE OF RADIO ENGINEERS WESCON CONVENTION RECORD, 1957
PART 4.

BURTIK-57-UUL BURNETT JR
THE USE OF UNITERMS IN LARGE COLLECTIONS.
CONFERENCE ON MULTIPLE ASPECT SEARCHING FOR INFORMATION
RETRIEVAL, WASHINGTON, DC, FEBRUARY 12-13, 1957, ASTIA,
ARLINGTON, VA., AD-147491, 1957, 83-90.

BURUNG-55-LCS BURTON NG LICKLIDER JCR
LONG-RANGE CONSTRAINTS IN THE STATISTICAL STRUCTURE OF
PRINTED ENGLISH.
AMERICAN JOURNAL OF PSYCHOLOGY, 1955, 68, 650-3.

BUSAR-57-EMM BUSA R
ELECTRONIC METHODS IN THE MECHANIZATION OF LINGUISTIC
ANALYSIS.
NACHRICHTEN FÜR DOKUMENTATION, MARCH 1957, 8, NO. 1, 20-6.

BUSAR-58-UPC BUSA R
THE USE OF PUNCHED CARDS IN LINGUISTIC ANALYSIS.
CHAPTER 16, IN PUNCHED CARDS, THEIR APPLICATIONS TO SCIENCE
AND INDUSTRY, 2D EDITION, REINHOLD, NEW YORK, 1958.

BUSBJC-61-SSA BUSBY JC
SASSIF (SELF ADJUSTING) SYSTEM OF SCIENTIFIC INFORMATION
FLOW.
AMERICAN INSTITUTE OF BIOLOGICAL SCIENCES, WASHINGTON, DC,
1961, 26P.

BUSHMC-62-SDI BUSHNELL RC
SELECTIVE DISSEMINATION OF INFORMATION (SDI) PROGRAM
PACKAGE FOR THE IBM 1401.
IBM CORP., DATA PROCESSING DIVISION, 340 S. BROADWAY,
AKRON 8, OHIO, 1962.

BUSHTM-62-CMF BUSH TH
COMPUTERS MAY FREE LAWYERS FROM MUCH DULL, ROUTINE
RESEARCH.
THE WALL STREET JOURNAL, NEW YORK, AUGUST 14, 1962.

BUSHV-45-WMT BUSH V
AS WE MAY THINK.
ATLANTIC MONTHLY, JULY 1945, 176, 101-8. ALSO CHAPTER 2, IN
ENDLESS HORIZONS, PUBLIC AFFAIRS PRESS, WASHINGTON, DC,
1946.

BUSHV-54-RSC BUSH V
REPORT TO THE SECRETARY OF COMMERCE BY THE ADVISORY
COMMITTEE ON APPLICATION OF MACHINES TO PATENT OFFICE
OPERATIONS.
US DEPT. OF COMMERCE, SALES AND DISTRIBUTION DIVISION,
WASHINGTON, DC, 1954.

BUSHV-54-STH BUSH V
SOME THINGS WE DON'T KNOW.
PAPER PRESENTED AT THE RICE INSTITUTE, HOUSTON, TEXAS,
MAY 6, 1954.

BUSHV-54-TRT BUSH V
TODAY'S RESEARCH AND TOMORROW'S WORLD.
PAPER PRESENTED BEFORE GUESTS OF THE BOARD OF DIRECTORS
OF STANFORD RESEARCH INSTITUTE, JANUARY 20, 1954. ALSO
PAMPHLET PUBLISHED BY STANFORD RESEARCH INSTITUTE,
MENLO PARK, CALIF.

BUSHV-55-CWD BUSH V
COMMUNICATIONS - WHERE DO WE GO FROM HERE.
MECHANICAL ENGINEERING, APRIL 1955, 77, 302-4. ALSO
ELECTRICAL ENGINEERING, MAY 1955, 74, 367-70.

BUSHV-55-MK BUSH V
FOR MAN TO KNOW.
ATLANTIC MONTHLY, AUGUST 1955, 196, 29-34.

BUSHV-57-PC BUSH V
PROFESSIONAL COLLABORATION.
SCIENCE, JANUARY 11, 1957, 125, NO. 3237, 49-54.

BUSHWE-62-ISR BUSHOR WE
INFORMATION STORAGE / RETRIEVAL.
ELECTRONICS, JUNE 29, 1962, 35, NO. 26, 39-62.

BUSNY-57-PA BUSNY G
A MULTIDIMENSIONAL ASSOCIATION.
SANTA TRAP, KATI, N. TECH. MEM. NO. 27, 197, 1972.

BUSNY-57-PAI BUSNY G
A NEW APPROACH TO THE INFORMATION RETRIEVAL PROBLEM.
SANTA TRAP, KATI, N. TECH. MEM. NO. 27, 197, 1972.

BUTCH-60-PUM BUTTON CH CROSSWAMP RJ KINNEY JR
MURPHY JR
PARTS IMAGE MAINTENANCE PROGRAM (PIMP).
IBM CORP., DATA SYSTEMS DIVISION, P.O. BOX 112, YTD,
1960, 44P.

CAHNJN-62-FIT CAHN JN
B.I.T.S. - BUREAU OF INFORMATION & TECHNOLOGY AND
SCIENCE PROPOSED AS A KEY STATION, OF RETRIEVAL
NETWORK, OF INFORMATION SERVICES TO THE FREE
ENTERPRISE SYSTEM.
PROCEEDINGS OF THE IJC ENGINEERING INFORMATION SYMPOSIUM,
NEW YORK, NY, JANUARY 17, 1962, 45-50.

CAHNJN-62-SIS CAHN JN
A SYSTEM OF INFORMATION SYSTEMS.
PAPER PRESENTED AT THE 4TH INTERNATIONAL INFORMATION
STORAGE AND RETRIEVAL, AMERICAN UNIVERSITY, WASHINGTON, DC,
FEBRUARY 1962.

CAMNL-57-LUE CAHN L
EXPERIMENTAL USE OF ELECTRONIC COMPUTERS IN PROCESSING
PICTORIAL INFORMATION.
PROCEEDINGS OF THE EASTERN JOINT COMPUTER CONFERENCE, 1957.

CAMMS-62-IRE CAMMENG
INFORMATION RETRIEVAL.
IBM CORP., DATA PROCESSING DIVISION, 1 CHERRY ST.,
GRAND RAPIDS 6, MICH., 1962.

CAMPJD-53-UNC CAMPBELL DJ
THE USE OF NOTCHES IN CARDS AS A MEANS OF SIGNALLING
INFORMATION.
JOURNAL OF DOCUMENTATION, DECEMBER 1952, 9, 224-5.

CAMPJD-61-DPO CAMPBELL JL CARON HS
DATA PROCESSING BY OPTICAL COINCIDENCE.
SCIENCE, 1961, 133, NO. 3461, 1533-8.

CAMPKC-50-PEC CAMPBELL JC CAMPBELL BF
PUNCHED CARD CODE FOR GENERAL ORGANIC RESEARCH.
INDUSTRIAL AND ENGINEERING CHEMISTRY, AUGUST 1956, 145P-61.

CAPKAF-57-UDD CARRIC AF
USE OF THE DASHW DATA PASTER, 1. AUTOMATIC NUMBER AND DATA
POSTING MACHINE.
CONFERENCE ON MULTIPLE ASPECT SEARCHING FOR INFORMATION
RETRIEVAL, WASHINGTON, DC, FEBRUARY 12-13, 1957, ASTIA,
ARLINGTON, VA., AD-147491, 1957, 59-61.

CARNCR-62-RRR CARNER CR
READY REFERENCE CENTER IN LIBRARY TWENTY-ONE.
LIBRARY JOURNAL, 1962, 17, NO. 6, 1557-8.

CARPA-62-SCP CARPENTER RA BULLER CC FINDLEY LI
A SYSTEM FOR THE CORRELATION OF PHYSICAL PROPERTIES AND
STRUCTURAL CHARACTERISTICS OF CHEMICAL COMPOUNDS WITH
THEIR COMMERCIAL USES.
AMERICAN DOCUMENTATION, APRIL 1962, 10, NO. 2, 135-43.

CARRCE-52-NRC CARROLL CE
NATIONAL RESEARCH COUNCIL - NATIONAL BUREAU OF STANDARDS
INFRARED PUNCHED CARD SYSTEM.
NATIONAL BUREAU OF STANDARDS, WASHINGTON, DC, 1952.

CARRHL-60-UMA CARROLL HL
USE OF MICROFILM APERTURE CARDS IN IBM EQUIPMENT.
PROCEEDINGS OF THE NATIONAL MICROFILM ASSOCIATION, 9TH
ANNUAL CONVENTION, 1960, 118-22.

CARRJM-58-PTC CARR JM SRD
PRECISION OF TRANSLATION BY COMPUTERS FROM OTHER LANGUAGES
TO RUSSIAN.
COMPUTERS AND AUTOMATION, 1958, 7, NO. 3, 12.

CARRJM-59-VCC CARR JM SRD PERLIS AJ ROBERTSON JR
SCOTT MR
A VISIT TO COMPUTATION CENTERS IN THE SOVIET UNION.
LANGUAGE TRANSLATION.
COMMUNICATIONS OF THE ACM, 1959, 2, NO. 6, 14-5.

CASERS-46-PCT CASEY RS BAILEY CF COX GJ
PUNCHED CARD TECHNIQUES AND APPLICATIONS.
JOURNAL OF CHEMICAL EDUCATION, 1946, 23, 495-9.

CASERS-58-PCT CASEY RS /ED./ PERRY JM /ED./ BERRY MR /ED./
KENT A /ED./
PUNCHED CARDS, THEIR APPLICATIONS TO SCIENCE AND INDUSTRY,
2D EDITION, REINHOLD, NEW YORK, 1958, 697P.

CASERS-61-KSC CASEY RS
KSC SERVICE FOR CHEMICAL SEARCHERS.
AMERICAN DOCUMENTATION, 1961, 12, NO. 1, 75-6.

CASS-62-BRT CASS S
BOOK REVIEW OF TAUBER'S "COMPUTERS AND COMMON SENSE" - THE
MYTH OF THINKING MACHINES.
AMERICAN DOCUMENTATION, 1962, 12, NO. 2, 223-4.

CASSWG-59-ALT CASS NG
AUTOMATIC LANGUAGE TRANSLATION.
AUTOMATION AND AUTOMATIC EQUIPMENT NEWS, 1959, 5, NO. 3, 148.

CAMKAE-62-CPT CAWKELL AE
CLASSIFICATION AND RETRIEVAL OF TECHNICAL INFORMATION.
1. CONSIDERATIONS LEADING TO THE CHOICE OF NOTCHED
CARDS.
WIRELESS WORLD, AUGUST 1962, 68, NO. 4, 352-7.

CECCS-56-SMT CECCATO S MARETTI E
SUGGESTIONS FOR MACHINE TRANSLATION.
INFORMATION THEORY, BUTTERWORTHS, LONDON, 1956.

CECCS-60-LAP CECCATO S
LINGUISTIC ANALYSIS AND PROGRAMMING FOR MECHANICAL
TRANSLATION.
UNIVERSITA DEGLI STUDI DI MILANO, MILANO, ITALY.
RACD TR 60-18, JUNE, 1960. ALSO GORDON AND BREACH, INC.,
NEW YORK, 1961, 246P.

CECCS-61-PEO CECCATO S
PRINCIPLES AND CLASSIFICATIONS OF AN OPERATIONAL GRAMMAR
FOR MECHANICAL TRANSLATION.
PAPER PRESENTED AT THE INTERNATIONAL CONFERENCE FOR
STANDARDS ON A COMMON LANGUAGE FOR MACHINE SEARCHING
AND TRANSLATION, SEPTEMBER 6-12, 1959.
CHAPTER 24, V. 3, PART 2, 603-713. OF ADVANCES IN
DOCUMENTATION AND LIBRARY SCIENCE.
INTERSCIENCE, NEW YORK, 1961.

CETODE-60-PMA CETONE DE /ED./
PROCEEDINGS OF THE MARCH AFENC CONFERENCE ON SCIENTIFIC
AND TECHNICAL INFORMATION.
AIR RESEARCH AND DEVELOPMENT COMMAND, WASHINGTON, DC,
AFBMD TR 60-143, 1960, 134P.

CEZARO-60-AES CEZARILIAN AC LYKOUDES PS TOULOUKIAN YS
ANALYTICAL AND EXPERIMENTAL STUDY OF A METHOD FOR LITERATURE
SEARCH IN ABSTRACTING JOURNALS.
PURDUE UNIVERSITY, LAFAYETTE, IND., TRC REPORT 11, PR-17147P,
1960, 19P.

A typical reference code is shown in Figure 7. It consists of the first four letters of the author's last name, his first two initials, the last two digits of the year of publication, and the initial letters of the first three words of the title of his document. (The following words are disregarded: the, of, in, from, by, at, as, and, an, a, on, or, to, with, for.) The choice of the composition of the reference code is practically unlimited, however. It may be assigned on the basis of report numbers, patent numbers, some arbitrary numbering system, characteristics of the document concerned, or even on the order in which the bibliographical (44) entries are to appear.

Computer manufacturers offer on request detailed instructions, programs and other essential data necessary for producing KWIC indexes. No attempt will be made here to duplicate their effort. The intricate data processing techniques inherent in the preparation of KWIC indexes do not conform with nor advance the aims of this thesis which are rather to subjectively evaluate this mode of indexing and to statistically test the hypothesis upon which it is based. Before concluding the general description of KWIC indexing, however, it is worth noting that in order to produce a KWIC index of titles, bibliographical entries consisting of at least, title, author and source are recorded in machine-readable form, such as punched cards, so that computer processing can produce the printed index with its bibliographical data. The cards making up the master record bibliography, the special word list, and the KWIC program constitute the input to the computer (see Figure 10). Upon processing, the computer determines the status (keyword or common) of each word in a document title by looking it

up in the stop list. All ordering and arranging of the input information is performed by the computer. A magnetic output tape containing the final arrangement of the KWIC index, the bibliography and the author index is (45) fed through a printout device to produce copy for subsequent reproduction.

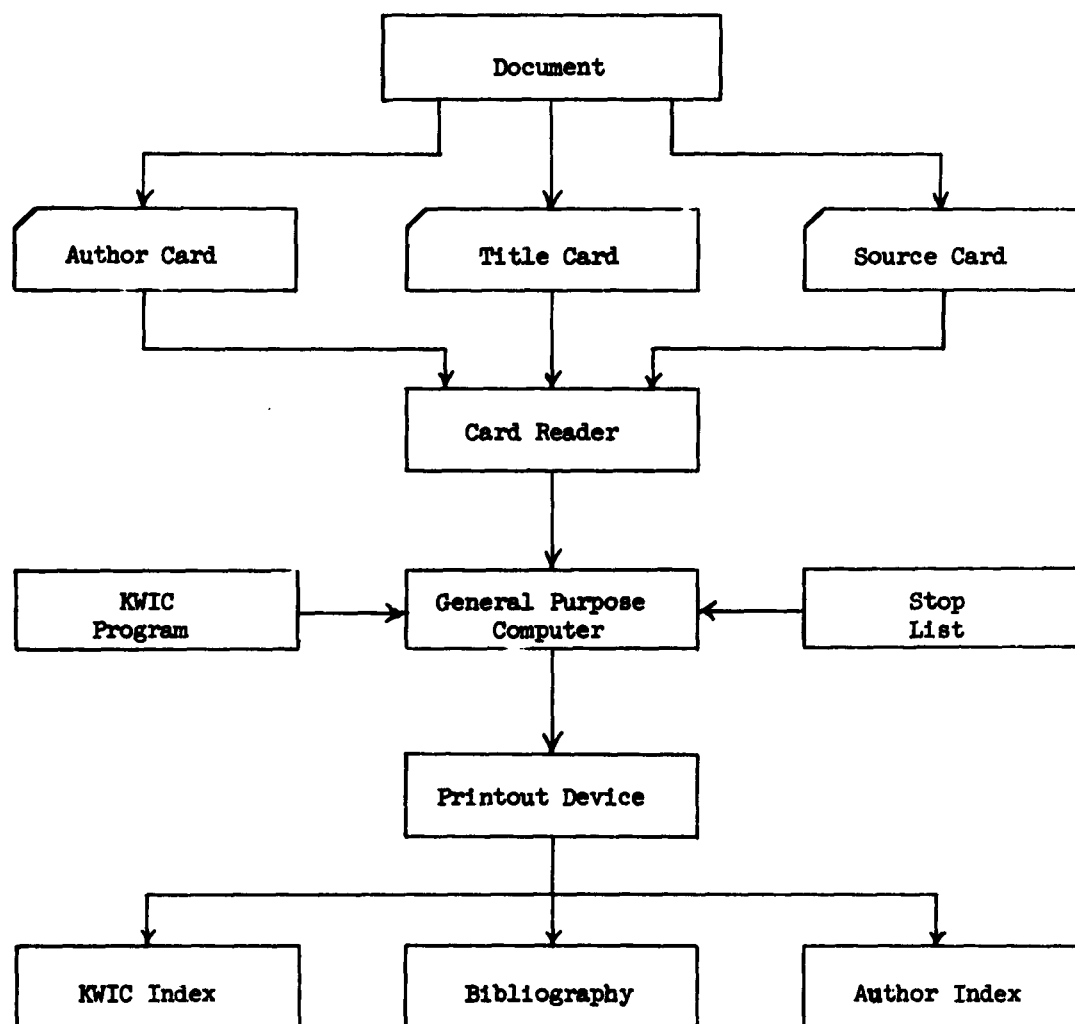


Figure 10: Generation of KWIC Index

III. STUDY OF ADEQUACY OF TITLES AS SEARCH TOOLS

A. General

"If the (KWIC) index is based on titles of documents, its quality depends on how well the authors have composed the titles of their papers", with these words H. P. Luhn almost matter-of-factly stated the hypothesis upon which KWIC indexing is based. ⁽⁴⁶⁾ Essentially, then, the efficacy this mode of indexing is dependent upon whether titles are in fact descriptive. Those who assail KWIC indexing claim that titles are not descriptive, or at least not sufficiently descriptive to merit serious consideration as sources of index entries; its exponents claim the converse. The natural result of this debate has been several studies aimed at assessing the adequacy of titles as search tools.

A study by Montgomery and Swanson showed that "the present indexing of medical literature as published in Index Medicus is, or could be, based largely on examination of titles and word-plus-synonym match of titles to subject headings." Titles were found to be about 50 per cent effective as a basis for judging the relevance of an article to a given ⁽⁴⁷⁾ information need.

Maizell found that in a sample of 25 articles, included both in Physics Abstracts and Chemical Abstracts, the titles alone contained about ^{(48), (49)} 50-70 per cent of the key terms under which the articles were indexed.

In the area of legal literature, Kraft examined 3428 titles from legal research projects, theses and the Index to Legal Periodicals. His findings, among other things, revealed that 64.4 per cent of the title entries contained as keywords one or more of the subject heading words

under which they were indexed; and 25.1 per cent contained logical equivalents. The remaining 10.5 per cent of the title entries had non-descriptive titles. He concluded that a KWIC index of legal titles produces an index which costs less than a subject heading system in both time and cost of production and which ranks high in "findability."⁽⁵⁰⁾

It seems intuitive to this author that highly descriptive titles will, in the main, be found in the technical fields of science and engineering and that less descriptive, or the least descriptive titles, will be found in the social sciences and humanities. As stated this is intuition, and regardless of its plausibility, it is an untested premise. This author decided to test the premise by examining titles from several fields, principally science and engineering, and in so doing continue the overall research effort at determining the adequacy of titles as search tools. In this regard, Kraft's study provided a worthy guide and reasonable point of departure - his procedure was modified slightly for this research. Six published indexes to periodical literature provided the source of titles for the study and the findings for each will be discussed in detail after the following description of the procedure employed.

B. Procedure

The scope of this study as indicated by the following list, spans the subject fields of science, engineering, social science and the humanities.

<u>Index</u>	<u>Subject Field</u>
Engineering Index	Engineering and Science
Applied Science and Technology Index	Engineering and Science
ASM Review of Metal Literature	Engineering and Science
Public Affairs Information Science, Annual Cumulated Bulletin	Economics and Information
Business Periodicals Index	Business
International Index	Social Science and Humanities

In each instance the most recent annual index was chosen on the assumption that each index mirrors the most current periodical literature in its respective field.

From each of these indexes a genuine random sample was taken. Randomness was achieved using numbers taken from tables of random numbers, initially. Later, to expedite the process of locating a desired title, random numbers were computer generated within limits, the limits being the total number of pages in the particular index and the maximum number of titles per page. Randomness was maintained in order that statistical inferences with a desired level of precision might be made from sample data. In this respect previous studies lacked precision or confidence limits because sampling was not conducted statistically.

After each title was randomly chosen, keywords were selected from it by the author using the rules under which a computer would have operated. All words were considered keywords with the exception of stoplist articles, prepositions, conjunctions and other "common" words (the stoplist was previously discussed). For purposes of analysis the titles were, in the context of acceptance sampling, considered as either effective (KWIC-able) or defective (not KWIC-able). Beyond this major dichotomous classification, the KWIC-able titles were grouped into four types. For a graphic illustration

of the logic used in sorting titles, see Figure 11. The examples which follow use titles from the Engineering Index (EI).

TYPE 1: A title which contains a keyword exactly or in some root form as a word in a subject heading of the index.

Examples:

EI Heading: FUEL CELLS

Title: "Fuel Cells and Their Effects on Utilities"

EI Heading: THERMODYNAMICS

Title: "'Laxation' Function Makes Entropy Easier to Understand"

The first title has a word which is the exact same as the subject heading under which it is indexed. The second title has a word "Entropy" which is a subject heading in the Engineering Index, however, "entropy" has a "see" reference to "Thermodynamics". In either case the title contains a keyword exactly or in some root form as a word in a subject heading of the index.

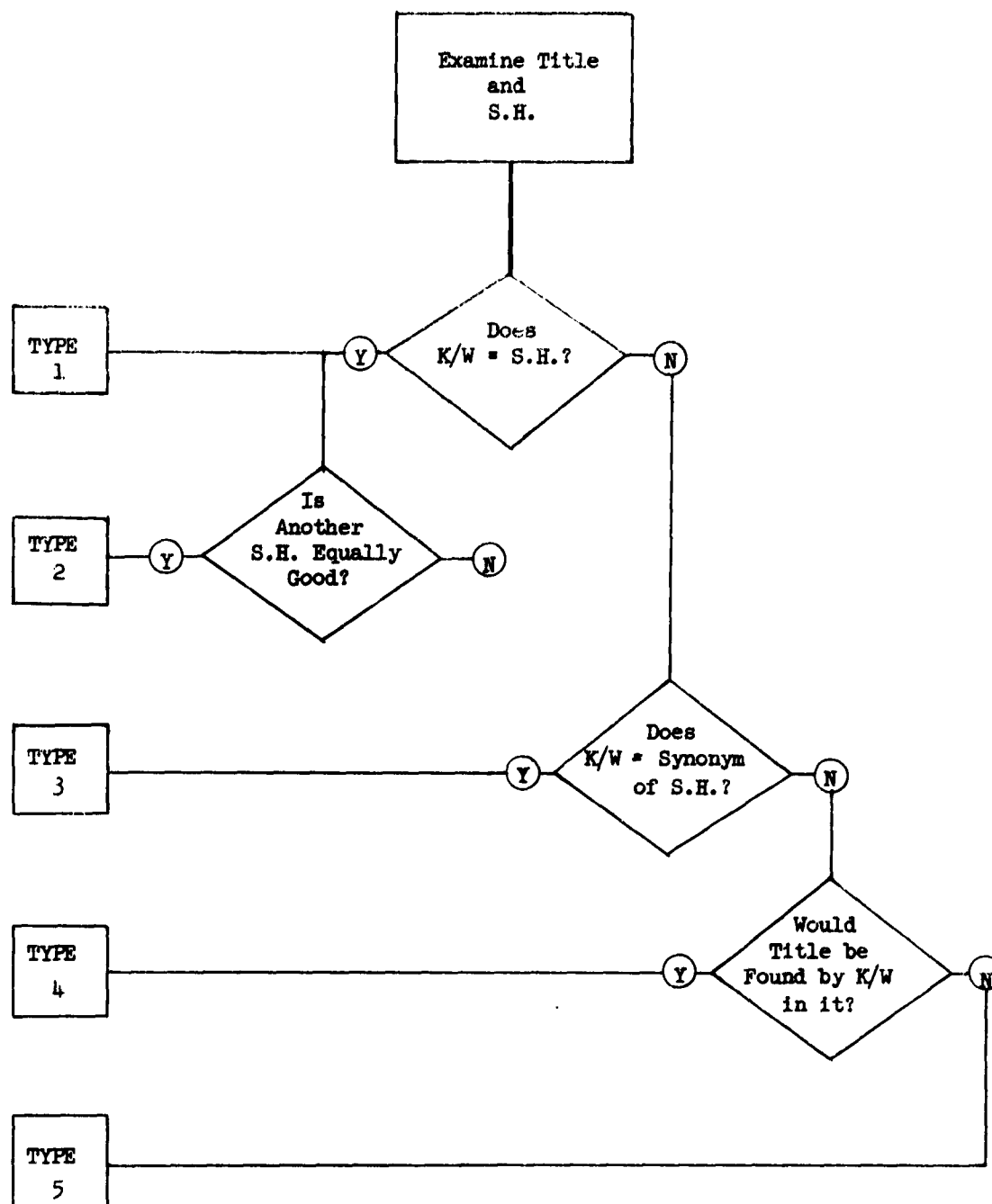
TYPE 2: A Type 1 title which could have been indexed equally well under another subject heading which appears in the title.

Example:

EI Heading: PYROMETRY and OPEN HEARTH FURNACE

Title: "Immersion Pyrometry and Open Hearth Furnace Productivity"

Titles of this type are even more "findable" than those of Type 1, hence very KWIC-able.



S.H. = Subject Heading
K/W = Any Keyword in Title

Types 1, 2, 3, & 4 = Effective Titles
Type 5 = Defective Titles

Figure 11: Logic Followed in Sorting Titles by Type

TYPE 3: A title not classified 1. or 2 but which contains a synonym of its subject heading:

Examples:

EI Heading: PLUTONIUM or URANIUM

Title: "Simultaneously Determining Pu and U in Dissolver Samples"

EI Heading: AUTOMOBILE

Title: "Atomic Age Car Concept"

This classification includes few titles because it is limited to abbreviations as in the first title and bona fide synonyms as in the second title.

TYPE 4: A title not of Types 1, 2 or 3, but which contain keywords that would enable an informed researcher to find it in an obvious manner under a KWIC indexing system.

Example:

EI Heading: HYDROCARBONS, PROCESSING

Title: "Isomerization or Normal Pentane over Platinum Alumina Catalysts"

EI Heading: SPACE VEHICLES, COMMUNICATIONS SYSTEMS

Title: "Transmission of Electromagnetic Waves through Ionized Layer in Presence of Strong Magnetic Field"

As Kraft points out, titles such as these, illustrate the problem inherent in all subject heading systems: How can the searcher "outguess" the

indexed. These titles would be lost if the searcher failed to think of the subject headings "Hydrocarbons, Processing" and "Space Vehicles, Communications Systems"; also the titles were not indexed under any other related subject headings. A KWIC index, however, would index the titles under each of the underlined words, some of which are very descriptive, (51) making them a cinch to find.

TYPE 5: A title not of Types 1, 2, 3 or 4. This is the class of non-descriptive ("defective" in acceptance sampling context) titles.

Examples:

EI Heading: MOTOR SHIPS

Title: "Nordvind"

EI Heading: PETROLEUM PROSPECTING, OKLAHOMA and TEXAS

Title: "Major Play Shaping Up in Palo Duro"

Admittedly this classification scheme is subjective, not so much in separating Types 1 and 2 titles (this is simply a matter of determining the presence or absence of a subject heading for each keyword in the title), but certainly in sorting titles into Types 4 and 5. The author attempted to be unbiased, if indifferent, in the sorting process. Whenever there was reasonable doubt about how to classify a title it was placed in that group (52) which was least advantageous for proving a case for KWIC indexing.

Now that the procedure has been elucidated, the sampling from each index will be discussed in turn.

C. Engineering Index Sample

1. General

The Engineering Index is published annually by Engineering Index, Inc., 345 East 47th Street, New York 17, New York, which "indexes and annotates, selectively, on the basis of engineering significance, the available current technical periodicals received by and permanently housed in the Engineering Societies Library. Included are the regular professional and trade journals, publications of engineering societies, scientific and technical associations, universities, laboratories and research institutions, government departments and agencies, and industrial organizations. Papers of conferences and symposia, separate and non-serial publications of various kinds, and selected books are also covered. Patents are excluded."⁽⁵³⁾

The particular issue chosen was 1962; it includes approximately 31,000 brief abstracts of articles from 1000-odd periodicals and serial publications reviewed. As stated previously it is considered by the author to contain a representative sampling of titles from the fields of science and engineering.

The Engineering Index lists each title only once, under what is presumed to be the most appropriate subject heading, but makes liberal use of the "see" and "see also" reference.

2. Estimation of Sample Size

Before the statistical sample could be performed on this index it was necessary to first get some indication of the required size of the

sample. In probability sampling the size of the sample can be determined mathematically once the desired precision and estimated variability of the result is decided.

With respect to sampling titles from the index, recall that the titles are sorted into two broad classes, KWIC-able and not KWIC-able. For purposes of computation, Q represents the proportion of KWIC-able titles, P represents the proportion of not KWIC-able titles, and p is the estimated proportion of P from the sample. The author chose a margin of error d in the estimated proportion p of titles in P and willingly incurred the small risk α that the actual error was larger than d; that is,

$$\Pr (| p - P | \geq d) = \alpha \quad (1)$$

Again, simple random sampling was assumed; p was taken as normally distributed. Since the sample (n) is very small in proportion to the total number of titles in the index (N), the standard deviation of p is,

$$\sigma_p = \sqrt{\frac{PQ}{n}} \quad (2)$$

Hence the formula that connects n with the desired degree of precision is,

$$d = t \sqrt{\frac{PQ}{n}} \quad (3)$$

where t is the abscissa of the normal curve that cuts off an area at the tails. Substituting the advance estimate p for P and solving for n, (54)
yields,

$$n = \frac{t^2 pq}{d^2} = \frac{pq}{V} \quad (4)$$

where $V = \frac{d^2}{t^2}$ = desired variance of the sample proportion.

Next the author set the precision requirements consistent with the critical result. Parenthetically, the critical result in any study is that result requiring the greatest sample size to achieve the desired precision in view of the estimated variability. (55), (56) Briefly, the specification of precision is that p is to lie in the range $(P \pm 2.5)$, except for a 1 in 20 chance. Kraft's work is used as a basis for an advance estimate p of P . Thus,

$$d = .025, \quad p = .10, \quad \alpha = .05, \quad t = 1.96$$

and

$$n = \frac{t^2 p q}{d^2} = \frac{(1.96)^2 (.10)(.90)}{(.025)^2} = 550 \quad (5)$$

It is worth noting at this point that 550 is an estimate of sample consistent with desired precision and advance estimate, p , of population proportion defective, P . As it turned out, upon sampling from the Engineering Index, the number of not KWIC-able titles was so few that the advance estimate of 10% defective titles proved to be disproportionate, hence all sampling was terminated after 390 titles had been examined.

3. Classification

The results of the sorting of titles are in the table below:

Type	"1"	"2"	"3"	"4"	"5"	Total
Number	324	124	8	50	8	390
% of Total	83.1*	31.8	2.05	12.8	2.05	100.0

* The reader should recall that Type 1 titles include Type 2 titles.

4. Confidence Limits

Since $n p$, or more specifically the number of titles observed in the smaller (defective) class, is small, only 8 and less than 10% of sample size, Duncan ⁽⁵⁷⁾ states that the binomial distribution is better approximated by the Poisson distribution than the normal distribution in setting up confidence for P . Confidence limits may be determined for this case in a variety of ways such as the Thorndike Chart, Molina's Table II or Dodge and Romig's Tables. The author took Cochran's suggestion, however, and used Table VIII₁ in Fisher and Yates' Statistical Tables (1957) which "gives binomial confidence limits for P for any value of n , and is a useful alternative to the ordinary binomial tables." ⁽⁵⁸⁾ The applicable portion of Table VIII₁ is included in order to illustrate the calculation of the confidence limits for P .

a	p a/N	Probability P of 'a' or more			Probability P of 'a' or fewer		
		.005	.025	.1	.1	.025	.005
8	.5	3.04	3.94	5.09	10.91	12.06	12.96
	.4	2.92	3.82	4.98	11.35	12.79	14.02
	.3	2.82	3.72	4.89	11.77	13.52	15.10
	.2	2.73	3.62	4.80	12.18	14.26	16.22
	.1	2.65	3.53	4.73	12.59	15.01	17.37
	0	2.57	3.45	4.66	12.99	15.76	18.58

To obtain the limits of the probability of an event (observed to occur 'a' times out of N) corresponding to a given probability level P , divide the tabulated values by N , first interpolating if necessary by linear interpolation with reference to $P = a/N$. The limits of the expectation of Poisson distributions are given directly, taking $p = 0$.

The limits for the probability of 8 defective titles occurring in a sample of size 390 may be gotten from the table directly. Taking $p=0$, the limits are 3.45 and 15.76 as circled above. Thus the 95% limits for P are estimated as:

$$\hat{P}_L = \frac{3.45}{390} \times 100\% = .885 \text{ per cent}$$

$$\hat{P}_u = \frac{15.76}{390} \times 100\% = 4.04 \text{ per cent}$$

From which it may be said that there is a very good chance, a 95% probability in fact, that the proportion of defective or not KWIC-able titles in the Engineering Index falls within the interval from .885 to 4.04 per cent.

Next confidence limits were determined for the Type 1 titles — those titles having at least one keyword exactly or in some root form as a subject heading in the Index. Now, however, the fraction defective, in this instance the fraction of Type 1 titles, is larger than 10% such that the Poisson distribution and the very convenient Fisher and Yates Table VIII₁ no longer apply. Duncan points out that in this case the binomial distribution can be approximated reasonably well by a normal distribution with a mean equal to p and a standard deviation equal to $\sqrt{\frac{pq}{n}}$. (59)

Thus the 95% limits for Type 1 titles are

$$83.1 \pm 1.96 \sqrt{\frac{(83.1)(16.9)}{390}} = 83.1 \pm 3.72 \quad (6)$$

$$\hat{P}_L = 79.3$$

$$\hat{P}_u = 86.7$$

This simply means that there is a 95% chance that the proportion of titles in the Engineering Index having at least one keyword exactly or in some root form as a subject heading in the Index will fall within the interval from 79.3 to 86.7 per cent.

Comments concerning the implications of this sample and the attendant confidence limits are being withheld until all sample data is presented. This permits a graphical compilation and a comparison of results.

D. Applied Science and Technology Index Sample

1. General

Applied Science and Technology Index is published monthly by H. W. Wilson Company, 950 University Avenue, New York 52, New York. This publication classified engineering and science articles according to a subject heading system, including hundreds of headings such as Ablation, Aeration, Todides, Magnetism, Polarograph, Transistors and Zwitterions.

The particular issue chosen was the 1962 annual cumulated index. The Applied Science and Technology Index indexes engineering and science articles from some 215 publications. This issue contained 1313 pages with 55-60 titles per page or some 75,000-odd bibliographic citations; however, since an article may be indexed under more than one subject heading, it is not known how many distinct titles the volume includes. (60)

No estimate of sample size was computed for sampling this index, the experience gained in sampling from the Engineering Index led the author to adopt a wait and see attitude; the assumption being that since both indexes are concerned with engineering and science that results should be similar. Sampling was terminated after 200 titles had been examined.

2. Classification

Type	"1"	"2"	"3"	"4"	"5"	Total
Number	176	135	2	15	7	200
% of Total	88	67.5	1.0	7.5	3.5	100.0

3. Confidence Limits

Using Fisher and Yates again the 95% confidence limits for P are estimated as:

$$\hat{P}_L = \frac{281}{200} = 1.4, \quad \hat{P}_U = \frac{1442}{200} = 7.2$$

Also the 95% limits for Type 1 titles are:

$$\frac{176}{200} \pm 1.96 \sqrt{\frac{(88)(12)}{200}} = 88 \pm 4.5 \quad (7)$$

$$\hat{P}_L = 83.5, \quad \hat{P}_U = 92.5$$

E. Metal Literature Sample

1. General

ASM Review of Metal Literature is published monthly by the American Society for Metals, Metals Park, Ohio. It is "an annotated survey of articles, technical papers and reports appearing in engineering, scientific and industrial journals and books, here and abroad." ⁽⁶¹⁾ This index professes to index based on the content of the article and not merely

on the title. To accomplish this a deeper and more detailed analysis of subject content is performed and more flexible cross-referencing is employed.

The particular issue chosen was Vol. 19, January-December 1962. This particular issue contained approximately 12,800 indicative abstracts from 900-odd publications. The author chose to sample from this index because it embraced a far more specific subarea, metal literature, of science and engineering. It was assumed that this index would yield even better results, that is, fewer defective titles.

2. Classification

Type	"1"	"2"	"3"	"4"	"5"	Total
Number	103	67	0	17	0	120
% of Total	86	56	0	14	0	100.0

3. Confidence Limits

Using Fisher and Yates again, the 95% confidence limits for P are estimated as:

$$\hat{P}_L = 0, \quad \hat{P}_u = \frac{369}{120} = 3.08$$

And the 95% limits for Type 1 titles are:

$$\frac{103}{120} \pm 1.96 \sqrt{\frac{(86)(14)}{120}} = 86 \pm 6.22 \quad (8)$$

$$\hat{P}_L = 80.0, \quad \hat{P}_u = 92.2$$

F. Public Affairs Information Service Sample

1. General

The Public Affairs Information Service, Annual Cumulated Bulletin is published by Public Affairs Information Service, Inc., 11 West 40th Street, New York, New York, 10018. "Its purpose is the publication of a weekly Bulletin, listing by subject current books, pamphlets, periodical articles, government documents, and any other useful library material in the field of economics and public affairs. The Bulletin is not primarily an index to periodicals, though selected articles in more than one thousand periodicals are listed each year. Publications of all kinds from all English-speaking countries are included, as well as many printed in English in other countries. Emphasis is placed upon factual and statistical information. Works in foreign languages are not mentioned.

(62)

This index or bulletin was chosen by the author as a first step in evaluating titles outside of the fields of engineering and science. The particular issue selected was Vol. 48, October 1962-September 1963 which included 845 pages of brief annotations with approximately 33,000 distinct titles. No attempt was made to predetermine a sample size; it was arbitrarily set by the author at 300. Further, because of the extra time involved in the additional look-up required of Type 2 titles, the Type 2 classification was dropped. Obviously titles which would have appeared under this classification still appear as Type 1 titles.

2. Classification

Type	"1"	"2"	"3"	"4"	"5"	Total
Number	166	-	10	113	11	300
% of Total	55.3	-	3.3	37.7	3.7	100.0

3. Confidence Limits

Using Fisher and Yates, the 95% confidence limits for P are:

$$\hat{P}_L = \frac{550}{300} = 1.83, \quad \hat{P}_u = \frac{1965}{300} = 6.55$$

The 95% limits for Type 1 titles are:

$$\frac{166}{300} \pm 1.96 \sqrt{\frac{(55.3)(44.7)}{300}} = 55.3 \pm 5.63 \quad (9)$$

$$\hat{P}_L = 49.7, \quad \hat{P}_u = 60.9$$

G. Business Periodicals Index

1. General

The Business Periodicals Index is published monthly by H. W. Wilson Company, 950 University Avenue, New York 52, New York. This publication indexes 118 periodicals in the fields of accounting, advertising, bank and finance, insurance, general business, labor, marketing public administration and taxation. Its citations are bibliographic and the particular issue chosen, which included material from July 1962 through

(63)
 June 1963, contained approximately 49,000 entries. It is not known how many distinct titles the volume includes because articles may be indexed under more than one subject heading.

As with the previous index, the author arbitrarily set sample size, in this case at 510. Also the Type 2 classification was dropped.

2. Classification

Type	"1"	"2"	"3"	"4"	"5"	Total
Number	229	-	38	164	79	510
% of Total	45.0	-	7.5	32.0	15.5	100.0

3. Confidence Limits

Now however the fraction defective is larger than 10%, therefore the normal approximation to the binomial distribution was employed. Thus the 95% confidence limits for P are:

$$15.5 \pm 1.96 \sqrt{\frac{(84.5)(15.5)}{510}} = 15.5 \pm 3.15 \quad (10)$$

$$\hat{P}_L = 12.4, \quad \hat{P}_U = 18.7$$

The 95% limits for Type 1 titles are:

$$\frac{229}{510} \pm 1.96 \sqrt{\frac{(45)(55)}{510}} = 45.0 \pm 4.33 \quad (11)$$

$$\hat{P}_L = 40.7, \quad \hat{P}_U = 49.3$$

H. International Index Sample

1. General

The International Index is published quarterly by H. W. Wilson Company, 950 University Avenue, New York 52, New York. It is a guide to periodical literature in the social sciences and humanities and as such was the most generic body of titles which the author examined.

The particular issue chosen was the annual cumulated index consisting of bibliographic citations of articles appearing in 173 international periodicals from April 1962 through March 1963. As with previous Wilson volumes, it is not known how many distinct titles are included because titles may be indexed under more than one subject heading. (64)

Again due to time constraints, the author set sample size at 510 and dropped the Type 2 classification.

2. Classification

Type	"1"	"2"	"3"	"4"	"5"	Total
Number	164	-	8	225	113	510
% of Total	32.2	-	1.5	44.2	22.1	100.0

3. Confidence Limits

Using the normal approximation again, the 95% limits for P are:

$$22.1 \pm 1.96 \sqrt{\frac{(22.1)(77.9)}{510}} = 22.1 \pm 3.66 \quad (12)$$

$$\hat{P}_L = 18.4, \quad \hat{P}_U = 25.8$$

The 94% limits for Type 1 titles are:

$$\frac{164}{510} \pm 1.96 \sqrt{\frac{(32.2)(67.8)}{510}} = 32.2 \pm 4.06 \quad (11)$$

$$\hat{P}_L = 27.9, \quad \hat{P}_U = 36.1$$

I. Interpretation of Data

The results of the preceding six samples are compiled in Table I, page 48.

At a glance it is obvious that titles within the fields of engineering and science are very effective or KWIC-able, beyond all expectation as far as the author is concerned. There is, in fact, a 95% probability that a vast percentage, between 79.3 and 92.2 per cent (the lower confidence limit of the Engineering Index and the upper confidence limit of the Applied Science and Technology Index), of the titles within these fields are Type 1 titles.

The Applied Science and Technology Index has the largest percentage (83.5 to 92.5 per cent confidence interval) of Type 1 titles. The author feels that this is due to the fact that titles may be indexed under more than one subject heading and that this index has the largest number of subject headings per se. The author offers no factual evidence to verify the latter comment; however, the fact that this index has the largest number of Type 2 and smallest number of Type 4 titles, seems to substantiate the observation.

The preponderance of Type 1 and 2 titles might suggest that indexers in the fields of engineering and science, either intentionally or unwittingly,

TYPE	1			2	3	4	5		Total
	Number	% of Total & Conf. Limits	324				8	50	
ENGINEERING INDEX	Number		--	--	124	8	--	6	390
	% of Total & Conf. Limits		79.3	86.7	31.8	2.05	.89	2.05	100.0
AP. SC. AND TECH. INDEX	Number		--	--	135	2	--	7	200
	% of Total & Conf. Limits		83.5	92.5	67.5	1.0	1.4	3.5	100.0
ASM REVIEW	Number		--	--	67	0	--	0	100
	% of Total & Conf. Limits		80.0	92.2	56	0	--	0	100.0
BULLETIN PAIS	Number		--	--	--	10	--	11	300
	% of Total & Conf. Limits		49.7	60.9	--	3.3	1.83	3.7	100.0
BUS. PERIOD. INDEX	Number		--	--	--	38	--	79	510
	% of Total & Conf. Limits		40.7	49.3	--	7.5	12.4	15.5	100.0
INTERNATL INDEX	Number		--	--	--	8	--	113	510
	% of Total & Conf. Limits		27.9	36.1	--	1.5	14.2	22.1	100.0

Table I: Compilation of Data

use titles as a basis for originating subject headings. This fact, if so, adds substantial support for KWIC indexing, because KWIC indexing, in contrast to human indexing, can be done by machines with little physical and infinitesimal intellectual effort being required.

That the ASM Review sample yielded no defective titles came as no surprise to the author. It was apparent while conducting the sample that this index bases its analysis on the content of the article and not merely on the title - something which was not apparent in the other samples. The author concluded that the ASM Review indexes metal literature thoroughly and deeply. Further, it and the Applied Science and Technology Index actually had more Type 2 titles than Type 1.

Comparison or contrast of the width of the confidence intervals for Type 1 titles in the engineering and science fields yields little because the sample sizes are unequal. In this instance the width of the intervals is not related to proportion of Type 1 titles but to sample size.

In the engineering and science area the percentage of defective titles was least in the ASM Review sample, increased in the Engineering Index sample and increased again in the Applied Science and Technology sample. The author feels that this order is plausible, if predictable. Assuming identical analysis, in this case KWIC indexing, the percentage of defective or not KWIC-able titles in a file of documents, even an amorphous file, is inversely proportional to the degree of technicality or, better stated, the technical character of the documents themselves.

The small number of not KWIC-able titles in the Public Affairs and Information Service sample came as a surprise to the author, for it imputes

a technical character to titles of documents in fields (economics and public affairs) which are unquestionably vastly less technical "in the usual sense" than science and engineering. "In the usual sense" deserves discussion. Technical in the usual sense has come to delimit the use of the word technical to science and in particular engineering. This is erroneous; technical may apply to any highly specialized field. In this sense medicine, law, the priesthood as well as the Navy and many trades are technical. Frequently such fields have languages or lingos which the non-cognoscenti find confusing or unintelligible. Jargon-filled titles from such fields, however, are exceedingly specific and descriptive for the profession or trade-oriented searcher. It is in this context that titles from the Public Affairs and Information Service Bulletin rank so high in findability. Its source material is indeed technical in character, much of it being the outputs of congressional committees, governmental departments and other bureaucratic agencies which characteristically title their exudations with balefully scientific and/or onerously long, but nonetheless descriptive, titles.

The International Index sample indicates the woeful lack of descriptive titles in the social sciences and humanities. Here authors employ metaphor, simile, metonymy, synecdoche and other unpronounceable artifices to be cleverly non-descriptive. For example, the title, "Saints and Sinners in Foreign Investment", which appeared under the subject heading, "Foreign Investment" contains metaphor, but remains descriptive and KWIC-able. Thus some artful languages in titles is tolerable. On the other hand titles such as "Milestone", "The Greeks Had a Name for It", and "Squeezplay" are

hopelessly non-KWIC-able in their present form. The author is not suggesting that all titles be descriptive. On the contrary, the figurative use of words in non-technical and transient literature is refreshing and desirable. Still, figurative ploys in titling are the bane of machine indexing as this sample and the Business Periodicals Index sample, which was little better, readily illustrate.

In summary this study reveals that a KWIC index, based on keywords which appear in the title alone will index documents in the fields of engineering and science in the same way as would a human between 79.3 and 92.5 per cent of the time, and also that the largest proportion of not KWIC-able titles sampled was a mere 3.5 per cent. Further, KWIC indexing is suitable for any highly specialized field (the more that it is jargon-filled the better), not just engineering and science. Clearly the adequacy of titles as search tools in any highly specialized field is certain. Finally the less specialized the field (as in social science and the humanities), the less descriptive titles are and the less effective KWIC indexing is.

IV. PROS AND CONS OF KWIC INDEXING

A. General

Up to this point this protracted discourse has embraced the background of KWIC indexing, a description of it and a complete evaluation of the hypothesis -- that titles are adequate search tools -- upon which it is based. However, without a discussion of the pros and cons of KWIC indexing, previous effort is singularly incomplete. This chapter is aimed at such a discussion and advantages shall be enumerated first.

B. Advantages

1. Heretofore this thesis has discussed KWIC indexing as applied to titles of documents. It may, however, be carried out on various levels, depending on the purpose an index is to serve. The process may be applied to either the title, abstract or entire text of a document.

2. Words of titles are ordered in their original sequence so that the reader has little difficulty in grasping the content of the document which the words describe. In printed form the indexes can be searched rapidly by eye, so that the need for machines in the retrieval phase is largely obviated.

3. The machine will index a document many other ways, the exact number of entries depending on the number of keywords in the title. This redundancy improves the probability of finding desired references, since the number of access points for conducting a search are increased. Certainly a human would index under multiple subject headings. It is doubtful though that he would use as many entries as a computer.

4. One of the major advantages of KWIC indexing is its flexibility.

If the title does not adequately describe the content of a document, descriptive words may be editorially added to improve title significance. Further, such words can be made to act as keywords under which the title will be indexed or they may function as stoplist words, which appear each time a title is indexed but which themselves do not become index entries. When the situation arises, editorial revisions may be made either to prevent unwanted terms from indexing or to index more than one part of some words. For example, the word mucopolysaccarides could be indexed under muco, poly and saccarides in order to give researchers more than one approach to the subject. Conversely the hyphen may be used to prevent the indexing of certain words contained in multiple-word expressions. For example very-high-speed instead of very high speed. The references "see" and "see also" are readily exploitable and can be inserted in proper alphabetic positioning automatically by the computer while it is preparing the KWIC index. (65), (66)

5. The acronym "KWIC" is well chosen because it conveys the concept of speed. It is not extravagant to say that this system produces indexes more rapidly and accurately than any other form of indexing. Because of the speed of electronic data processing equipment, elapsed time between the date of publication of a document and its appearance in an index can be reduced from many months to a few days. (67)

6. A KWIC index system has the capacity of providing the users a double convenience in having an index in book form at their desks.

7. 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. These and a whole host of indexing decisions inherent in any human indexing effort are eliminated by the KWIC index. Once the index is developed and in production, indexing can be done rapidly on a routine, essentially clerical, basis with the exception of adding descriptive words.

8. Printouts may have unlimited formats and may be reproduced for distribution by any convenient process, such as photo-offset. (68)

9. Generally, a KWIC index contains three sections as previously stated; the keyword index, a bibliography of documents, and an author index. Other sections may be added if desired, however. For example, corporate entries, lists of periodicals index, etc. (69)

10. KWIC indexing can be adopted to the literature of any scientific discipline, as well as to such areas as correspondence files, files of internally generated memorandums, legal papers, procedure manuals, etc.

11. Costs with respect to the preparation of a KWIC index are very encouraging. Kraft says that the approximate cost of producing a KWIC index containing 2000 titles on rented equipment is \$580 per month. Naturally on in-house equipment the cost would drop drastically. Kraft makes the comment:

"The cost per article indexed is only 29¢. This compares most strikingly with the cost of manually indexing an article, which averages \$4. ...Since each title will average six keywords, the cost per index entry is 5¢ (29¢/6)." (70)

C. K. Bauer, manager of the Scientific and Technical Information Department of the Lockheed Aircraft Corporation, Marietta, Georgia, claims that Lockheed's newly installed KWIC system which processes 150 new books, 1250

reports and 425 periodicals monthly and prepares monthly overdue notices, accession lists and financial reports saves the company \$600 per month. Further, total run time in data processing for all outputs of the system is less than two and one-half hours per month, including taping, sorting and printing.⁽⁷¹⁾ Materials from Mr. Bauer's system are exhibited in Appendix A.

C. Disadvantages

1. The most common objection to KWIC indexing is that titles are not descriptive enough. In view of the results of this and other studies, the author does not feel that this is a legitimate objection. KWIC is feasible in highly specialized fields; its wholesale use is not suggested. On the other hand, when a title is found to be non-descriptive, or not sufficiently descriptive, it may be repaired by the scant editorial effort of adding descriptors.

2. KWIC indexing does not index deeply enough. This is less a disadvantage than a limitation. Certainly KWIC's hands are tied by the title when it comes to depth of indexing. By design a KWIC index is a shallow index, and consequently it can not hope to compete with the likes of the ASM Review of Metal Literature. Each, the shallow and the deep index, serves a useful purpose.

3. A title will sometimes have adequate keywords but still cause confusion because it is not descriptive. For example, an article titled, "Fixed Points and Antipodal Points", which concerns topology, appeared in the American Mathematical Monthly in the 1963 issue and could just as easily have appeared in the Engineering Index or the Applied Science and Technology

Index. If this title were to appear in a KWIC index comparable to either the Engineering Index or the Applied Science and Technology Index it would cause confusion. The title, "Fixed Points and Antipodal Points", could easily be assumed by radio engineers - particularly one who was working in the currently active research area of the antipodal reception of VLF waves - to belong to an article about electromagnetic wave propagation. (72)

Another title that could be confusing is "Ray Paths of Whistling Atmospherics - Differential Geometry:", that of an article on the propagation of radio waves. An editor, lacking sufficient subject-field expertise, would undoubtedly KWIC both titles and thereby generate "noise" in the system. But the total elimination of noise can not be achieved except at great cost. The author feels that titles such as these appear infrequently. Hence, the noise they generate can easily be lived with, at least until interested cognoscenti inform the editor.

4. Another objection to KWIC indexing is that bibliographic information must be in machine readable form, that is, must be punched on IBM cards. This of course is a drawback; however, the time and manpower required for, and the cost of this operation is low. Key punching is faster than typing. Also, since bibliographic entries are in machine readable form, they may be used for other purposes, such as the preparation of library catalog cards, accession lists, etc.

5. Two objections which involve the format of the KWIC index are that it is not very readable and that too much of the title is chopped or lost. These objections involve programming and are easily remedied. One technique which is employed by the American Bar Foundation to make its Index to Current State Legislation more readable is shown in Figure 12.

PERSONS.—AMENDS 1260.3, AGRICULTURAL CODE, TO REVISE AND ENL. CAL. C0574-2
AMENDS 1260.3, AGRICULTURAL CODE, TO INCREASE APPLI. CAL. C0579-4
MENDS 1299.10, 1299.10, CODE, TO REDEFINE FARM CAL. C0822-1
ION PROGRAM.—ADDS 257, CODE, RE BUYING TUBERCU CAL. C1334-4
8% REPEALS 2910, 3407, CODE, TO REVISE PROVIDE CAL. C1806-4
MEDICAL FERTILIZERS AND MINERALS AND REVISES CAL. C0277-7
TAL.—REPEALS 2910, 3407, TECHNICAL AND VETERINARY CAL. C0278-7
WITH THE U.S. DEPT. OF AGRICULTURE AND CERTAIN TOWNS FOR CU R.1. C0561-1
FUNCTIONS FROM DEPT. OF AND CONSERVATION, STATE R.1. C0704-2
WITH THE U.S. DEPT. OF TO CONSTRUCT RESERVOIRS MASS C0569-5
THORIZE COMMISSIONER OF TO REGULATE STANDARDS FOR S.C. A0255-1
MEMBERSHIP ON THE PUBLIC AND ADVISORY BOARD FROM OPERATION OF ILL. C0049-1
OWNS FURNISHING OLD AGE AND TO USE OF PROVISIONS IN U.S. NEWM H0005-A1
OR EXPENSES FOR MEDICAL FUR INDIGENT RELATIVES.—AMENDS PA. A0361-2
D PROVIDE FOR FINANCIAL FUR INDUSTRIAL, AGRICULTURAL, TE MASS C0562-1
OM, PROVIDING FOR STATE FUR TEMPORARY UNEMPLOYMENT COMPE NICH P0276-AR
E OF PERSONS RECEIVING IN CERTAIN CASES AND PROVIDES FOR PA. A0373-1
F NATURAL RESOURCES AND IN ALL REGULAR FEDERAL CASES, DEL. C1010-6
L CH. 460, REPEALS 2910, 3407, IN ALL REGULAR FEDERAL CASES, DEL. C1010-6
OF THE STATE PRISONERS TO THE BUREAU OF UNEMPLOYMENT AND PAID OFF BY CI NEWM H0005-A1
FUNDS TO THE PRISONERS SOCIETY AS A REHABILITATION CENT DEL. H0112-7
ANGE WORKING OF LAWS RE SOCIETY OF THE STATE FOR OPERATE DEL. H0113-7
ACTS 1902, RE FINANCIAL TO FAMILIES WITH DEPENDENT CHILD MASS C0432-2
VIDE A PLAN FOR MEDICAL TO NEEDY DEPENDENTS OF VETERANS MASS C0537-1
PROVIDES FOR FINANCIAL TO QUALIFIED PERSONS 65 OR OLDER NEB. L0100-6
ETC.—PROVIDES FURTHER TO TO SCHOOLS OFFERING SUCH CLASSES OHIO A0194-4
TO VARIOUS VETERANS ORGANIZATIONS, DEL. H0000-1
E UNDER LAW FOR OLD AGE —AMENDS 1, CH. 110A, RE PUBLIC WELFA MASS C0550-1
LIGATION FOR DISABILITY —AMENDS 2, 110D, RE PUBLIC WELFA ALA. A0574-2
E OF RECIPIENT FUR SUCH ETC.—PROVIDES FURTHER AID TO I ALA. A0574-2
RIOUS FUNDS FOR OLD AGE MENTAL HEALTH, SCHOLARSHIP AND ALA. A0589-4
F AND RECEIVING FEDERAL TO EXTEND COVERAGE TO LOWER EMP DEL. C0561-2
TION OF REGENERATION AND AIR CONDITIONING SYSTEMS AND DEL. C0561-2
RATE RECIPIENT OF SUCH DEFENSE TECHNICIANS EMPLOYED ON MASS C0408-1
DS OF HEALTH TO CONTROL POLLUTION, TO PRESCRIBE FINES.—MASS C0408-1
PIRE AUTHORITY TO LEASE TO NIGHTS UN BOSTON EXTENSION AND PA. A0361-2
E OF AERONAUTICS AS TO AIRCRAFT AWARDED CERTIFICATES IN CITY ILL. H0013-1
PERATOR, APPRENTICE AND FOR USE IN BUSINESS OF PEST CAL. C1027-1
CEMENT OF REGULATION OF MOTOR, RAIL AND PIPELINE NEB. L0087-7
NG ALCOHOLIC LIQUORS ON AIRPLANE UNDER ON-PREMISES LICENSE, CAL. C0115-1
ER OF CITY EMPLOYEES TO AIRPLANE UNDER ON-PREMISES LICENSE, CAL. C0115-1
N POWERS AS TO CLASSES OF AUTHORITIES FOR PROMOTION OF NEB. L0044-1
AS TO LANDING FEES AND FACILITIES ON PRIVATE LAND NEWM C0314-4
ON CONVERTING NORTHWEST ALA. JUNIOR COLLEGE TO A TRADE SCHOL ALA. A0540-2
STRUCTION OF LIBRARY AT ALABAMA AGRICULTURAL AND MECHANICAL ALA. A0586-1
ENALTY FOR GIVING FALSE ALARMS IF REPORTS IN RURAL FIRE PROT NEB. L0599-1
AND ORGANIZATION OF ALBUQUERQUE METROPOLITAN ARROYO FLOW NEWM C0318-7
DRIZING USE AND SALE OF ALCOHOLIC BEVERAGE, TO CLAS ALA. A0589-4
CENSE FUND OF THE STATE ALCOHOLIC BEVERAGE, TO CLAS NEWM C0332-2
RE REGULAR LICENSE BY LIQUOR BY STATE LIQUOR COM NEWM H0065-2
BUSINESSES SELLING MO LIQUOR EXCEPT REEK.—AMEND ILL. H0685-2
ER OF ON-SALE OR 43, RE LIQUOR GENERAL LICENSE.—A CAL. C1089-5
D 133, 194, CH. 43, RE LIQUOR LICENSES, TO REVISE ILL. H0681-7
OF ITS ON-SALE GENERAL LIQUOR LICENSE TO ANOTHER, CAL. C1086-3
FOR TRANSFER OF CERTAIN LIQUOR LICENSES ISSUED ON CAL. C1086-3
S BY HOLDERS OF CLASS A LIQUOR LICENSES, TO ADD A CAL. C1086-3
CLASS B CLASS B, RE LIQUOR LICENSES, TO PROVIDE NEB. L0109-2
INTEREST IN MORE THAN LIQUOR RETAIL LICENSES APT WIS. L0105-2
NO PROFESSIONS CODE, RE LIQUOR, TO REGULATE CREDIT CAL. C1891-3
NILE UNDER INFLUENCE OF LIQUOR, NARCOTICS, BARBITU MASS C0369-1
S 4-141, RE SHIPMENT OF LIQUOR, TO REVISE REGULATI S.C. A0240-2
S GOVERNING THE SALE OF LIQUORS WITHIN SPECIFIED ILL. H0685-2
SE PROVISION RE SERVING LIQUORS ON AIRPLANES UNDER CAL. C1086-3
PROHIBITION OF SALE OF LIQUORS WITHIN CERTAIN STATE CAL. C1086-3
OF ALCOHOL, RE LIQUORS ON PRESCRIPTION.—MASS C0449-1
194A SEC. 9, RE SALES OF LIQUORS TO MINORS AND INCU NER. L0109-2
DOS SEC. 9, VITAL 29, RE LIQUORS, TO PRESCRIBE NEB. A0589-2
AND PROFESSIONS CODE RE LIQUORS, TO MAKE EDITORIAL CAL. C1040-6
E 53-162.01, 53-164, RE LIQUORS, TO ELIMINATE USE NEB. L0044-4
AND THOSE RE SINGLE USE ALE OR BEER BOTTLES.—AMENDS AND REP CAL. C1040-6
PROCEDURES CONSTITUTING ALTHOUGH DECAY UNDER PROHIBITION AND I CAL. C1040-6
NCES SERVING BEVERAGES TO ALLIANCE OF BOTTLE DIET CHARITIES AND MASS C0439-1
LIGHTING FORMULA FOR ALLOCATING AND DISTRIBUTING THE EQUA MASS C0310-AR
FOR THEM, TO REQUIRE ALLOCATION OF PART OF FUNDS FOR STD. NO.D C0192-2
N DETERMINING AMOUNT OF ALLIANCE ACCORDING TO NEED AND ESTA ALA. A0574-2
D DETERMINING AMOUNTS OF AND CONTRIBUTION BY STATE MASS C0537-1
SE DEFERRED RETIREMENT AND DEFINITION OF ACTUARIA R.1. C0685-2
SE COMPENSATION, TRAVEL AND EXPENSES OF MEMBERS OF R.1. C0685-2
E ADDITIONAL RETIREMENT FUR MEMBERS OF UNIFORMED B MASS C0441-2
CASE SERVICE OR RETIREMENT FUR MUNICIPAL EMPLOYEES, U.1. C0212-2
10 TO INCREASE PER OIUM FUR TRAVEL ON OFFICIAL BUS MONT H0142-2
SUPPLEMENTAL RETIREMENT FUND FOR STATE EMPLOYEES, PA. A0359-1
S, TO PRESCRIBE HOUSING IN LIEU OF A RESIDENCE.—A WIS. S0016-1
OVISION LIMITING TRAVEL OF SUPERVISORS TO MILEAGE CAL. C1148-1
V AND TO LIMIT MILEAGE OF THOSE SERVING IN PLACE CAL. C1148-1
EASE SERVICE OR RETIREMENT TO INCLUDE PUBLIC SCHOOLS, I R.1. C0109-2
ELECT LESSER RETIREMENT TO PROVIDE A SURVIVOR BENEF MASS C0526-2
LETTURE TIME ACTIVITIES UNDER LAW FOR OLD AGE AID, MASS C0520-1
JUSTMENT OF RETIREMENT —AMENDS 9357.66, GOVERNMENT CAL. C0206-3
PENSIONS AND RETIREMENT ALLOWANCES OF CERTAIN FORMER PUBLIC MASS C0478-2
IN SALARIES AND EXPENSE OF LEGISLATORS AND OF VAN WIS. C0225-AR
OPTIONS FOR PAYMENT OF TO INCLUDE BENEFICIAL MASS C0402-2
TENDERS OR RETIREMENT TO INCLUDE BENEFICIAL MASS C0402-2
CODE, RE CONSTRUCTION, R.1. C0211-4
F PLANS FOR ERECTION OF ALIEN, AMEND EXISTING OR TO FRAME A CAL. C0884-5
IN APPLICATION FOR, OR ALTERATION AND REPAIR OF COUNTY BUILD CAL. C0884-5
OR ANOR OF USE OF FALSE OF INSTITUTIONS UNDER THE CAL. C0193-2
VACANCIES AND CHOSING UP, GAME AND FISH LICENSE NO.D C0193-2
NO HOUR AND PROVIDE NEB. L0109-2
CONDITIONS MAKING THIS TO REPEAL THE OLD AND ENAL NEWM H0032-A1
1. RESOLUTION PROPOSING TO REPEAL THE OLD AND ENAL NEWM H0032-A1
SOLUTION, PROVIDING FOR 9, REPEAL SEC. 18 ART. 7, WIS. C0127-2
TUTION, RE LEGISLATIVE CONSTITUTION OF 1105, TO R.1. C0127-2
AL ELECTION ON PROPOSED TO REPEAL THE OLD AND ENAL NEWM H0032-A1
1. RESOLUTION PROPOSING TO REPEAL THE OLD AND ENAL NEWM H0032-A1
T. RESOLUTION PROPOSING TO REPEAL THE OLD AND ENAL NEWM H0032-A1
S TO INCREASE PER OIUM FUR TRAVEL ON OFFICIAL BUS MONT H0142-2
SUPPLEMENTAL RETIREMENT FUND FOR STATE EMPLOYEES, PA. A0359-1
S, TO PRESCRIBE HOUSING IN LIEU OF A RESIDENCE.—A WIS. S0016-1
OVISION LIMITING TRAVEL OF SUPERVISORS TO MILEAGE CAL. C1148-1
V AND TO LIMIT MILEAGE OF THOSE SERVING IN PLACE CAL. C1148-1
EASE SERVICE OR RETIREMENT TO INCLUDE PUBLIC SCHOOLS, I R.1. C0109-2
ELECT LESSER RETIREMENT TO PROVIDE A SURVIVOR BENEF MASS C0526-2
LETTURE TIME ACTIVITIES UNDER LAW FOR OLD AGE AID, MASS C0520-1
JUSTMENT OF RETIREMENT —AMENDS 9357.66, GOVERNMENT CAL. C0206-3
PENSIONS AND RETIREMENT ALLOWANCES OF CERTAIN FORMER PUBLIC MASS C0478-2
IN SALARIES AND EXPENSE OF LEGISLATORS AND OF VAN WIS. C0225-AR
OPTIONS FOR PAYMENT OF TO INCLUDE BENEFICIAL MASS C0402-2
TENDERS OR RETIREMENT TO INCLUDE BENEFICIAL MASS C0402-2
CODE, RE CONSTRUCTION, R.1. C0211-4
F PLANS FOR ERECTION OF ALIEN, AMEND EXISTING OR TO FRAME A CAL. C0884-5
IN APPLICATION FOR, OR ALTERATION AND REPAIR OF COUNTY BUILD CAL. C0884-5
OR ANOR OF USE OF FALSE OF INSTITUTIONS UNDER THE CAL. C0193-2
VACANCIES AND CHOSING UP, GAME AND FISH LICENSE NO.D C0193-2
NO HOUR AND PROVIDE NEB. L0109-2
CONDITIONS MAKING THIS TO REPEAL THE OLD AND ENAL NEWM H0032-A1
1. RESOLUTION PROPOSING TO REPEAL THE OLD AND ENAL NEWM H0032-A1
SOLUTION, PROVIDING FOR 9, REPEAL SEC. 18 ART. 7, WIS. C0127-2
TUTION, RE LEGISLATIVE CONSTITUTION OF 1105, TO R.1. C0127-2
AL ELECTION ON PROPOSED TO REPEAL THE OLD AND ENAL NEWM H0032-A1
1. RESOLUTION PROPOSING TO REPEAL THE OLD AND ENAL NEWM H0032-A1
T. RESOLUTION PROPOSING TO REPEAL THE OLD AND ENAL NEWM H0032-A1
S TO INCREASE PER OIUM FUR TRAVEL ON OFFICIAL BUS MONT H0142-2
SUPPLEMENTAL RETIREMENT FUND FOR STATE EMPLOYEES, PA. A0359-1
S, TO PRESCRIBE HOUSING IN LIEU OF A RESIDENCE.—A WIS. S0016-1
OVISION LIMITING TRAVEL OF SUPERVISORS TO MILEAGE CAL. C1148-1
V AND TO LIMIT MILEAGE OF THOSE SERVING IN PLACE CAL. C1148-1
EASE SERVICE OR RETIREMENT TO INCLUDE PUBLIC SCHOOLS, I R.1. C0109-2
ELECT LESSER RETIREMENT TO PROVIDE A SURVIVOR BENEF MASS C0526-2
LETTURE TIME ACTIVITIES UNDER LAW FOR OLD AGE AID, MASS C0520-1
JUSTMENT OF RETIREMENT —AMENDS 9357.66, GOVERNMENT CAL. C0206-3
PENSIONS AND RETIREMENT ALLOWANCES OF CERTAIN FORMER PUBLIC MASS C0478-2
IN SALARIES AND EXPENSE OF LEGISLATORS AND OF VAN WIS. C0225-AR
OPTIONS FOR PAYMENT OF TO INCLUDE BENEFICIAL MASS C0402-2
TENDERS OR RETIREMENT TO INCLUDE BENEFICIAL MASS C0402-2
CODE, RE CONSTRUCTION, R.1. C0211-4
F PLANS FOR ERECTION OF ALIEN, AMEND EXISTING OR TO FRAME A CAL. C0884-5
IN APPLICATION FOR, OR ALTERATION AND

Observe that the left portion of each entry is shaded grey to highlight the keyword column and that each keyword appears only once for quicker recognition. Figure 13 illustrates what can be done to remedy the chopped title problem. Here 103 characters instead of the conventional 60 characters is available for the title. "An analysis of the lengths of titles in a 4500 - article computer literature index of periodicals, books and proceedings showed that 30 per cent of the titles would have been chopped in order to produce a 60 character single line, but only 2 per cent of the titles were (73) chopped by using a 103 character line." The 103 character format is the obvious solution to the chopping problem.

Thus, objections regarding format are easily rectified.

KEYWORD INDEX		REP - 510
PROGRAMMING OF DIGITAL COMPUTERS	SIGNAL CORPS RESEARCH AND DEVELOPMENT ON AUTOMATIC	CACM522 22
TIONAL APPROXIMATIONS FOR THE ERROR FUNCTION AND FOR	SIMILAR FUNCTIONS NOTE ON THE CONSTRUCTION OF KA	CACM618 394
SALE, A	SIMPLE ALGEBRAIC LANGUAGE FOR ENGINEERS	CACM520 22
A DECISION MATRIX AS THE BASIS FOR A	SIMPLE AUTOMATIC COILING SYSTEMS	CACM587 5
PARAMETER ESTIMATION FOR	SIMPLE DATA INPUT ROUTINE	CACM620 577
ENCY IN SOLVING LINEAR PROGRAMMING PROBLEMS WITH THE	SIMPLE NONLINEAR MODELS	CACM577 23
REPRESENTATION OF CHEMICAL KINETICS	A SIMPLE TECHNIQUE FOR CODING DIFFERENTIAL EQUATIONS	CACM620 616
SOLUTION OF DIFFERENTIAL EQUATIONS	SIMPLEX ALGORITHM DECISION RULE FOR IMPROVED EFFICI	CACM617 507
ANALYSIS AND PATTERN RECOGNITION	SIMULATION AND ANALYSIS OF BIOCHEMICAL SYSTEMS, I,	CACM610 577
	SIMULATION AND ANALYSIS OF BIOCHEMICAL SYSTEMS, II,	CACM621 63
	SIMULATION AND ANALYSIS OF BIOCHEMICAL SYSTEMS, III,	CACM622 115
	SIMULATION OF A COMPUTER TIMING DEVICE	CACM627 363
	SIMULATION OF CITY TRAFFIC	CACM624 274
	SIMULATION OF DISCRETE FLOW SYSTEMS	CACM620 659
	SIMULATOR FOR THE IBM 650 AND BURROUGHS 220	CACM520 20
	SIMULTANEOUS EQUATIONS	CACM627 377
	SIMULTANEOUS LINEAR EQUATIONS WITH POLYNOMIAL COEFFIC	CACM526 16
	SIMULTANEOUS NONLINEAR EQUATIONS	CACM520 12
	SIMULTANEOUS ORDINARY DIFFERENTIAL EQUATIONS USING A	CACM626 355
	SIZE GENERAL PURPOSE DIGITAL COMPUTER USING MAGNETIC	CACM520 3
	SLANG SYSTEM	CACM611 75
	SMALGOL-61	CACM614 479
MAGNETIC (FERRITE) ELEMENTS	SMALL SIZE GENERAL PURPOSE DIGITAL COMPUTER USING	CACM520 3
REMARKS ON THE PRACTICAL	SOLUTION OF CHARACTERISTIC VALUE PROBLEMS	CACM526 34
SIMULATION AND ANALYSIS OF BIOCHEMICAL SYSTEMS, II,	SOLUTION OF DIFFERENTIAL EQUATIONS	CACM621 63
KNOWN EIGENVECTORS	SOLUTION OF EIGENVALUE PROBLEMS WITH APPROXIMATELY	CACM627 351
	SOLUTION OF POLYNOMIAL EQUATIONS	CACM620 644
	SOLUTION OF SIMULTANEOUS EQUATIONS	CACM627 377
	SOLUTION OF SIMULTANEOUS LINEAR EQUATIONS WITH POLYN	CACM526 16
	SOLUTION OF SIMULTANEOUS ORDINARY DIFFERENTIAL EQUATI	CACM626 355
	SOLUTION OF TRIAGONAL MATRICES	CACM617 314
	SOLVING A TRANSCENDENTAL EQUATION	CACM627 377
TRIANGULAR WALK PATTERN FOR THE DOWN-HILL METHOD OF	SOLVING LINEAR PROGRAMMING PROBLEMS WITH THE SIMPLEX	CACM620 502
ALGORITHM A DECISION RULE FOR IMPROVED EFFICIENCY IN	SOME* HAS BEEN PREVENTED AUTOMATICALLY	
INDEXING OF THE WORD	SORTING METHODS	CACM624 613
STIMATION OF THE RELATIVE EFFICIENCY OF TWO INTERNAL	SORTING OF LARGE NETWORKS	CACM624 558
TOPOLOGICAL	SORTING PROCEDURE	CACM621 20
A HIGH-SPEED	SORTING PROCEDURE	CACM527 30
A HIGH-SPEED	SOURCES OF INFORMATION ON CAREER OPPORTUNITIES IN MAT	CACM627 472
HEMATICS, PROGRAMMING AND ELECTRONIC DATA PROCESSI	SOVIET COMPUTER TECHNOLOGY, 1959	CACM603 131
	SOVIET CYBERNETICS AND COMPUTER SCIENCES, 1960	CACM610 566
	SOVIET UNION	CACM526 8
A VISIT TO COMPUTATION CENTERS IN THE	SPECIAL FUNCTIONS FROM DIFFERENTIAL EQUATIONS	CACM523 3
ELIMINATION OF	SPECIAL FUNCTIONS FROM DIFFERENTIAL EQUATIONS*	CACM526 21
REMARKS ON *ELIMINATION OF	SPECIAL IR LANGUAGE	CACM621 8
THE PROS AND CONS OF A	SPECIFICATION LANGUAGES FOR MECHANICAL LANGUAGES AND	CACM610 512
THEIR PROCESSORS, A BAKER'S DOZEN	SPEEDS OF FLEXIBLE SHAFTS ON AN AUTOMATIC COMPUTER	CACM526 27
A TECHNIQUE FOR COMPUTING CRITICAL ROTATIONAL	SPHERE LEAST SQUARES	CACM624 611
FITTING OF A GREAT CIRCLE THROUGH POINTS ON A	SPHERE *EFFICIENT METHOD FOR GENERATING UNIFORMLY D	CACM524 17
DISTRIBUTED POINTS ON THE SURFACE OF AN N-DIMENSIONAL	SPHERE* *EFFICIENT METHOD FOR GENERATING UNIFORMLY D	CACM520 26
DISTRIBUTED POINTS ON THE SURFACE OF AN N-DIMENSIONAL	SPHERES A NOTE ON A METHOD	CACM524 13
FOR GENERATING POINTS UNIFORMLY ON N-DIMENSIONAL	SPHERES BY THE METHOD OF LEAST SQUARES	CACM614 431
	SPREADING	CACM614 437
	SPLITTING	CACM622 102
	SPLITTING IN AN ITERATIVE PROGRAM	CACM614 172
	SQUARE ROOT IN THE QUARTER-IMAGINARY NUMBER SYSTEM	CACM605 319
	SQUARE ROOTS THROUGH CONTINUED SQUARING	CACM524 23
	SQUARE ROOTS WITHOUT USING DIVISION	CACM602 86
	SQUARE ROOTS WITHOUT USING DIVISION*	CACM524 13
	SQUARE-ROOT APPROXIMATIONS	CACM521 5
	SQUARE-ROOT COMPUTATION	CACM618 353
	SQUARES	CACM614 431
	FITTING SPHERES BY THE METHOD OF LEAST SQUARES	CACM624 611
SPHERE	LEAST SQUARES FITTING OF A GREAT CIRCLE THROUGH POINTS ON A	CACM527 29
	MULTI-DIMENSIONAL LEAST-SQUARES POLYNOMIAL CURVE FITTING	CACM626 351
	A SHORT METHOD FOR MEASURING ERROR IN A LEAST-SQUARES POWER SERIES	CACM605 319
	COMPUTATION OF SQUARE ROOTS THROUGH CONTINUED SQUARING	CACM620 537
	PROGRAMS FOR THE IBM 650, DATATRON 205, AND UNIVAC	CACM616 279
	SS-80 A LIST OF COMPUTER SYSTEMS	CACM627 505
	COMBAT VEHICLE FIRING	CACM523 17
	IMPLEMENTING A	CACM624 547
	PROPOSED	CACM622 70
	REITERATION OF ACM POLICY TOWARD	CACM626 321
	A PROPOSAL FOR A SET OF PUBLICATION	CACM616 236
	A START AT AUTOMATIC STORAGE ASSIGNMENT	CACM626 361
	THE	CACM616 362
CESS	STATE TO TOTAL TIME RATIO IN A DOUBLE EXPONENTIAL PR	CACM627 418
THE DEPARTMENT OF COMPUTER MATHEMATICS AT MOSCOW	STATE UNIVERSITY	CACM627 418
COMBINING ALGOL	STATEMENTS	CACM627 418
AN ALGORITHM DEFINING ALGOL ASSIGNMENT	STATEMENTS IN ALGOL 60	CACM627 418
TECHNIQUES FOR BOOLEAN EXPRESSIONS AND CONDITIONAL	STATEMENTS TO PRODUCE A TRUTH FUNCTION TABLE	CACM627 418
ALGORITHM FOR ANALYZING LOGICAL	STATIC STORAGE ALLOCATION	CACM627 418
STOCHASTIC EVALUATION OF A	STATIONARY NOISE	CACM627 418
DIGITAL SYNTHESIS OF CORRELATED	STATISTICAL PROGRAMS AT THE UNIVERSITY OF NORTH	CACM627 418
CAROLINA	STATISTICAL PROGRAMS FOR THE IBM 650, PART I	CACM524 13
	STATISTICAL PROGRAMS FOR THE IBM 650, PART II	CACM527 32
1962)	STATUS OF IPL-V FOR THE PHILCO 2000 COMPUTER (JUNE	CACM627 477
	STOCHASTIC EVALUATION OF A STATIC STORAGE ALLOCATION	CACM610 460
	STORAGE ALLOCATION	CACM611 417
	A GENERAL FORMULATION OF	CACM610 417
PROGRAM ORGANIZATION AND RECORD KEEPING FOR DYNAMIC	STORAGE ALLOCATION	CACM610 422
EXPERIENCE IN AUTOMATIC	STORAGE ALLOCATION	CACM610 436
STOCHASTIC EVALUATION OF A STATIC	STORAGE ALLOCATION	CACM610 460
TECHNIQUES FOR STORAGE ALLOCATION ALGORITHMS	STORAGE ALLOCATION	CACM610 469

Figure 13: The 103 Character, Single Line KWIC Index Format

SYSTEM	A PREPLANNED APPROACH TO A	STORAGE ALLOCATION COMPILER	CACM611 417
D SYSTEM	DYNAMIC	STORAGE ALLOCATION FOR AN INFORMATION RETRIEVAL	CACM610 431
AN AUTOMATIC USE OF A BACKING STORE	PROBLEMS OF	STORAGE ALLOCATION IN A MULTIPROCESSOR MULTIPROGRAMME	CACM610 431
	DYNAMIC	STORAGE ALLOCATION IN THE ATLAS COMPUTER, INCLUDING	CACM610 435
	A	STORAGE ALLOCATION SCHEME FOR ALGOL 60	CACM610 441
	A SEMI-AUTOMATIC	STORAGE ALLOCATION SYSTEM AT LOADING TIME	CACM610 446
NUMBER OF 16 RANDOM/	A METHOD OF REPRESENTATION,	STORAGE AND RETRIEVAL OF 13 RANDOM CODES IN A 4-DIGIT	CACM623 165
	A START AT AUTOMATIC	STORAGE ASSIGNMENT	CACM625 321
	ALLOCATION OF	STORAGE FOR ARRAYS IN ALGOL 60	CACM611 500
	USE OF MAGNETIC TAPE FOR DATA	STORAGE IN THE ORACLE-ALGOL TRANSLATOR	CACM611 15
	SUCCESSIVE APPROXIMATIONS AND COMPUTER	STORAGE PROBLEMS IN ORDINARY DIFFERENTIAL EQUATIONS	CACM615 222
	COMPUTER, INCLUDING AN AUTOMATIC USE OF A HACKING	STORE DYNAMIC STORAGE ALLOCATION IN THE ATLAS	CACM611 435
	MULTIPROGRAMMING	STRETCH, FEASIBILITY CONSIDERATIONS	CACM624 13
ALGOL 60	A	A STRING LANGUAGE FOR SYMBOL MANIPULATION BASED ON	CACM621 54
	EVALUATING NUMBERS EXPRESSED AS	STRINGS OF ENGLISH WORDS	CACM600 541
	INDEXING OF THE WORD	*STRUCTURE* HAS BEEN PREVENTED AUTOMATICALLY	
	LANGUAGE PROBLEMS POSED BY HEAVILY	STRUCTURED DATA	CACM621 28
INTRODUCTORY PROBLEM IN SYMBOL MANIPULATION FOR THE	STUDENT		CACM609 403
AN IR LANGUAGE FOR LEGAL RETRIEVAL	STUDIES		CACM617 300
MECHANICAL PRAGMATICS, A TIME-MOTION	STUDY OF A MINIATURE MECHANICAL LINGUISTIC SYSTEM		CACM620 576
	A SHORT	STUDY OF NOTATION EFFICIENCY	CACM608 466
	ALGOL	SUB-COMMITTEE REPORT-EXTENSIONS	CACM577 24
COMPILER	TWO	SUBROUTINE METHOD FOR CALCULATING LOGARITHMS	CACM545 5
	LOW LEVEL LANGUAGE	SUBROUTINES FOR SYMBOL MANIPULATION WITH AN ALGEBRAIC	CACM612 102
	ON THE COMPILATION OF	SUBSCRIPTED VARIABLES	CACM614 422
	RECURSIVE	SUBSCRIPTING COMPILERS AND LIST-TYPE MEMORIES	CACM614 102
	FLOW OUTLINING, A	SUBSTITUTE FOR FLOW CHARTING	CACM572 4
	EXTRACTION OF ROOTS BY REPEATED	SUBTRACTIONS FOR DIGITAL COMPUTERS	CACM574 17
MS IN ORDINARY DIFFERENTIAL EQUATIONS	SUCCESSIVE APPROXIMATIONS AND COMPUTER STORAGE PROBLE		CACM580 6
ETERMINATION OF THE OPTIMUM RELAXATION FACTOR OF THE	SUCCESSIVE OVER-RELAXATION METHOD	ZNHIQUE FOR THE D	CACM615 222
	THE	SUMADOR CHINO	CACM614 149
D FOR GENERATING UNIFORMLY DISTRIBUTED POINTS ON THE	SURFACE OF AN N-DIMENSIONAL SPHERE	EFFICIENT METHO	CACM604 621
D FOR GENERATING UNIFORMLY DISTRIBUTED POINTS ON THE	SURFACE OF AN N-DIMENSIONAL SPHERE	EFFICIENT METHO	CACM574 17
ALGORITHM	SURGE, A RECODING OF THE COBOL MERCHANDISE CONTROL		CACM570 26
	NATIONAL ACM MEMBERSHIP	SURVEY	CACM622 43
	AUTOMATIC ABSTRACTING AND INDEXING,	SURVEY AND RECOMMENDATIONS	CACM629 470
	ACM MEMBERSHIP	SURVEY JANUARY 1, 1962	CACM615 226
		SURVEY OF CODED CHARACTER REPRESENTATION	CACM626 277
RETRIEVAL	A	SURVEY OF LANGUAGES AND SYSTEMS FOR INFORMATION	CACM600 639
F AUTOMATIC DATA PROCESSING IN BUSINESS AND MANAC/	SURVEY OF PROGRESS AND TREND OF DEVELOPMENT AND USE O		CACM621 43
F AUTOMATIC DATA PROCESSING IN BUSINESS AND MANAC/	SURVEY OF PROGRESS AND TREND OF DEVELOPMENT AND USE		CACM574 22
F AUTOMATIC DATA PROCESSING IN BUSINESS AND MANAC/	SURVEY OF PROGRESS AND TREND OF DEVELOPMENT AND USE		CACM575 17
	SURVEY OF PUNCHED CARD CODES		CACM579 34
	SURVEY OF PUNCHED CARD CODES		CACM600 608
	FURTHER	SURVEY OF PUNCHED CARD CODES	CACM614 12
COMBAT VEHICLE FIRING STABILITY (ACTIVE	SUSPENSION)		CACM616 277
REMARKS ON ALGOL AND	SYMBOL MANIPULATION		CACM577 25
A GENERALIZED TECHNIQUE FOR	SYMBOL MANIPULATION AND NUMERICAL CALCULATION		CACM613 147
A STRING LANGUAGE FOR	SYMBOL MANIPULATION BASED ON ALGOL 60		CACM621 54
	SYMBOL MANIPULATION BY THREADED LISTS		CACM634 145
AN INTRODUCTORY PROBLEM IN	SYMBOL MANIPULATION FOR THE STUDENT		CACM629 408
	SYMBOL MANIPULATION IN XTRAN		CACM604 213
TWO SUBROUTINES FOR	SYMBOL MANIPULATION WITH AN ALGEBRAIC COMPILER		CACM612 102
ACM CONFERENCE ON	SYMBOL MANIPULATION, PROGRAM AND PREPRINTS		CACM604 133
RECURSIVE FUNCTIONS OF	SYMBOLIC EXPRESSIONS AND THEIR COMPUTATION BY MACHINE		CACM624 124
PROPOSED STANDARD FLOW CHART	SYMBOLS		CACM570 17
HANDLING IDENTIFIERS AS INTERNAL	SYMBOLS IN LANGUAGE PROCESSORS		CACM576 21
A PROCEDURE FOR INVERTING LARGE	SYMMETRIC MATRICES		CACM628 445
EIGENVALUES OF A	SYMMETRIC 3X3 MATRIX		CACM614 168
	SYNTACTIC ANALYSIS BY DIGITAL COMPUTER		CACM620 515
AUTOMATIC-PROGRAMMING-LANGUAGE TRANSLATION THROUGH	SYNTACTIC AND SEMANTIC AUGMENTS TO ALGOL		CACM634 211
	SYNTACTICAL ANALYSIS		CACM623 145
	A SYNTACTICAL CHART OF ALGOL 60		CACM619 333
	SYNTACTICAL CHARTS OF COBOL 61		CACM625 200
	A SYNTAX DIRECTED COMPILER FOR ALGOL 60		CACM611 51
DIGITAL	SYNTHESIS OF CORRELATED STATIONARY NOISE		CACM627 400
INDEXING OF THE WORD	*SYSTEM* HAS BEEN PREVENTED AUTOMATICALLY		
ABBREVIATING WORDS	SYSTEMATICALLY		CACM605 323
LOGICAL STATEMENTS TO PRODUCE A TRUTH FUNCTION	TABLE	ALGORITHM FOR ANALYZING	CACM583 4
	TABLE LOOK-AT TECHNIQUES		CACM614 172
	TABLES		CACM616 272
AN ENGINEERING APPLICATION OF LOGIC-STRUCTURE	TABLES		CACM614 516
	TABLES FOR AUTOMATIC COMPUTATION		CACM581 8
	TABLES WITH BINARY-SEARCH FACILITY		CACM582 1
VARIABLE-WIDTH	TALL, A LIST PROCESSOR FOR THE PHILCO 2000 COMPUTER		CACM629 484
A NOTE ON SAMPLING A	TAPE FILE		CACM626 343
FURTHER REMARKS ON SAMPLING A	TAPE FILE, I		CACM620 507
FURTHER REMARKS ON SAMPLING A	TAPE FILE, II		CACM620 508
PROCESSING MAGNETIC	TAPE FILES WITH VARIABLE BLOCKS		CACM610 505
USE OF MAGNETIC	TAPE FOR DATA STORAGE IN THE ORACLE-ALGOL TRANSLATOR		CACM611 15
THE BASIC SIDE OF	TAPE LABELLING		CACM602 35
IBM 709	TAPE MATRIX COMPILER		CACM599 31
	TAPE SPLITTING		CACM614 477
	TAPE SPLITTING IN AN ITERATIVE PROGRAM		CACM622 102
OPTIMUM	TAPE WRITING PROCEDURES		CACM619 339
THE APPLIED MATHEMATICS LABORATORY OF THE DAVID W.	TAYLOR MODEL BASIN		CACM619 372
GUIDES TO	TEACHING COBOL		CACM625 272
INDEXING OF THE WORD	*TECHNIQUE* HAS BEEN PREVENTED AUTOMATICALLY		
THE STATE OF DIGITAL COMPUTER	TECHNOLOGY IN EUROPE		CACM616 256
SOVIET COMPUTER	TECHNOLOGY, 1959		CACM603 121
SOME BASIC	TERMINOLOGY CONNECTED WITH MECHANICAL LANGUAGES AND		CACM618 336
THEIR PROCESSORS	TERMINOLOGY PROJECT		CACM607 409
	A TERMINOLOGY PROPOSAL		CACM612 72

V. APPLICATIONS OF KWIC INDEXING

KWIC indexes are used in such a variety of applications, that it would be difficult to enumerate all of them. Donald Kraft, information retrieval expert of the IBM Corporation has summarized several uses.

Kraft says:

"KWIC indexes are used regularly by a number of companies and organizations. Perhaps the most well known is Chemical Titles, published twice a month by the American Chemical Society as a current awareness program. Each issue indexes approximately 2,800 titles of chemistry. Biological Abstracts use a KWIC index in each monthly issue. Bell Telephone Laboratories use a KWIC index to disseminate technical information to its staff. Trans-Canada Airlines indexes its procedural manuals with KWIC indexes. In January 1963 the American Bar Foundation began publication of a KWIC Index of bills enacted by the 50 state legislations. This current awareness service is published approximately every two weeks during the height of the legislative sessions. In July 1962 the American Bar Foundation published its first KWIC Index, Index to Legal Theses and Research Projects - 1960-1961. ...IBM uses KWIC indexes for its collection of marketing publications and procedure manuals as well as for indexes to current IBM research projects and computer program libraries." (74)

In a letter to the author dated April 20, 1964 Mr. Kraft mentioned additional applications of KWIC indexing. One of these called a "Responsibility Index", seemed especially novel. The index, used at IBM Headquarters, permutes the job titles of Headquarters personnel and instead of a reference code a man's name is used. "Consequently one can use this to determine who is responsible for what. ...The American Medical Association uses a KWIC index to the signs and symptoms of diseases in their publication entitled "Current Medical Terminology." Northwestern University, Department of Languages (Greek), used a KWIC index to study the writings

of Galen. In this example, the entire text of Galen was keypunched in a transliterated form and treated as if it were one title. The listing proved to be very helpful. Kansas State University produced a KWIC index called "Kansas Slavic Index", which is a transliteration of Russian sociological titles. In closing, Kraft mentioned the work of Professor Kenneth Janda of the Political Science Department at Northwestern University. Professor Janda, to the surprise of this author, has KWIC indexed 236 recent political science doctoral dissertations. The author assumes that these titles were edited but from Professor Janda's index it is impossible to discern either the presence or amount of editing. Perhaps Professor Janda's political science index suggests an ever broadening scope of activity for KWIC indexing.

VI. CONCLUSIONS

The flood of research papers, scientific articles, government reports and documents of every description, which began with World War II and continues unabated today, compounded by a multiplicative analysis factor (the number of analytics, extracts, index entries, etc. per document) has generated an alarming information morass. The morass per se is harmless fire consumes paper. The harm is in the result of the morass - the difficulty in processing and retrieving useful information. The severity of the information problem has sent librarians, documentalists and information scientists in quest of means to expedite the handling of literature. Dr. Dwight Gray of the National Science Foundation, who equates information retrieval with experimentation, recently wrote "...the processing and dissemination of the results of research - that is of scientific information - is as integral a part of the total research sequence as experimentation is." He then concluded, "the prerequisite to a solution to this problem is realistic recognition of information dissemination as a blood brother of experimentation..."⁽⁷⁵⁾

Science, at the appeal of the information handlers and given impetus by interested constituents and government funding, is applying mechanical innovations - notably today's newest breeds of lightning fast computers - to all aspects of the information handling problem. Clearly there are vast horizons for mechanization in information handling, for in any information storage and retrieval system there are only two tasks which are exclusively assignable to man - the choice of documents to go into the system and the final evaluation of the usefulness of answers by the user of the information.

In between these two points is a series of unit operations - all of which may be performed by human beings and some of which may be performed wholly or in part by various machines.

Within the analysis unit operation, science has applied machines to abstracting, extracting, classifying and with distinct success to indexing. And of the various forms of machine indexing none has been as successful and widely adopted as keyword-in-context indexing (KWIC), the subject of this thesis.

After traversing the deductive path to machine indexing and journeying along the path stopping in succession at a description of KWIC indexing, a study of the adequacy of titles as search tools, pros and cons, and most recently at applications of KWIC indexing, the reader might conclude that the author regards KWIC indexing as a panacea. Not so! KWIC indexing does not impinge on nor will it replace other modes of indexing, with the possible exception of the shallow alphabetic subject index. A cure-all it is not: but a new and efficacious weapon to attack the information problem it is!

The author feels that the following principal conclusions may be elicited from the preceding discourse of KWIC indexing:

- (A) Titles are adequate search tools, particularly in highly specialized fields.
- (B) KWIC indexing, despite the repairability of non-descriptive titles, is best suited for highly specialized fields.
- (C) The KWIC index is useful for either dissemination or retrieval.

(D) "KWIC" offers many advantages, such as speed, ease of preparation, flexibility and cost, but few disadvantages.

(E) Any specialized library or information center, given sufficient funds, or direct access to the machines necessary for preparing a KWIC index, or both, should give serious consideration to initiating this mode of indexing.

And finally, a comment regarding the study of the adequacy of titles as search tools is appropriate. As stated earlier, several similar studies to test titles have been conducted in the past. However, their results are not readily comparable and suggest limited inferences. This is because these studies covered very specialized fields, i.e., medicine, law, etc. and were performed by different individuals using different procedures and criteria. The study in this thesis spanned science, engineering, economics, business, social science, public affairs and the humanities. It was performed exclusively by the author, thus eliminating inter operator error and, in the author's estimation, lending additional support to its results.

VII. FUTURE

In the future, new applications of KWIC indexing will appear with great frequency now that the advantages of this system have been demonstrated. If the information to be indexed can be described in roman letters, it can be handled by a computer by using a keyword-in-context program. As KWIC indexes become available in more and more areas of knowledge, it is possible that authors will bear in mind to an even greater extent the later retrievability of their papers when writing the title. At Bell Telephone Laboratories, to encourage further improvement in titling, its Technical Information Libraries recently prepared and distributed to more than 5000 of the Laboratories' scientific and engineering personnel a concise guide on choosing titles for their papers. (76)

W. W. Youden, data processing expert of the National Bureau of Standards, sees a bright future for computer-produced indexes. The author includes Mr. Youden's comment because it provides the desired prophetic note on which to end this thesis. Here is what Mr. Youden has to say on the subject:

"The use of computers in the production of indexes will increase. By feeding the output of the computer into a photo composition machine there will be an improvement in both the quality of the printing and the size of the character set. Today, the entire cost of a computer run to produce one copy of an index should be less than the cost of keystroking the input. The facility with which indexes may be updated and the ease of selecting items for special bibliographies will result in the majority of indexes being computer produced before many years. As more and more keyword indexes are published the author's choice of title words will improve, which will lead to more and better keyword indexes. Because of the intellectual labor involved

it is unlikely that even with computers the amount of subject-indexing will increase greatly.

...The day can not be too far off when printed indexes to scientific and technical libraries will be available to the individual researcher at his desk at a reasonable cost." (77)

APPENDIX A

Lockheed Aircraft Corporation Materials

APPENDIX A

This appendix offers four examples of computer-produced outputs from the newly installed mechanized information service system at the Lockheed Aircraft Corporation, Marietta, Georgia. The outputs are: (a) an accession list announcing all information material received, prepared on Multilith stencils for large reproduction (see Figure 12); (b) a cumulative index of report numbers in lieu of a conventional card file, with a carbon copy for each information center within company (see Figure 13); (c) a mechanically prepared subject index (for each information center through carbon copy print-out process) in lieu of conventional card file, using a "Keyword-In-Context" title system (see Figure 14); and (d) a cumulative personal author index for books and reports in lieu of a conventional card file with a carbon copy for each information center (see Figure 15).

MATHEMATICS

R64-00002544
O Z

MATHEMATICAL MODELS FOR SYSTEM-ANALYSIS

AD 419609

AF INST OF TECH, WPAFB, OHIO.

513.83M966E
S

ELEMENTARY DIFFERENTIAL TOPOLOGY

514.083P482E
S

EIGHT-PLACE TABLES OF TRIGONOMETRIC FUNCTIONS FOR EVERY SECOND OF ARC

517.52H393M
S

MULTILINEAR ANALYSIS FOR STUDENTS IN ENGINEERING + SCIENCE

517J13M
A Z

MATHEMATICS FOR QUANTUM MECHANICS

MECHANICAL EQUIPMENT

621.822B628A
S

ADVANCED BEARING TECHNOLOGY

MEDICAL SCIENCE

R64-00002308
A Z

THE EFFECTS OF TRANSIENT WEIGHTLESSNESS ON BRIGHTNESS DISCRIMINATION

AMRL TDR 64 12

CORNELL AERON LAB, INC, BUFFALO.

MEDICAL SCIENCES

R64-00002303
A Z

EFFECTS OF SIMULATED AEROSPACE SYSTEM ENVIRONMENTS ON THE GROWTH OF SELECTED ANGIOSPERMS

AMRL TDR 63 131

BOEING CO, SEATTLE.

METALLURGY

R64-00002075
L Z

EFFECT OF POLYSLIP ON STRAIN-HARDENING IN ALUMINUM SINGLE CRYSTALS

UCRL 11160

CALIF. UNIV, BERKELEY, LAWRENCE RAD LAB.

METALLURGY

R64-00002156
L Z PREPARATION + EVALUATION OF HIGH-PURITY BERYLLIUM
Q-81933-5 FRANKLIN INST. LABS FOR RES + DEVL. PHILA.

R64-00002158
L Z THE ANALYSIS OF FRACTURE SURFACES BY ELECTRON-MICROSCOPY
D1-82-0169 BOEING SCI RES LABS. SEATTLE.

R64-00002190
S SOVIET-RESEARCH IN POWDER-METAL-MATERIALS * METALLURGY METALS
JPRS 17 887 JOINT PUB RES SERV. NY.

R64-00002191
S ARTICLES ON SOVIET POWDER-METALLURGY * METALLURGY
JPRS 18 407 JOINT PUB RES SERV. NY.

R64-00002239
L Z METASTABLE AMORPHOUS PHASES IN TELLURIUM-BASE ALLOYS
CIT TR 22 CALIF INST OF TECH. PASA.

R64-00002362
S SOME ASPECTS OF THE HYDROSTATIC PRESSING OF POWDERS
JPRS 18 407 JOINT PUB RES SERV. NY.

R64-00002417
S SOVIET-EXPERIMENTS WITH TUNGSTEN MOLYBDENUM AND ALLOYS OF REFRACTORY-METALS * METALS
JPRS 17 955 JOINT PUB RES SERV. NY.

R64-00002435
S DEVELOPMENTS IN CZECHOSLOVAK POWDER-METALLURGY
JPRS 20 291 JOINT PUB RES SERV. NY.

R64-00002437
S POSSIBILITIES + PROBLEMS OF POWDER-METALLURGY IN THE DEVELOPMENT OF NEW-MATERIALS
JOINT PUB RES SERV. NY.

R64-00002445
S METALLURGICAL PROGRESS-REPORT * CERAMIC-MATERIAL METALS
USBM U 1109 BUREAU OF MINES. ALBANY. ORE.

CONTROL NO.

INDEX ALPHABETIC

CONTROL NO.

[illegible][illegible]

ARC	151	02	003
ARC	152		
ARC	153		
ARC	211		
ARDC	TR	57	20
ARDC	TR	59	66
ARDC	TR	59	66
ARDC	TR	59	66
ARDC	TR	59	66
ARF	B183		13
ARF	B241		13
ARF	B241		13
ARF	B241		13
ARF	B241		13
ARF	B241		13
ARF	PR	7	
ARF	PROJ	D11H	
ARF	112C	55	
ARF	2211	52	
ARF	2241	52	
ARF	3200	12	
ARF	5134	88	
ARF	5134	88	
ARF	5186	12	
ARF	5186	23	
ARF	5186	33	
ARF	-B785	-2E	
ARF	-1214	-TR	
ARF	-2212	-S	
ARINC		548	
ARINC	CHA	52	10
ARINC	CHAR	555	
ARINC	CHAR	557	
ARINC	NO	409A	
ARINC	NO	552	
ARINC	234		325
ARL	MK	47	
ARL	NC	1	
ARL	SM	287	
ARL	SU	287	
ARL	SM	4	
ARL	TCH	62	24

R63-0000566
R63-00002966
R63-00002966
R63-00002965
R63-00003101
R63-00005155
R63-00002952
R63-00002110
R63-00002255
R62-00008351
R62-00008350
R63-00008327
R64-00003065
R63-00006497
R63-00008141
R63-00007009
R63-00007308
R63-00005294
R63-00002341
R63-00001685
R63-00001124
R63-00006490
R63-00006673
R64-00006673
R64-00003209
R64-00003209
R63-00003209
R63-00005159
R63-00007286
R63-00008139
R64-00001240
R62-00007615
R62-00007680
R63-00006960
R63-00008925
R64-00008978
R64-00000779
R62-00007407
R62-00007681
R63-00001610
R63-00007859
R63-00001508
R63-00005715
R63-00002457
R64-00002456
R64-00002163

DPR 58-768

QUARTERLY KWIC INDEX

DATE 04-09-

KEY WORD

TITLE

CONTROL

ACADEMY	CE-ABSTRACTS PROCEEDINGS	
ACCELERATORS	REPORTS OF THE ACADEMY OF SCIENCES OF USSR PHYSICAL CHEMISTRY (SELECTED ARTICLES)	R 64-00
ACCELETRA	1963 SUMMER STUDY REPORT PART ONE SUMMER 1963 • ACCELERATORS NUCLEAR-PHYSICS	R 64-00
ACCELERATING	PHOTOELECTRIC EARPIECE RECORDINGS + OTHER PHYSIOLOGIC VARIABLES AS OBJECTIVE METHODS OF MEASURING THE INCREASE IN TOLERANCE TO HEADWARD ACCELERATION	R 64-00
ACCELERATION	THE ACCELERATING EFFECT OF ADDITIVES ON RADIATION-INDUCED GRAFT POLYMERIZATION • SOLVENT-ACTION	R 64-00
ACCELERATION	VESTIBULAR SUPPRESSION • ACCELERATION	R 64-00
ACCELERATION	DECREASES IN ARTERIAL OXYGEN SATURATION AS AN INDICATOR OF THE STRESS IMPOSED ON THE CARDIORESPIRATORY SYSTEM BY FORWARD ACCELERATION	R 64-00
ACCELERATIONS	WEIGHTLESSNESS AND ACCELERATION	R 64-00
ACCELERATIONS	SOVIET-STUDIES ON REACTIONS TO ACCELERATION	R 64-00
ACCELERATIONS	HUMAN TRACKING PERFORMANCE UNDER TRANSVERSE ACCELERATIONS	R 64-00
ACCELERATIONS	THE VERTICAL LONGITUDINAL + LATERAL ACCELERATION EXPERIENCED BY AN S. E. 5A AIRPLANE WHILE MANEUVERING • AERODYNAMIC-CONFIGURATIONS	R 64-00
ACCELERATOR	TWO-MILE ACCELERATOR PROJECT QUARTERLY STATUS REPORT 1 JULY TO 30 SEPTEMBER 1963	R 64-00
ACCELERATORS	BOEING RADIATION EFFECTS LABORATORY • RESEARCH LINEAR ACCELERATORS	R 64-00
ACCEPTANCE-CRITERIA	C-141A SONIC FATIGUE-TEST OF BRAZED HONEYCOMB SANDWICH-PANELS FOR ACCEPTANCE-CRITERIA C-141A TAILPIPE (AIRCRAFT) (RELIABILITY)	R 63-ER
ACCEPTANCE-PROBABILITY	SINGLE SAMPLING INSPECTIONS-PLANS WITH SPECIFIED ACCEPTANCE-PROBABILITY + MINIMUM-COSTS • PROBABILITY	R 64-00
ACCESS-DOOR	C-141A FATIGUE TEST OF A LOWER SURFACE ACCESS-DOOR CENTER-WING PANEL (BL 19.25L)	R 64-ER
ACCESSION-LIST	RADIATION-EFFECTS INFORMATION-CENTER MONTHLY ACCESSION-LIST ABSTRACT NO 21127 TO 21177 • BIRL IOGRAPHY BATTELLE	R 64-00
ACCIDENTS	ARGONNE-NATIONAL-LABORATORY RADIOLOGICAL-ASSISTANCE TEAM • ACCIDENTS SAFETY	R 64-00
ACCIDENTS	A SUMMARY OF INDUSTRIAL-ACCIDENTS IN USAEC-FACILITIES 1961-1962 ACCIDENTS SAFETY	R 63-00
ACCOMMODATIONS	CONTINUED USE OF UNECONOMICAL FIRST-CLASS AIR-TRAVEL ACCOMMODATIONS BY EMPLOYEES OF DEFENSE CONTRACTORS	R 63-00
ACCOMPLISHMENTS	THE ACCOMPLISHMENTS OF AIRPOWER IN THE MALAYAN-EMERGENCY (1918-1960) • GUERRILLA-WARFARE WARFARE	R 64-00
ACCOUNTING	HANDBOOK OF SAMPLING FOR AUDITING + ACCOUNTING	311.21
ACCURATELY	A LIGHTWEIGHT FOLDING BOOM FOR ACCURATELY PLACING SCIENTIFIC EXPERIMENTS UP TO 25 FEET FROM LARGE SPACECRAFT	R 64-00
ACETONITRILE	ELECTRON-SPIN RESONANCE ESR STUDY OF GAMMA-IRRADIATED SOLID ACETONITRILE • RADIATION-EFFECTS CHEMISTRY	R 63-00

CPR 59-762

PERSONAL AUTHOR	CONTROL NO.
CHAPMAN	D R R63-00008244
CHAPMAN	D R R63-00003739
CHAPMAN	D R R63-00003427
CHAPMAN	D R R63-00001263
CHAPMAN	G R64-00001312
CHAPMAN	G T R63-00004897
CHAPMAN	G T R63-00008035
CHAPMAN	M C R63-00007730
CHAPMAN	R F R64-00001274
CHARCZENKO	N R63-00002900
CHARCZENKO	N R63-00006366
CHARLESRY	A R63-00002485
CHARLTON	H R63-00002091
CHARNES	A R64-00001199
CHARNES	A R64-00000927
CHARTENER	W H R63-00002896
CHARTENER	W H R63-00004711
CHARTERS	A C R64-00000852
CHARZENKO	N R63-00007606
CHASE	L F R64-00002160
CHASTELLUX	M C 917C489T
CHATLAND	H R63-00000147
CHATLAND	H R63-00000147
CHAUMONT	J M R64-00002169
CHAUVIN	L T R63-000026723
CHAVCHANIDZE	V V R63-00007771
CHEATHAM	P G R63-00006179
CHEESMAN	H K R63-00000518
CHEEVER	J 813.54C515W
CHEN	H H R63-00005873
CHENEY	K B R63-00006039
CHENG	H K R63-00002926
CHENG	H K R63-00000462
CHENG	H K R64-00002455
CHECRVAS	W R62-00007757
CHECRVAS	W R63-00004131
CHECRVAS	W R63-00000875
CHERENKOV	P A R63-00001427
CHERNEV	I 794.1C521W
CHERNIACK	J R R63-00004227
CHERNICK	J R63-00003180
CHERRY	W H R63-00002254
CHERRY	W R R64-00001222
CHERVAS	W R63-00001596
CHESNUT	M B 973.782C524D
CHESSEMAN	C E R64-00001664

PERSONAL AUTHOR INDEX

PERSONAL AUTHOR	CONTROL NO.	PER
CHESSICK	J J R62-00007688	CHR
CHESSICK	J J R63-00002575	CHR
CHESWORTH	R H R64-00001066	CHR
CHESWORTH	R H R63-00007500	CHR
CHESWORTH	R H R63-00004766	CHR
CHESWORTH	R H R63-00007882	CHR
CHESWORTH	R H R63-00006378	CHR
CHESWORTH	R H R63-00001820	CHR
CHESWORTH	R H R63-00002756	CHR
CHESWORTH	R H R64-00001634	CHR
CHESWORTH	R H R64-00002092	CHR
CHESWORTH	R H R64-00001071	CHR
CHEVALIER	H L R62-00007493	CHR
CHEVALIER	H L R63-00007750	CHU
CHEW	G F 530.12C529S	CHU
CHEW	W L R63-00007407	CHU
CHEZEM	C G R63-00006071	CHU
CHI	S H R62-00008056	CHU
CHICK	B R64-00000591	CHU
CHICK	B R62-00007958	CHU
CHICK	H J R62-00008313	CHU
CHIEN	C Y R63-00003089	CHU
CHILDRESS	S R63-00001897	CHU
CHILDRESS	S R64-00000930	CIC
CHILDS	C B R63-00001712	CIF
CHILDS	W O R63-00002940	CID
CHIN	G Y R63-00003585	CKH
CHINOWSKY	W R63-00002435	CLA
CHISMAN	S W R63-00000354	CLA
CHISMAN	P H R63-00003193	CLA
CHIU	H H R63-00005549	CLA
CHIU	H H R63-00003341	CLA
CHIU	H Y 531.5C543G	CLA
CHOW	K T R62-00002906	CLA
CHOW	K T R62-00002906	CLA
CHOW	K T R63-00005308	CLA
CHOW	K T R63-00000757	CLA
CHOW	R R R63-00004440	CLA
CHOW	W L R63-00003271	CLA
CHRISTENBERRY	W W R63-00007435	CLA
CHRISTENSEN	A J R63-00007158	CLA
CHRISTENSEN	G R62-00007846	CLA
CHRISTENSEN	J A R64-00000628	CLA
CHRISTENSEN	J A R64-00000629	CLA
CHRISTENSEN	J M R63-00006171	CLA
CHRISTENSEN	J M R63-00008106	CLA

1

PERSONAL AUTHOR INDEX

5-11-64

PAGE 28

CONTROL NO.	PERSONAL AUTHOR	CONTROL NO.	PERSONAL AUTHOR	CONTROL NO.
3-00000244	CHESSICK	J J R62-00007688	CHRISTENSEN	L M R63-00001100
3-00000379	CHESSICK	J J R63-00002575	CHRISTENSEN	L M R64-00000545
3-000003427	CHESWORTH	R H R64-00001066	CHRISTENSEN	L M R63-00002909
3-000001263	CHESWORTH	R H R63-00007500	CHRISTENSEN	L M R63-00005384
3-000001312	CHESWORTH	R H R63-00004766	CHRISTENSEN	R H R63-00002892
3-000004897	CHESWORTH	R H R63-00007882	CHRISTENSEN	W I R63-00004549
3-000008135	CHESWORTH	R H R63-00006378	CHRISTIAN	J B R63-00007095
3-000007730	CHESWORTH	R H R63-00001820	CHRISTIAN	J L R63-00003334
3-000001274	CHESWORTH	R H R63-00002756	CHRISTIAN	J L R63-00000505
3-000002900	CHESWORTH	R H R64-00001634	CHRISTIANA	J J R63-00006819
3-000006466	CHESWORTH	R H R64-00002092	CHRISTIANA	J J R63-00004776
3-000002485	CHESWORTH	R H R64-00001071	CHRISTIANS	H R63-00004571
3-000002091	CHEVALIER	H L R62-00007493	CHRISTMA	R J R63-00001855
3-000001199	CHEVALIER	H L R63-00007750	CHU	F L R63-00000643
3-000000927	CHEW	G F 530.1205295	CHUAN	R L R63-00003713
3-000002896	CHEW	W L R63-00007407	CHUNG	K T R63-00007960
3-000004711	CHEZEM	C G R63-00006071	CHUNN	F H R64-ER005649
3-000000852	CHI	S H R62-00008056	CHURCH	G W R62-00007829
3-000007406	CHICK	B R64-00000591	CHURCH	J S R63-00001833
3-000002160	CHICK	B R62-00007958	CHURCHILL	D R R63-00001253
3-000004897	CHICK	H J R62-00008313	CHURCHILL	D R R64-00002278
3-000000147	CHIEN	C Y R63-00003089	CHURCHILL	N C R63-00001149
3-000000147	CHILDRESS	S R63-00001897	CHUSHKIN	P I R64-00002472
3-000002169	CHILDRESS	S R64-00000930	CICOLANI	L S R64-00000272
3-000006723	CHILDS	C B R63-00001712	CIEPLUCH	C C R62-00008084
3-000007771	CHILES	W D R63-00002940	CIOBKOWSKI	S R63-00003669
3-000006179	CHIN	G Y R63-00003585	CKHEED CALIF	CO R63-00006742
3-00000518	CHINOWSKY	W R63-00002435	CLAPP	C W R62-00007448
3-00000515W	CHISMAN	S W R63-00003354	CLAPPER	W E R62-00007379
3-000005873	CHISMAN	P H R63-00003193	CLARE	K G R63-00001206
3-000006029	CHIU	H H R63-00005549	CLARE	V R63-00000779
3-000002026	CHIU	H H R63-00003341	CLARK	A G R63-00004937
3-00000462	CHIU	H Y 531.50543G	CLARK	B R63-00008610
3-000002455	CHOW	K T R62-00002906	CLARK	C K R63-00007732
3-000007747	CHOW	K T R62-00002906	CLARK	D A R62-00007850
3-000004131	CHOW	K T R63-00005308	CLARK	D C R62-00008117
3-000002875	CHOW	K T R63-00000757	CLARK	D F R63-00004287
3-000001427	CHOW	R R R63-00004440	CLARK	F L R64-00001503
3-00000521	CHOW	W L R63-00003271	CLARK	F H 671.37C592A
3-000004227	CHRISTENBERRY	W R63-00007435	CLARK	F L R63-00007608
3-000003180	CHRISTENSEN	A J R63-00007158	CLARK	G L 539.752C593F
3-000002254	CHRISTENSEN	G R62-00007846	CLARK	J A 536.2C593T
3-000001222	CHRISTENSEN	J A R64-00000628	CLARK	J S R64-00000340
3-000001596	CHRISTENSEN	J A R64-00000629	CLARK	L R R63-00008037
3-000005240	CHRISTENSEN	J M R63-00006171	CLARK	L R R63-00006798
3-000001664	CHRISTENSEN	J M R63-00008106	CLARK	L V R63-00000459

2

BIBLIOGRAPHY

1. Smithsonian Institution. Board of Regents. Report to the Senate and House of Representatives Showing the Operations, Expenditures, and Condition of the Institution During the Year 1851. Washington, D.C.: A Boyd Hamilton, pp. 22, 1852.
2. Adams, Henry. The Education of Henry Adams. Boston: Houghton Mifflin, p. 498, 1913.
3. Shera, J.H., The Propaedeutic of the New Librarianship. In "Information Retrieval Today" papers presented at the institute conducted by the Library School and the Center for Continuation Study, University of Minnesota, pp. 5-19, September 1962.
4. Bush, V. "As We May Think", Atlantic Monthly, Vol. 176, pp. 101-108, July 1940.
5. Bourne, C.P., "The World's Technical Journal Literature: An Estimate of Volume. Origin Language, Field, Indexing, and Abstracting", American Documentation, Vol. 13, pp. 159-168, April 1962.
6. Schultz, L., "The Information System: Too Big and Growing", American Documentation, Vol. 13, p. 288, July 1962.
7. "Current Concentrates of the Library World", Special Libraries, Vol. 53, No. 6, p. 338, July-August 1962.
8. Bello, F., "How to Cope with Information," Fortune, Vol. 62, Pt. 1, p. 162, September 1960.
9. Ibid
10. Bourne, C.P., "The World's Technical Journal Literature: An Estimate of Volume, Origin Language, Field, Indexing, and Abstracting," American Documentation, Vol. 13, pp. 159-168, April 1962.
11. Swanson, D.R., "Library Goals and the Role of Automation", Special Libraries, Vol. 53, No. 8, pp. 466-471, October 1962.
12. Bar-Hillel, Y., "Is Information Retrieval Approaching a Crisis?," American Documentation, Vol. 13, pp. 95-98, April 1963.
13. Bourne, C.P., "The Historical Development and Present State-of-the Art of Mechanized Information Retrieval Systems", American Documentation, Vol. 11, pp. 108-110, April 1961.

14. Wooster, H., Long Range Research in Information Retrieval. In "Information Retrieval Today" papers presented at the institute conducted by the Library School and the Center for Continuation Study, University of Minnesota, pp. 149-164, September 1962.
15. "Current Concentrates of the Library World," Special Libraries, Vol. 53, No. 6, pp. 338, July-August, 1962.
16. Heald, J.H., "ASTIA and the Information Revolution", Special Libraries, Vol. 54, No. 1, p. 40, January 1963.
17. Taube, Mortimer and Webster, Harold. Information Storage and Retrieval. New York: Columbia University Press, pp. 15-17, 1958.
18. Kent, Allen. Textbook on Mechanized Information Retrieval. New York: Interscience Publishers a Division of John Wiley and Sons, Inc., p. 9, 1962.
19. Kent, Allen, Textbook on Specialized Information Centers, Unpublished, copyright, 1964.
20. Kent, Allen. Textbook on Mechanized Information Retrieval. New York: Interscience Publishers a Division of John Wiley and Sons, Inc., p. 83 1962.
21. Kent, Allen. Textbook on Specialized Information Centers. Unpublished, copyright, 1964.
22. Kent, Allen, Information Retrieval and Machine Translation, Part 2. New York: Interscience Publishers, Inc., pp. 789-791, 1961
23. Ibid.
24. Mooers, C.N., The Indexing Language of an Information Retrieval System. In "Information Retrieval Today" papers presented at the institute conducted by the Library School and the Center for Continuation Study, University of Minnesota, p. 21, September 1962.
25. Kent, Allen. Textbook on Specialized Information Centers. Unpublished, copyright, 1964.
26. Kent, Allen. Textbook on Mechanized Information Retrieval. New York: Interscience Publishers a Division of John Wiley and Sons, Inc., p. 9, 1962.
27. Jahoda, G., "A Technique for Determining Index Requirements", American Documentation, Vol. 14, pp. 82-85, 1964.
28. Jahoda, G., "The Development of a Combination Manual and Machine-Based Index to Research and Engineering Reports", Special Libraries, Vol. 53, No. 2, pp. 74-78, February 1962.

29. Jahoda, G., Indexing with Edge-Notched and Internally Punched Cards. In "Information Retrieval Today" papers presented at the institute conducted by the Library School and the Center for Continuation Study, University of Minnesota, pp. 37-39, September 1962.
30. Jonker, T., "The Descriptive Continuum", AFOSR TN57-287 (AD132358), June 1957.
31. McKenna, T.E., "Readin', Ritin', and Reproduc'n: Tools for the Special Librarian", Special Libraries, Vol. 53, No. 9, November 1962.
32. Kent, Allen. Information Retrieval and Machine Translation, Part 2. New York: Interscience Publishers, Inc., pp. 789-791, 1961.
33. Artandi, S., "A Selective Bibliographic Survey of Automatic Indexing Methods", Special Libraries, Vol. 54, No. 10, pp. 630-634, December 1963.
34. Ibid.
35. Climenson, W.D., Hardwick, N.H., and Jacobson, S.N., "Automatic Syntax Analysis in Machine Indexing and Abstracting," American Documentation, Vol. 12, pp. 178-183, July 1961.
36. Perrin, Porter G. Writer's Guide and Index to English. Third ed. Chicago: Scott Foresman and Co., pp. 214-217, 1959.
37. Ibid.
38. Artandi, S., Op. cit.
39. Climenson, W.D., et.al., Op cit.
40. Borko, H., "Progress Report of the Information Retrieval and Linguistics Project", SDC document TM-676, January 1962.
41. Luhn, H.P. Keyword-In-Context Index for Technical Literature (KWIC Index). IBM Advanced System Division, Yorktown Heights, New York. pp. 4-6, August 1959.
42. Wadding, R. V., "Keyword-In-Context (KWIC) Indexing on the IBM 7090 DPS", Report 62-825-440, IBM Federal Systems Division, Owego, New York. p. 8, September 1962.
43. Balz, Charles F. and Stanwood, Richard H. Literature on Information Retrieval and Machine Translation. IBM Federal Systems Division, Owego, New York. Tearsheets, November 1962.
44. Keyword-In-Context (KWIC) Indexing. E20-8091, IBM Technical Publications Department, White Plains, New York. pp. 10-11, 1962.

45. Ibid, p. 17.
46. Luhn, H.P., Op. cit.
47. Montgomery, C. and Swanson, D.R., "Machine Indexing by People", American Documentation, Vol. 13, pp. 359-366, October 1962.
48. Maizell, R.E., "Value of Titles for Indexing Purposes", Reveu de la Documentation, Vol. 27, pp. 126-127, August 1960.
49. Artandi, S., Op. cit.
50. Kraft, D.H., "A Comparison of Keyword-In-Context (KWIC) Indexing of Titles with a Subject Heading Classification System," American Documentation Vol. 15, pp. 48-52, January 1964.
51. Ibid.
52. Ibid.
53. Engineering Index. New York: Engineering Index Inc., 1962.
54. Cochran, William G. Sampling Techniques. Second edition, New York and London: John Wiley and Sons, Inc., pp. 71-75, 1963.
55. Levy, N.P., "A Survey on the Information Practices of Engineers at Western Electric", American Documentation, Vol. 15, pp. 86-88, April 1964.
56. Semon, T.J., Cohen, R. and Richmond, S.B., "Sampling in Marketing Research", Journal of Marketing, pp. 263-273, January 1959.
57. Duncan, Acheson J. Quality Control and Industrial Statistics. Homewood, Illinois: Richard D. Irwin, Inc., pp. 438, 1959.
58. Cochran, William G., Op. cit., p. 58.
59. Duncan, Acheson J., Op. cit., p. 436.
60. Applied Science and Technology Index. New York: H.W. Wilson Company, 1962.
61. ASM Review of Metal Literature. Metals Park, Ohio: American Society for Metals, Vol. 19, 1962.
62. Public Affairs Information Service, Annual Cumulated Bulletin. New York: Public Affairs Information Service, Inc., Vol. 48, October 1962-September 1963.
63. Business Periodicals Index. New York: H.W. Wilson Company, July 62-June 1963.

64. International Index. New York: H.W. Wilson Company, April 1962-March 1963.
65. Balz, C.F. and Stanwood, R.H., "On Preparing Information for KWIC Indexing," Report 62-816-729, IBM Federal Systems Division, Owego, New York, pp. 1-17, January 1962.
66. Balz, C.F. and Stanwood R.H., "Some Applications of KWIC Indexing," Report 62-825-475, IBM Federal Systems Division, Owego, New York, pp. 1-12, June 1962.
67. Keyword-In-Context (KWIC) Indexing. E20-8091, IBM Technical Publications Department, White Plains, New York, p. 3, 1962.
68. Ibid.
69. Balz, C.F. and Stanwood R.H., "Some Applications of KWIC Indexing", Report 62-825-475, IBM Federal Systems Division, Owego, New York, pp. 1-12, June 1962.
70. Kraft, D.H., "Application of IBM Equipment to Library Mechanization, Keyword-In-Context (KWIC) Indexing and the Selective Dissemination of Information (SDI)", paper presented to the clinic on Library Applications of Data Processing, University of Illinois, February 1964.
71. Bauer, C.K., "Practical Application of Automation in a Scientific Information Center - A Case Study", Special Libraries, Vol. 55, No. 3, pp. 137-142, March 1964.
72. Wyllys, R.E., "On Adequacy of Titles as Search Tool", SDC document TM-1640, 17 December 1963.
73. Youden, W.W., Characteristics of Programs for KWIC and other Computer Produced Indexes. In "Automation and Communication" short papers contributed to the theme sessions of American Documentation Institute, 26th Annual meeting, Part 2: pp. 331-332, October 1963.
74. Kraft, D.H., "A Comparison of Keyword-In-Context (KWIC) Indexing of Titles with a Subject Heading Classification System," American Documentation Vol. 15, pp. 48-52, January 1964.
75. Gray, Dwight E., Science, Vol. 137, July 1962.
76. Kennedy, R.A., Writing Information Titles for Technical Papers - A Guide to Authors. In "Automation and Communication" short papers contributed to the theme session of American Documentation Institute 26th annual meeting, Part 2: pp. 133-134, October 1962.
77. Youden, W.W., Op. cit.