AUTOMATED INFORMATION STORAGE AND RETRIEVAL

A STUDY AND DESIGN
OF AN AUTOMATED
INFORMATION STORAGE AND RETRIEVAL SYSTEM

By

RAYMOND JAMES HICK, B. Eng.

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Engineering

McMaster University

October 1964

MASTER OF ENGINEERING (1964)
(Mechanical Engineering)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: A Study and Design of an Automated Information Storage and Retrieval System

AUTHOR: Raymond James Hick, B. Eng. (McMaster University)

SUPERVISOR: Professor J. N. Siddall

NUMBER OF PAGES: v, 86

SCOPE AND CONTENTS:

The purpose of this project has been twofold. First, a thorough literature survey has been conducted in the field of information storage and retrieval with primary attention given to computer oriented techniques. The systems used and the theory behind them have been of more concern to the author than the actual equipment used because it is felt that equipment will change very rapidly as it has in the past whereas the basic system, in most cases, will prevail.

The second part of this project has been the design of an overall automated trial information system with the needs of the designer being of prime importance.

ACKNOWLEDGEMENTS

THE AUTHOR WISHES TO THANK HIS SUPERVISOR, MR. J.N. SIDDALL FOR HIS CONTINUED ENCOURAGEMENT, ADVICE AND AID THROUGHOUT THE ENTIRE PROGRAM.

IN ADDITION, VALUABLE ASSISTANCE WAS GIVEN BY DR. G. KEECH IN COMPUTER TECHNIQUES AND OPERATION.

THE AUTHOR WISHES TO ACKNOWLEDGE THE RECEIPT OF A MCMASTER UNIVERSITY GRADUATE SCHOLARSHIP AND THE AID OF A RESEARCH GRANT FROM THE NATIONAL RESEARCH COUNCIL.

TABLE OF CONTENTS

TEXT

TEXT

1. INTRODUCTION

Man has always been concerned with uncovering new knowledge and information. He has found over the years, and especially in the last two centuries, that knowledge and actions based on that knowledge will help him in his efforts whether his cause is that of his own comfort or that of war. Consequently, a great many people have devoted themselves to study and the search for knowledge. As a result, one of the major problems facing the world today is that of storing and making accessible the great wealth of information that has been produced to date and all that will be produced in the future. There has been a virtual explosion in the production of literature in the sciences and the humanities and there is no indication of decline in the ever increasing rate of such production.

Recently, man has become more aware of this problem and many have turned their attentions to search for a solution. Many suggestions have been advanced involving ingenious systems of filing, cross referencing and co-ordinate indexing. The great successes of electronic computers and associated equipment in various types of data processing have given great hope to those working in the field of information storage and retrieval. Varying degrees of success have been achieved in the application of such data processing equipment to the field. Many phases of existing systems have been greatly speeded up through use of the computer. New systems have been proposed that are

BASED PRIMARILY ON THE COMPUTER. THESE AND OTHERS WILL BE PUT TO THE TEST AND IT IS HOPED THAT A PRACTICAL AND WORKABLE AUTOMATED SYSTEM WILL EVOLVE IN THE NEAR FUTURE.

2. LITERATURE SURVEY

2.1 RECOGNITION OF NEED

Much has been written and said on the topic of information storage and retrieval. Most authors are generally concerned that with increased rates of population expansion, increased publishing skills and the increased recognition for research, we are approaching a point where we will not be able to store and retrieve all the information being produced using current methods. Usually this is accompanied by a set of figures bearing testimony to the rate at which scientific and technical works are being turned out. It is felt that unless current techniques of information storage and retrieval are not grossly improved, much of the new literature will be stored in libraries and will become virtually inaccessible to future researchers and designers. In fact, in some circles, it has even been advocated that once a literature search reaches a certain size, it is cheaper to repeat original research than it is to do a thorough literature survey. This may be a trifle extreme because one never knows the true value of a search until it has been performed. Nevertheless, this is the thinking of some top industrial people in the United States. Literature searching causes more delay among their designers and researchers than any other single item.

The work that has been done in this field has been directed toward research more than toward design. It is felt that perhaps the information

RETRIEVAL PROBLEM OF THE DESIGNER IS MORE CRITICAL TODAY THAN THAT OF THE RESEARCHER. MOST RESEARCH TODAY IS CARRIED ON BY TEAMS AND A SINGLE PIECE OF RESEARCH IS USUALLY EXTENDED OVER A LONG PERIOD OF TIME. WHEN A RESEARCH PROJECT IS BEGUN, THERE ARE INFORMATION RETRIEVAL PROBLEMS AND THE RESEARCH TEAM MUST BE SURE THAT THEY HAVE EXAMINED EVERY BIT OF PERTINENT INFORMATION. THOROUGHNESS RATHER THAN TIME IS MOST IMPORTANT AT THIS STAGE. ONCE THE LITERATURE SEARCH HAS BEEN DONE AND THE DIRECTION OF RESEARCH DECIDED UPON, THE TEAM HAS APPROACHED THE 'LEADING EDGE' OF THE FIELD. WHATEVER FUTURE PROGRESS IS MADE WILL BE MADE THROUGH THEIR OWN EFFORTS.

THE DESIGNER, ON THE OTHER HAND, IS IN A SOMEWHAT DIFFERENT POSITION. WHEREAS THE RESEARCHER IS USUALLY A SPECIALIST IN ONE FIELD, THE DESIGNER IS USUALLY EXPECTED TO BE A GENERALIST OF POSSIBLY SEVERAL FIELDS. HE MAY BE CALLED ON TO DESIGN SOMETHING THAT ONLY BORDERS ON HIS OWN FIELD. IN SUCH CASES, ACCESS TO PERTINENT INFORMATION IS CRITICAL. AT THIS STAGE, HIS PROBLEM IS SIMILAR TO THE RESEARCHER. HE MUST EXAMINE THE STATE OF THE ART. TO HIM, THIS WILL MEAN A SURVEY OF GENERALLY ACCEPTED ENGINEERING INFORMATION IN THE FIELD IF HE IS NOT COMPLETELY FAMILIAR WITH IT. HE WILL ALSO EXAMINE OTHER DESIGNS AND PATENTS IN THE SAME LINE. HE IS THEN READY TO PROCEED WITH THE DESIGN STAGE. WHILE DESIGNING, HE MUST HAVE CONSTANT ACCESS TO INFORMATION PERTAINING TO SUCH THINGS AS STANDARDS, ENGINEERING CODES AND EQUIPMENT OF MANY SUPPLIERS. THESE ARE ALL ESSENTIAL TO THE DESIGN PROCESS. THEREFORE, IT IS FELT THAT THE DESIGNER'S INFORMATION PROBLEM IS SOMEWHAT MORE CRITICAL THAN THAT OF THE RESEARCHER AND IN ADDITION BECAUSE OF THE FACT

THAT MOST OF THE PROGRESS IN THE FIELD HAS BEEN WITH THE RESEARCHER IN MIND.

2.2 CLASSICAL LIBRARIANSHIP

Many partial solutions and helpful aids have been proposed by people of various backgrounds. These have been for the most part summarized in a textbook by Becker and Hayes (1). Their book begins with a history of the librarian who has over the years developed techniques and systems that have to this day served the purposes of information storage and retrieval. Their task has been lightened because of the spirit of co-operation and beneficial standardization between all libraries. The primary tools of the librarian are the catalog or index cards. These are used with standardized format and content according to the Dewey Decimal System or the Library of Congress in the United States. A brief discussion is given on these classification systems. These are rather basic and are founded on the attempt to classify all knowledge into ten general categories and then further sub-categories are made as necessary. These techniques are very well known and can be found in any book on library science. It must be noted, however, that these classification systems require that an indexer arbitrarily pick the headings to be used on the index cards. 'See' and 'see also' references help a great deal.

2.3 CO-ORDINATE INDEXING AND THE UNITERM SYSTEM

Next, Becker and Hayes (1) give a concise description of the Uniterm System as put forward by Mooers and Taube. This method of indexing requires an indexer, who would have to be qualified enough to read the book, select a single word (Uniterm) or a set of single words that in his estimation describe the book or paper that is to be indexed. Key words are selected for each article or book in the library and then all key words are kept in a file. In this system, each book or article is given an accession number. These are usually given in the order that the books are read. In other words, the order isn't really important as long as each book has one and only one number. The key word file is comprised of a card for each key word. On this card is listed each accession number of the books from which the key word has been selected. Thus the file has been inverted from having each book with a set of key words to each key word having reference to many books. In a way, this method is very similar to simple subject classification where all the books of the same subject heading are listed together. This technique has been called co-ordinate indexing. It is also covered by P. Logue (2), R. Morse (3), B. E. Holm (4), and R. A. Jacobson (5) with minor variations. Figure (1) shows a simple example of cards in a key word file and how they may be used to locate information.

Now, retrieval in this system consists of matching for key words. For example, a person may desire information on a certain topic that can be described by five or six uniterms or key words. Then the key word file is consulted and the accession numbers found thereon are examined. Whenever we have an accession number that is common to all the cards or at least to several of them, we are reasonably sure that this book concerns the topic of search. The file search and comparison can be done either manually or

BY MACHINE. WHEN A MATCH HAS BEEN MADE, THE SEARCHER IS PROBABLY REFERRED TO AN ABSTRACT FILE. HE CAN THEN READ THE ABSTRACT AND DECIDE WHETHER OR NOT TO READ THE COMPLETE ARTICLE. THE BIGGEST PROBLEM WITH THIS SYSTEM HAS BEEN THAT OF FALSE DROPS. THIS OCCURS WHEN A GROUP OF CERTAIN TERMS ARE CHOSEN SO THAT A MATCH IS MADE ON A CERTAIN ARTICLE THAT HAS NOTHING TO DO AT ALL WITH THE REQUEST FOR INFORMATION. THIS HAPPENS MORE OFTEN THAN MIGHT BE INITIALLY EXPECTED. MOOERS AND TAUBE HAVE ATTEMPTED MANY WAYS OF TRYING TO REDUCE OR ELIMINATE THE NUMBER OF FALSE DROPS. THESE USUALLY CONSIST OF LINKS OR ROLE INDICATORS THAT ARE USED TO GIVE FURTHER QUALIFICATION TO THE KEY WORDS USED. THE ADDITION OF THESE LINKS GREATLY INCREASES THE COMPLEXITY OF THE SYSTEM BUT DOES HELP WHEN THERE ARE A GREAT MANY FALSE DROPS. THE NUMBER OF FALSE DROPS DEPENDS TO A GREAT EXTENT ON THE INITIAL INDEXING. IF, OF COURSE, EVERY BOOK HAD THE EXACT NUMBER OF KEY WORDS REQUIRED TO DESCRIBE IT UNIQUELY, WE MIGHT HAVE A SYSTEM WITH NO FALSE DROPS. THEREFORE, WE SEE THAT THE QUALITY OF THE INDEXING IS OF GREAT IMPORTANCE TO THE END RESULTS. MUCH CONTROVERSY HAS RESULTED OVER THE COST AND DEPTH OF THE INDEXING AND THE TIME AVAILABLE FOR THE SEARCH. A COMPLETE SYSTEM WOULD HAVE HIGHLY TRAINED AND COMPETENT PEOPLE DOING THE INDEXING IN THEIR PARTICULAR FIELDS AND HIGHLY TRAINED PEOPLE FAMILIAR WITH THE WHOLE SYSTEM DOING THE SEARCH. THIS USUALLY TURNS OUT TO BE QUITE EXPENSIVE. IT IS FELT THAT ANY AID IN THIS RESPECT THAT CAN BE OBTAINED BY USING COMPUTER TECHNIQUES WILL CERTAINLY HELP. THERE IS ALSO CONTROVERSY WITH RESPECT TO THE TIME OF SEARCH. SOME PEOPLE FEEL THAT A RESEARCHER MAY NOT BE PARTICULARLY RUSHED TO FIND THE information requested. This is the attitude taken by Bar-Hillel (6). However, a design engineer working to a schedule may be very concerned with time. He may be willing to accept more false drops as long as the references are in his hands sooner. He may feel that he can weed out the false drops by quick examination of the abstracts and thereby save himself considerable waiting time.

References (2), (3) and (4) are reports from such companies as the Monsanto Chemical Co., Philip Morris, Inc., and the DuPont Co., on systems that they have actually put into practice with considerable success. Groups such as the American Institute of Chemical Engineers and the Engineers Joint Council have favoured the co-ordinate indexing technique and each has prepared a thesaurus to be used in conjunction with this system. The AIChE Thesaurus is made up primarily from the field of chemical engineering. The EJC have tried to expand on this and create a Thesaurus of Engineering Terminology. A thesaurus is of considerable aid when searching for information with respect to the problem of synonyms. It is characteristic, however that nowhere in the literature can one find detailed information on computer systems and/or languages used.

2.4 THE WESTERN RESERVE UNIVERSITY CONCEPT SYSTEM

Becker and Hayes (1) also discuss the Western Reserve University System. Its chief advocates are J.W. Perry and A. Kent who explain the system in great detail in references (7) and (8). Perry and Kent seem to believe in machine <u>aided</u> literature searching. It appears to be their feeling

THAT SOME PARTS OF THE INFORMATION STORAGE AND RETRIEVAL PROCESS CANNOT BE REPLACED BY MACHINE. IN PARTICULAR, THEY SEEM TO FEEL THAT INDEXING AND ALSO ABSTRACTING MUST ALWAYS BE DONE BY PEOPLE.

MUCH OF THEIR WORK IN RETRIEVAL IS BASED ON CLASS DEFINITION. USED IN CONJUNCTION WITH THIS IS BOOLEAN ALGEBRA WHICH IS PRIMARILY AN ALGEBRA OF LOGIC. FOR EXAMPLE, SUPPOSE THAT A, B, AND C REPRESENT WORDS OR CONCEPTS. THEN A SEARCH QUESTION USING (A. B–C) WOULD SEARCH FOR AN ARTICLE CONTAINING BOTH A AND B BUT NOT C. A COMPLETE ALGEBRA THAT CAN COMBINE WORDS AND/OR CONCEPTS IN ANY LOGICAL FASHION HAS BEEN ESTABLISHED.

PERRY AND KENT (7) GO ON TO DISCUSS OPERATION CRITERIA ON INFORMATION SYSTEMS THAT CAN BE USED TO COMPARE DIFFERENT SYSTEMS. THIS IS DONE MAINLY BY COMPARING RATIOS. IF N IS THE TOTAL NUMBER OF DOCUMENTS EMBRACED BY A CERTAIN SYSTEM AND M· IS THE NUMBER OF DOCUMENTS WHICH USE OF THE SYSTEM INDICATES TO BE OF POSSIBLE PERTINENT INTEREST AND W IS THE NUMBER OF DOCUMENTS THAT ARE FOUND TO BE OF ACTUAL PERTINENT INTEREST BY PERSONAL INSPECTION OF THE M SELECTED DOCUMENTS, THEN WE CAN DEFINE FRACTION SUCH AS M/N WHICH IS THE FRACTION OF TOTAL DOCUMENTS TO WHICH ATTENTION IS DIRECTED. THIS FRACTION IS ALSO CALLED THE RESOLUTION FACTOR. WE HAVE ALSO THE FRACTION (N–M)/N, WHICH IS THE FRACTION OF TOTAL DOCUMENTS FROM WHICH ATTENTION IS DIVERTED AND IS JUSTLY CALLED THE ELIMINATION FACTOR. BY COMPARING VARIOUS RATIOS BETWEEN TWO SYSTEMS MAKING AN IDENTICAL SEARCH, WE CAN EVALUATE ONE SYSTEM AGAINST THE OTHER. THIS IS RATHER THEORETICAL AS YET BECAUSE THE AUTHOR KNOWS OF NO CASE WHERE TWO OR MORE SYSTEMS HAVE BEEN USED ON A GIVEN SET OF DOCUMENTS

FOR COMPARISON PURPOSES. HOWEVER, IT IS THOUGHT THAT IN THE FUTURE WHEN SYSTEMS ARE BEYOND THE EXPERIMENTAL STAGES OF TODAY, THAT THESE CRITERIA WILL PROBABLY BECOME OF USE.

PERRY AND KENT CLAIM THAT THERE IS NO 'NATURAL' LOGIC IN THE ENGLISH LANGUAGE OR IN ANY OTHER NATURAL LANGUAGE. EACH LANGUAGE IS BASED ON CERTAIN POSTULATES AS TO WHICH TYPES OF OBSERVATIONS AND CORRELATIONS ARE IMPORTANT. SISTER LANGUAGES WOULD TEND TO CONTAIN THE SAME BASIC POSTULATES BUT WOULD BE COMPLETELY DIFFERENT FROM OTHERS. THE BASIC POSTULATES OF EACH LANGUAGE ARE RATHER ARBITRARY AS ARE MANY OF THE CONVENTIONS IN SENTENCE STRUCTURE. AN ATTEMPT TO USE A NATURAL LANGUAGE WOULD, AT BEST, LEAD TO FORMALIZATION OF THE BASIC POSTULATES OF THE LANGUAGE. THE BASIC THEORETICAL STRUCTURE OF CHEMISTRY, PHYSICS OR ANY OTHER BRANCH OF SCIENCE WOULD NEITHER BE REVEALED NOT TAKEN INTO ACCOUNT IN SUCH A STUDY OF A NATURAL LANGUAGE.

IN THEIR ESTIMATION, IT IS NECESSARY TO HAVE A COMPLETELY DIFFERENT MACHINE LANGUAGE BASED ON AN ALGEBRA OF LOGIC. REFERENCES (1) AND (7) GIVE A GOOD DESCRIPTION OF THE SYSTEM BUT IT IS THOUGHT THAT A BRIEF DESCRIPTION OF THE LANGUAGE SHOULD BE INCLUDED HERE FOR COMPLETENESS. THE LANGUAGE ADOPTED BY THE WESTERN RESERVE UNIVERSITY GROUP IS ONE BASED ON CONCEPTS, SO THAT IN A SEARCH, RETRIEVAL IS BASED ON MATCHING OF CONCEPTS RATHER THAN WORDS. THEY STATE THAT A SYSTEM CANNOT BE BASED MERELY ON WORDS BECAUSE WORDS ARE ONLY SYMBOLS OF CONCEPTS. THEY DO RECOGNIZE, HOWEVER, DIFFERENCES IN CONCEPTUAL-IZATION IN DIFFERENT LANGUAGES AND EVEN GEOGRAPHICALLY WITHIN THE SAME LANG-UAGE. THEIR LANGUAGE IS BASED ON A SET OF CAREFULLY SELECTED AND DESIGNED generic concepts with some allowance for class inclusion. For example, the word thermometer is in the class of measuring device.

The letters (Group 1) B, C, D, F, G, H, L, M, N, P, R, S, T are used in groups of three at a time with a space in between, to define a concept. For example, B-CT means bacteria, B-LD means building or house, C-PT means to compute, count or calculate and M-CH means machine, device or tool. The letters (Group 2) A, E, I, O, U, W, X, Y, Z are used in conjunction with the foregoing to signify an analytical relationship. For example, A means class inclusion, U means a process related to a device, and W means a thing to which a concept is submitted. Various minor concepts are grouped together to form a major concept. For example, MUSRMACHTWMP is formed from the three concepts MUSR, MACH, and TWMP. When taken together, it means machine for measuring temperature. That is, a thermometer. However, it could also mean pyrometer, so we see that a concept has been defined rather than just an object. The system becomes even more complex because to completely define some concepts, role indicators must be used in this system as well.

The basic system may be outlined as follows–

I Preparation Phase

1. Indexing Step– a person must scan the articles and decide what is pertinent
2. Editing Step– another person familiar with the overall system is needed to take account of important relationships between substances, processes, etc., which are designated by index

ENTRIES PREVIOUSLY ESTABLISHED.

3. ENCODING STEP– THE KEY WORDS (CONCEPTS) AND ABSTRACTS MUST BE ENCODED AS EXPLAINED ABOVE.

4. RECORDING STEP– THE SYMBOLIZED DATA IS PUT ON COMPUTER CARDS OR TAPE FOR FURTHER PROCESSING. AN EXAMPLE OF AN ENCODED ABSTRACT IS SHOWN IN FIGURE (2) AND FIGURE (3) IS AN EXPLANATION OF THE TERMS.

II SEARCHING PHASE

1. STATEMENT OF SCOPE OF SEARCH– THE INFORMATION REQUEST MUST BE ACCURATELY STATED.

2. INTERPRETATION OF SCOPE OF SEARCH– A PERSON WITH KNOWLEDGE OF THE WHOLE SYSTEM MUST EXAMINE THE SEARCH QUESTIONS TO SEE THAT IT IS PROPERLY STATED. IT MUST BE PUT IN BOOLEAN ALGEBRA FORM AND ENCODED.

3. MACHINE SEARCHING OPERATION– THE MACHINE DOES ITS OWN SCANNING AND MATCHING HERE.

4. PROVISION OF ACTUAL RECORDS– THIS CAN VARY ACCORDING TO COMPLEXITY REQUIRED. IT CAN GO FROM AN ABSTRACT FILE AND A BOOKSHELF, TO A MICROFILM SYSTEM FOR PRINTING OF COPIES. THIS SECTION IS RATHER SECONDARY TO THE ACTUAL SYSTEM.

SUPPOSE WE HAVE AN INFORMATION REQUEST ON THE USE OF ALLOYS FOR TREATING GRAY IRON TO IMPROVE ITS PROPERTIES. THE REQUEST MAY BE SYMBOLIZED IN THE FOLLOWING MANNER–

First Search Requirement		$A . B = A'$
A	Use of (By means of)	KQJ
B	Alloy	LALLØ1
Second Search Requirement		$C . D = B'$
C	Material Processed	KEJ
D	Gray Iron	RARNØ4
Third Search Requirement		$E . D = C'$
E	Properties given for	KOV
D	Gray Iron	RARNØ4
Fourth Search Requirement		$F . G = D'$
F	Properties	KWV
G	Improvement	KQV

Now the search requirement in total is to find the four individual search requirements in any order in any sentence in an abstract. The request may now be symbolized as–

$$\langle \# \text{"} (A' . B' . C' . D') \text{"} \# \rangle$$

The symbols $\langle \ \rangle$ denote that the machine must find #, ", the logical product $(A' . B' . C' . D')$ and then " and # in that order. The regular parentheses show that the order of A', B', C' or D' is not important. Such is an example of a generic question interpreted and encoded in the machine language.

Becker and Hayes (1) say that this method of producing terms, since it is based on a highly organized concept of the relationship among terms, allows for recovery in response to questions phrased within this structure of terms. Thus generic questions within this framework will be handled without difficulty. However, it must also be realized that relationships outside the structure, defined by the chosen semantic relations will be as hard to handle in the Western Reserve System as they are in any other system and perhaps more so.

It seems to the author that Perry and Kent have achieved a certain degree of rigour but have also succeeded in making the system more complex. They claim that words are no good and yet they must search for concepts in exactly the same way any other system would search for words. It is true that they have eliminated many problems such as synonymns, word roots, etc., but they have done this only at the expense of having to encode everything. There has been no provision made for either automatic encoding or abstracting. They have been successful in eliminating most of the bottlenecks in the machine searching section of their program but have also been successful in transferring those bottlenecks to other places in the program. In addition, there has been no allowance for the fact that usefulness of an article may change in time. Their system would involve re-abstracting, etc., which would, of course, be very time consuming. This points up the desirability of machine abstracting if it is at all possible. A machine abstract could be prepared in very short order and the change in value over time might be allowed for by simply a change in input parameters.

It should be noted that Perry and Kent do not give much indication as to the success they have had with their system with respect to the computer. They do hint that they have had success with simple programs but neglect to indicate types of computer programs used.

Reference (8) by Perry and Kent is a book which tries to set up a mathematical model that will examine and grade various information storage and retrieval systems with respect to time and efficiency. This contains a great deal of additional work on theoretical operational criteria such as pertinent factors, recall factors and omission factors as previously explained. They do go on to present some cost analysis figures for standard operating systems such as the Dewey Decimal System and the United States Patent Office Pigeon-Hole System. This is rather interesting to the whole field of information storage and retrieval, but for the most part, is quite specialized and not of much concern at the present time to those interested in automatic systems. In such an early stage of development, we must concern ourselves first with whether or not a system will work satisfactorily at all. When many choices are available, we can begin to think about efficiencies. In addition, some books make reference to Russian systems and indicate that it is supposed that the Russians are more advanced in this field than we are. Therefore, in the opinion of many citizens and politicians, we must keep up, and cost and efficiency are of secondary importance.

2.5 TOOLS AND EQUIPMENT

Becker and Hayes (1) present a brief but thorough review of the equipment used in the field of information retrieval today. They begin with a discussion of the modern forms of printing which includes letterpress and offset. Another section is devoted to office duplication and photocopy. The foregoing is generally encompassed by the field but would not normally be thought of as prime tools. A discussion is presented on electronic data processing machines. This includes an explanation of punch card and punch tape codes along with the associated key punch machines. Reference is made to a print reading machine developed by the Rome Air Development Centre (Rome, N.Y., U.S.A.) which utilizes television-type video impulses to recognize different letters. It is equipped to read at the rate of approximately two full single spaced pages per minute and punch the information on tape. This, of course, is an experimental machine but eventually this type of machine will overcome one of the major drawbacks of automated systems. This is the problem of rapid input.

Also included, is a discussion of special purpose devices using principles of magnetic recording via magnetic drums, disks, tape and cards. The general theory behind each is given and this provides a grounding in any of the basic systems in use today.

An excellent discussion of microphotography is presented next. Microfilm and many of the various viewers are encountered. Many problems such as preparation and accessability have been overcome by systems such as the Rapid Selector and Kodak's Minicard system. These systems generally contain both index and document file information on the same card. A card in the Minicard system is approximately half the size of a postage stamp. These systems are advantageous in that they are quite compact and undoubtedly find many specialized applications.

In summary, Becker and Hayes have presented an excellent general review of equipment used in the field today. More detailed information can easily be obtained from suppliers of the equipment but this is usually quite cluttered up with impressive superlatives.

2.6 AUTO-ABSTRACTING

Becker and Hayes (1) then present a brief of the auto-abstracting process originated by H. P. Luhn (9, 10). Luhn seems to be more concerned with the time aspect of literature searching than anyone else. He has been trying to automate some parts of the total system that have been avoided by others, such as the topics of automatic abstracting and indexing. His proposals are based somewhat on classical co-ordinate indexing as presented in section (2.3) of this report. Associated with this, he wishes to create a dictionary of words, ideas and notions similar to creating families of words in a thesaurus. It is felt that in building this thesaurus, a finite number of ideas would be achieved just as there are approximately one thousand word categories in Roget's Thesaurus. He proposes that all documents be edited and transcribed into machine readable form with possibly nouns and certain verbs identified. Then a card index of all transcribed sentences would be set up. The cards would be grouped in 'notional families' and experts would have to weight the notions. For example, it is possible that nouns are enough to represent a notion. These can be grouped into families as in a thesaurus. As

NOTED LATER IT IS FELT THAT THESE PRINCIPLES ARE NON-FORMULATED IDEAS BY THE AUTHOR AND POSSIBLY THIS IS A REASON WHY NO FURTHER INFORMATION OR EXAMPLES ARE OFFERED BY LUHN.

HE ESTIMATED THAT THE NUMBER OF NOTIONAL CATEGORIES WOULD BE LESS THAN ONE THOUSAND. THEN, NEW DOCUMENTS CAN BE ENCODED WITH THIS DICTIONARY OF NOTIONS BECAUSE IT WILL BE POSSIBLE TO RECOGNIZE THE NOTIONS IN THE AUTHOR'S USE OF SENTENCES, PARAGRAPHS, ETC. THEN, IT IS SUPPOSED THAT THE MOST FREQUENTLY USED NOTIONS ARE THE MOST IMPORTANT. WE WOULD THEN END UP WITH A 'NOTIONAL ABSTRACT' THAT COULD BE STORED ON TAPE OR CARDS FOR FUTURE SEARCHES. THE FOREGOING IS THE SYSTEM ORIGINALLY PROPOSED IN REFERENCE (9) AND LUHN ADMITS THAT THIS IS AN UNTESTED PROPOSAL TO BE DEVELOPED IN THE FUTURE. MANY QUESTIONS ARE STILL TO BE ANSWERED AND MANY OF THE ASSUMPTIONS ARE YET TO BE VERIFIED. ONE OF THE PRIMARY ADVANTAGES THAT HE HAS IN MIND IS THAT THE ORIGINAL IDEAS OR NOTIONS OF THE AUTHOR WOULD NOT BE DESTROYED BY THE INDEXER WHO MUST INTERPRET THE ARTICLE ACCORDING TO HIS OWN BACKGROUND. IT IS FELT THAT IN THIS ARTICLE, LUHN IS DREAMING OF AN OVERALL SYSTEM BASED ON THESE IDEAS. HOWEVER, IT IS ALSO FELT THAT THE IDEAS PRESENTED MUST COME DOWN SEVERAL LEVELS OF SOPHISTICATION IN ORDER TO BE TRIED.

IN A LATER ARTICLE (10), LUHN DOES COME DOWN IN SOPHISTICATION TO TRY SOME OF HIS IDEAS WITH PRONOUNCED SUCCESS. HE MAKES THE POINT THAT WHATEVER TIME THAT AN AUTO-ABSTRACT CAN SAVE IS TIME SAVED FOR A HIGHLY QUALIFIED INDEXER OR ABSTRACTER WHO IN MOST SYSTEMS MUST HAVE ABILITY COMPARABLE TO THE AUTHOR OF THE ARTICLE BEING ABSTRACTED. HE AGAIN POINTS OUT THAT AN abstracter is influenced by his background, attitude and disposition whereas a machine is impartial according to its instructions. Luhn admits that one of the great drawbacks to automatic abstracting today is the fact that an article must be transcribed into machine readable form. This means usually that the article must be punched onto cards. However, it is noted that many publishing houses in the United States today have paper tape machines attached to their automatic type-stters. Thus as the type is being set for publication of a book, a computer tape is being prepared at the same time which can be easily converted into cards if desired. In addition, there is some work going on today on character recognition devices that will be developed into machines for reading print. Therefore, it is hoped that this one large drawback will be overcome sometime in the near future.

Luhn's basic idea behind his auto-abstracting program is that of word significance. It is his feeling that a writer normally repeats certain words in advancing his arguments. A writer, conscious of his writing, will try to use synonyms to reduce the boredom in his article, but usually there aren't too many alternate choices so he is eventually forced back into repeating himself. Luhn also feels that the more often certain words are found in each others' company, the more significant they are. For example, it is his feeling that if two key words are found side by side, they hold more significance than if they were found separated.

Common words can be eliminated through the use of a common word list. There are some words that may be very common to some field and yet

THEY ARE NOT FOUND IN THE COMMON WORD LIST. FOR EXAMPLE, WORDS LIKE MACHINE AND POWER ARE COMMON IN ENGINEERING LITERATURE. WITH THIS IN MIND, LUHN THEN SUGGESTS THAT THERE BE A CUT-OFF OF VERY HIGH AND LOW FREQUENCY WORDS AND ONLY THE WORDS WITH FREQUENCIES BETWEEN THESE TWO LIMITS BE USED AS SIGNIFICANT KEY WORDS. HE DOES SAY, HOWEVER, THAT THESE CUT-OFF POINTS WOULD HAVE TO BE DETERMINED BY EXPERIENCE AND HE ALSO NEGLECTS TO SAY IF HE USED THIS PRINCIPLE IN HIS OWN PROGRAM.

ONCE HAVING ESTABLISHED WORD SIGNIFICANCE, HE GOES ON TO ESTABLISH SENTENCE SIGNIFICANCE. HE CONSIDERS THAT THE TWO MOST IMPORTANT FACTORS ARE KEY WORD OCCURANCE IN THE SENTENCE AND KEY WORD PROXIMITY IN THE SENTENCE. PSYCHOLOGICAL RESULTS (NOT REFERENCED) LEAD HIM TO THINK THAT SIGNIFICANT WORDS PLACED MOST CLOSELY TOGETHER ARE IMPORTANT FOR TRANSMISSION OF IDEAS. THEREFORE, HE DIVIDES EACH SENTENCE INTO BRACKETS AND GIVES EACH BRACKET A WEIGHT ACCORDING TO THE NUMBER AND PROXIMITY OF KEY WORDS. THEN HE WEIGHTS THE WHOLE SENTENCE ACCORDING TO THE BRACKET WEIGHTS. HE SUGGESTS THAT THE PRINT-OUT CAN BE THOSE SENTENCES ABOVE A CERTAIN WEIGHT OR A CERTAIN NUMBER OF THE MOST IMPORTANT SENTENCES.

IT IS HIS FEELING THAT AUTO-ABSTRACTS, WHEN DEVELOPED A LITTLE FURTHER WILL HAVE A HIGHER DEGREE OF STABILITY THAN THOSE PREPARED BY HUMANS. HE FEELS THAT THE AUTHOR'S OWN WORDS SHOULD NOT BE ALTERED BY HUMANS. THE ABSTRACT WILL BECOME A SORT OF 'INDICATIVE' ABSTRACT. WHEN PEOPLE BECOME MORE FAMILIAR WITH THEM THEY WILL BECOME MORE ACCUSTOMED TO THE FACT THAT THE SENTENCES SELECTED FOR THE AUTO-ABSTRACT INDICATE MOST FREQUENTLY USED WORDS AND WILL not expect that the sentences selected form a complete resume of the article. They will see which are the key words and expect more of the same in the same general context as the sentences presented.

2.7 KWIC, KWOC, WADEX

Becker and Hayes (1) discuss briefly, machine prepared indexes as well as abstracts. The KWIC (Key Word in Context) is perhaps the first and most common. Usually this type of indexing concerns itself primarily with the title of the article because it is felt that the title should contain an indication as to what is in the article. Occasionally, one finds articles with short and rather meaningless titles. It has been widely advocated by people responsible for KWIC, that these titles be changed with or without the author's consent, to something more meaningful so that their indexing system will operate more satisfactorily.

The idea of the KWIC index is to list each book or article in its system, once for every key word used in its title. The index first examines all the titles and picks out all the key words. It is supposed that this is done simply by eliminating common words. The key words are arranged in alphabetical order, and for each key word in the title, the title is printed in context, in a vertically aligned position. The key word is in the middle of the line. An example of a KWIC index is shown in figure (4). A person can then simply scan down the column of key words in the centre of the page until he finds the ones he wishes. If the title indicates that the book is required, a reference location is also supplied. Illustrations of KWIC indexes can be found in references (1) and (11).

The KWOC system (Key Word Out of Context) is very similar to the KWIC system. The arranging in alphabetical order according to the key words is identical. The main difference is that the key word is printed out at the left hand side of the page and then the whole title follows it. This essentially then is the same index with a different format that makes it slightly easier to read.

The WADEX system (Word and Author Index) is similar to the KWOC system but has some major differences. Titles are printed fully including authors' names. Each title begins at the left hand side and occupies as many lines as are required. Key words and authors' names are arranged together in alphabetical order and printed out of context to the left of the title. WADEX is arranged in two columns with pagination at the bottom and a dictionary entry at the top of the page. This system is quite easy to read and includes authors' names. It is felt that many people remember articles or books by author rather than by title. Figure (5) is an illustration of a WADEX index. Reference (11) contains a verbal description of the steps taken in preparing a WADEX index and this is aided by a flow chart. No indication is given as to what types of languages are used but it is assumed that these indexes are prepared by computer experts that would use basic machine languages rather than some prepared language such as IBM's FORTRAN or COBOL.

These computer prepared indexes are not expected to serve the purpose of a complete information storage and retrieval system. However, they are examples of the efforts for time saving going on in the field. It might be pointed out that these indexes are very well suited for browsing for the searcher who is not completely sure of what he requires.

2.8 SYSTEMS THEORY

Becker and Hayes (1) and B.C. Vickery (12) make some attempts at the formalization of information system theory. Reference (1) breaks an information storage and retrieval system into separate functions for the user, operator, designer and supplier of equipment. A brief description of the responsibilities of each group is presented. Vickery (12) makes reference to some of the established systems (Luhn, Perry and Kent) but it seems that he gets very involved with theories concerning structural modes, file organizations and search procedures. A discussion of operational criteria of systems is also presented.

2.9 A CRITICISM

Many authors are quick to criticize the few established systems because they are in such a crude stage of development. Perhaps one of the best criticisms of the field of information storage and retrieval is given by Y. Bar-Hillel (6). He gives some praise to developers of such systems as the Minicard system for having dared to step over the accepted four steps of accession numbers, citations, abstracts and duplicate copies. These systems make use of microfilm mounted in computer cards. Then if a card is selected because of key words and it is sorted out by a machine sorter, the searcher has the document or possibly the abstract on microfilm immediately. This can then be read through a microfilm viewer. There is, of course, much effort required in preparation of these cards but many companies in the United States are producing cards and microfilm mounters that greatly simplify the task. He feels that this system may develop into something that will eliminate the reference system. It seems that he fails to realize that the Minicard system is not a system at all. Rather it is merely a tool to be incorporated into some other system.

He is very critical of auto-indexing and auto-abstracting systems. He feels that sets of key words obtained by machine will not be of as good quality as those obtained by an experienced indexer because of problems such as synonyms, plurals (word bases), etc. These would not be any better than a crude uniterm system in which he places no faith at all. He says that an auto-abstract can do·no better than the average of human ones and that an author's extract would be the best. It seems that he doesn't put any value on time required for human indexing and also fails to recognize that the practice of authors indexing their own works is not yet very common. In the future, this would certainly be a help however.

He claims that the only steps of the literature search process which are amenable to performance by a digital computer are those steps which follow the assignment of the Boolean function over the topic terms, up to and including the printing of the reference list.

2.10 COMMENTS OF THE AUTHOR

The field of automated information storage and retrieval is in its initial stages of development. The problems encountered can not and will not be solved overnight. There will be a period of evolution in which theories and systems will grow and perish according to their merits. Criticisms such as Bar-Hillel's (6), although they do not contribute anything directly, will serve as inducements for the design of better systems.

It has been seen in this literature survey that many people have been working in this field and have had success in finding ways to aid the problems of information systems. The Western Reserve System is perhaps the most formal and most complex of all. The basic fault here is that it has been made virtually impossible to have a completely automated system. The encoding and abstracting has to be done by highly trained personnel. They have succeeded in solving some of the problems of concepts in language but in so doing have developed a partially automated system that is quite complex and is not felt to be suitable from a designer's point of view because of his time limitations.

The uniterm system and auto-abstracting system have progressed along somewhat similar lines. They have tried to use natural language and have varying degrees of complexity. Usually, a search can be done very quickly but only by getting many false drops. To eliminate false drops involves time and more complicated systems. From the designer's point of view, he may be perfectly willing to take many false drops to save time. It is probable that he could eliminate most of the false drops by the title alone and many more by a quick examination of the abstract. From the point

OF VIEW OF THE DESIGNER, SYSTEMS THAT WORK BETTER WILL BE DEVELOPED IN THE FUTURE THAT WILL GIVE HIM FAST AND ACCURATE INFORMATION BUT TODAY HE WOULD LIKE A COMPLETELY AUTOMATED SYSTEM TO GIVE QUICK INFORMATION EVEN THOUGH IT IS NOT COMPLETELY RELEVANT.

It should be noted that it is the author's feeling that the literature in the field of information storage and retrieval is in an unsatisfactory state. One reason for this is that people that have had some success using a computer oriented system have not published details concerning types of computers used, computer language used and programs used. The second reason is that the literature promises and implies much more than it delivers. It is possible that this is characteristic of a new field. It seems that many of the aids designed for the field are written up in the literature as being complete solutions to all the problems.

It is also characteristic of the work that has gone on in the field that a complete system has not been attempted. Various portions of systems have been tried and found successful. To the author's knowledge, no-one has attempted to incorporate many of the successful features and ideas into an overall system design. It is his hope to establish plans for an overall information storage and retrieval system that makes full use of digital computers and that as much of the detail as possible will be provided.

3. DESIGN

3.1 SCOPE OF DESIGN

The two basic faults with the current literature found by the author have been stated as the omission of computer programs used and the failure of anyone to attempt an overall automated system design. It is with these two shortcomings in mind that this design has been attempted.

The basic system proposed will be outlined. Due to time limitations, all of the detail in the system could not be included. The key aspects of the system will be detailed and the items neglected will be described with an indication given as to how they could be completed.

It should be noted that the proposed system is not meant to be a complete solution of all the problems in the field of information storage and retrieval. The system would have to be greatly complicated to solve many of the problems. However, it is hoped that the system would serve as a basic foundation for further advances in this field. In addition, the system was designed with particularily the needs of the designer in mind. It is felt that designers, more than any other group, are in need of rapid information retrieval. In fact, it is felt that they would be very willing to accept many false drops as long as the response of the system was quick.

The system designed is meant to be a completely automated system. New information coming into the system is analyzed for key words and key sentences that form an abstract. Various files are kept of key words, abstracts etc. When there is a request for information from a user, the key words in his request are identified and by matching for key words, the computer can pick the pertinent references out of its library. The computer used in this case was an IBM 7040, recently installed in the Engineering Building at McMaster University. Since the memory of the computer is finite and only a certain amount of information can be stored, it was necessary to break the design down into several stages as will be explained in the following sections.

3.2 DESIGN OUTLINE

Before analyzing any part of the design in detail, it would perhaps be advantageous to examine the overall design concept with reference to figure (5). Any new information coming into the system must first be converted into machine readable form (A). At present, this means simply punching the information manually onto either computer cards or tape. It is unfortunate that this is such a time consuming process but as mentioned in the literature survey, there is considerable work going on in the United States on print reading machines. Experimental machines have been developed and it is hoped that the process of preparing material for the computer will be greatly speeded up in the near future. In addition, as explained previously, a computer tape can now be made from an automatic type-setting machine for an article or book being put into print.

The machine readable form of the information in this case is a series of IBM computer cards. This becomes the input data to the first computer

PROGRAM WHICH IS PHASE I OF THE AUTO-ABSTRACTING PROCESS (B). THE KEY WORDS IN THE ARTICLE ARE SELECTED ON THE BASIS OF WORD FREQUENCY AND THE RESULTS ARE PRINTED OUT AND ALSO PUNCHED OUT ONTO CARDS. ALONG WITH THE WORD ITSELF, THE WORD LENGTH AND FREQUENCY OF USE ARE ALSO OUTPUTS OF PHASE I.

THE OUTPUT CARDS OF PHASE I BECOME PART OF THE INPUT DATA OF PHASE II ALONG WITH THE FULL ARTICLES AS USED IN PHASE I. PHASE II (C) EXAMINES EACH SENTENCE OF THE ARTICLE AND EVALUATES IT ACCORDING TO THE WORDS USED IN IT. THE KEY SENTENCES WHICH THE PROGRAM SELECTS THEN BECOME THE AUTO-ABSTRACT OR AUTO-EXTRACT AS IT IS SOMETIMES CALLED. THE OUTPUT OF PHASE II IS DEPENDENT ON THE TYPE OF STORAGE USED FOR THE ABSTRACTS. IT CAN BE EITHER A PRINT OUT OR A PUNCH OUT ON CARDS. IF THE SYSTEM, INCLUDING THE STORAGE SECTION, WERE COMPLETELY AUTOMATED AT SOME TIME, AN OUTPUT FROM PHASE II WOULD NOT BE NECESSARY. THE ABSTRACT WOULD SIMPLY BE STORED AWAY EITHER IN THE MEMORY OF THE MACHINE OR ON VIDEOTAPE DEPENDING ON THE SYSTEM OF STORAGE IN USE.

THEN, BOTH THE ABSTRACT AND THE FULL ARTICLE ARE STORED ACCORDING TO A REFERENCE NUMBER GIVEN THEM. THE ACTUAL MEDIA OF STORAGE IS SOMEWHAT ARBITRARY. THE METHOD SELECTED WILL DEPEND ON SUCH FACTORS AS THE NUMBER OF ARTICLES AND ABSTRACTS ON FILE AND THE NUMBER OF SEARCHES. THIS IS A FIELD IN WHICH THERE HAS BEEN MUCH DEVELOPMENT OF EQUIPMENT. WE HAVE MICROFILM EQUIPMENT THAT WILL STORE ABSTRACTS OR SHORT ARTICLES ON MICROFILM MOUNTED IN A COMPUTER CARD THAT CAN BE CODED WITH THE ACCESSION NUMBER FOR MACHINE SORTING PURPOSES. ALSO AVAILABLE ARE SMALL CARDS ABOUT HALF THE SIZE OF A POSTAGE STAMP MADE UP OF MICROFILM. ON THE MICROFILM ARE LIGHT AND DARK SPOTS that encode information for sorting. Usually associated with this type of system is a series of necessary equipment such as cameras, film mounters, sorters and viewers. There is also equipment that uses a magnetic tape system similar to the videotape system used in television. This system, too, would require a good deal of specialized equipment associated with it.

It is the author's feeling that when the information system is being set up, the full articles should be stored as original copies and the abstracts could be kept in an abstract file of cards. The abstract could be typed on the face of the card and stored by the accession number given to the article. Then as the library grew, it would be possible to switch to a system employing microfilm that would make sorting and retrieval much easier. It would be up to the judgement of the operators of the system as to when the microfilm or videotape additions would be needed.

It is also possible that sometime in the future we would have information storage and retrieval centres that would contain the original documents and the necessary computers. Then outside areas could ask for and receive information by means of a teletype system and receive copies of the information, if necessary, over the wire. This, of course, would eliminate a great deal of duplication of effort. Information centres in the United States and Russia that are already existant could possibly grow into such district centres.

So far, only the abstracting process has been dealt with. From Phase I (B), the key words, along with the accession number assigned to them must be stored in a file. The file is inverted so that instead of having a group of key words for each accession number, there is rather a group o f accession numbers for each key word. The file of key words has been arranged for searching procedures to be discussed later. This, essentially, is the completion of work on new information coming into the system.

However, as a by-product of the system, it would be very convenient to prepare a computer index (G). This is not essential to the system and could be held as a sub-program to be run only when required. The three most common indexes in existance are KWIC (Key Word in Context), KWOC (Key Word Out of Context), and WADEX (Word and Author Index). These have been covered in section (2.7) of this report. The type of index here would be one very similar to WADEX which lists each article once for the author's name and once for every key word in the title. It is felt that not all titles contain all the pertinent key words and since the key words have already been determined, the article should also be listed for each of the key words found in the text of the article.

The process of information retrieval must begin with a request for a certain type of information. In this system, the request may be of two forms. First of all, the user of the system may just write down a list of key words from which the search would be conducted (H). This would likely be done with the aid of the system operator. The other form of request would be to have the user compose a request in two or three paragraphs that would completely express his requirements (K). This would be put in machine readable form as done in part (A) and then would be subjected to a Phase I type operation (L) that would pick out the key words from which the search would be conducted.

By one of these two methods, the search key words have been found. These would be prepared on cards either automatically from part (L) or manually from part (H). Now, there must be a program to search through the key word file for matches between the search key words and the key words on file. By methods discussed in section (3.7), the accession numbers to all the pertinent articles are determined. This process can be greatly aided by the addition of a thesaurus. Thus a search key word may be looked up in this thesaurus to give a number of synonyms to include in the search. Such methods usually would be very helpful in increasing the number of references found by the system. At any rate, the accession numbers of supposedly pertinent articles have been obtained and these of course stand for references which can be looked up in an accession number file (J). This can be done either by machine or manually. The level of automation used at this stage is greatly dependent on the amount used in the storage portion of the process. If a manual file was used in the storage portion, then a manual look up would be indicated here. If the storage has been done on microfilm or videotape and prepared for machine sorting, then the computer can do the looking up instead.

Now the user of the system has his references and/or copies of the abstracts and the full articles. By scanning titles and abstracts he can easily eliminate the greatest portion of the false drops incurred.

This has been a discussion of the overall system and in the next few subsections, the details of some of the more important aspects of the system will be discussed.

3.3 KEY WORD DETAIL (PHASE I)

The purpose of Phase I of this system is to pick out key words in a given article. It operates on the theory that the more important words in an article, save for common words, are used more often than others. This is felt to be due to an author's tendency to repeat himself when bringing out an important point. Therefore, the program must remember each word in an book or article and how often it is used. All the computer programming in this system has been done in the Fortran IV language. A copy of the Phase I part of this system is presented in Appendix 1.

The program begins by reading in data which must be stored for continual use throughout each run. This includes such things as a list of symbols used such as the period, the comma, parentheses, etc. In addition, a list of all common words such as the, and, if, etc., which are to be eliminated from the frequency counting are read into the memory at this time. The machine begins its processing work by searching for a blank in the input data (a space which doesn't contain any letter or symbol). When it finds a blank, a word has been isolated in a crude form. Actually the word isolated may have a punctuation symbol as the first or last letter. This occurs in a word ending a sentence, for example. The period occurs immediately after the last letter and by just picking out blanks, the computer has included the punctuation with the word. Therefore, it was necessary to have a small routine in the program to check the first and last letter of each crude word isolated, against the symbol list read in. In this manner it was possible to eliminate the punctuation from the words.

Since the word has now been truly isolated, it is then compared to the list of common words in the memory of the machine. If a match is found, the computer ignores the word completely and goes on to the next word. If, however, a match is not found in the list of common words, then this word must be remembered and counted. In order to do so, the machine first compares this word to all the words stored previously. If a match is found here, the frequency of that word is simply increased by unity. If a match is not found, then the word is stored in a new location and given the frequency one.

This basic process is repeated until every word in the article has been examined. Then the results are printed out and punched out on cards. This includes the word, its frequency and length. Also at this point, the machine ignores all words that have only been used once for it is felt that they hold no real significance.

Since the memory of the computer is finite, it can handle an article of only a certain size. In this case, using the Fortran IV language on an IBM 7040 machine with a sixteen thousand unit memory, the practical limit is an article which uses seven hundred different, non-common words. This can not be translated into article lengths because of authors' differences in word selection, etc. Methods to increase the capacity of the program are discussed later.

3.4 ABSTRACT DETAIL (PHASE II)

The output data from Phase I is a group of cards containing the words used, their frequencies and lengths. This serves as input data to Phase II along with the cards containing the original article. This program begins in much the same manner as Phase I in that blanks are sought and words isolated without punctuation. When a word is isolated, a search is made through all the key words found in Phase I. If a match is found, a counter is increased by the number of times that the word was used in the article. For example, if a word was used five times in the article and found in a particular sentence, then the count for that sentence is increased by five. The program continues to isolate words and keep a tally of the frequencies until such time as it finds a period which designates the end of a sentence. When the end of a sentence has been found, a tally of frequencies for the whole sentence has been compiled.

In this program, only the highest counting five sentences were kept but this could be expanded to more or even to all sentences over a given tally. However, in this program, when a sentence tally has been made, it is compared with the tally of the five highest counting sentences previously stored. If it is higher than that of any of the other sentences, it replaces the lowest counting sentence in the memory. If the tally is lower, then the sentence is ignored and the program proceeds to the next sentence. This process continues until the whole article has been examined. At the end, the five highest counting sentences are remaining. These are the sentences containing the highest

WEIGHTED COUNT OF KEY WORDS. THEY ARE CONSIDERED TO BE THE KEY SENTENCES IN THE ARTICLE AND AS SUCH FORM THE AUTO-ABSTRACT.

A MAXIMUM OF 480 CHARACTERS, INCLUDING BLANKS AND PUNCTUATION HAS BEEN ALLOWED FOR EACH SENTENCE FOR PURPOSES OF CONSERVING THE MEMORY STORAGE CAPACITY. THIS, TOO, CAN BE INCREASED ACCORDING TO METHODS DISCUSSED LATER.

WHEN THE WHOLE ARTICLE HAS BEEN SCANNED, THE PRINT OUT IS EXECUTED WHICH PRINTS OUT ALL THE KEY WORDS AND THEIR FREQUENCIES AND THEN ALL THE KEY SENTENCES AND THEIR TALLIES. THIS PRINT OUT IS RATHER ARBITRARY. IN THIS CASE, THE ARTICLES PROCESSED WERE NOT MEANT FOR STORAGE BUT ONLY FOR EXPERIMENT. IN MORE SOPHISTICATED SYSTEMS, THE STORAGE MIGHT BE DONE AUTOMATICALLY AND NO OUTPUT FROM PHASE II WOULD BE REQUIRED.

A COPY OF PHASE II IS PRESENTED IN APPENDIX 2 AND AN EXAMPLE RUN THAT WAS ACTUALLY PROCESSED IS PRESENTED IN APPENDIX 3.

3.5 CO-ORDINATE INDEXING AND FILE INVERSION

EACH PIECE OF NEW INFORMATION COMING INTO THE SYSTEM MUST BE GIVEN AN ACCESSION NUMBER UNDER WHICH IT IS STORED. THESE WOULD NORMALLY BE GIVEN IN THE ORDER OF ARRIVAL OF THE INFORMATION. THIS MEANS THAT NO ATTEMPT AT NUMERICAL CLASSIFICATION IS MADE. THERE WOULD BE A RECORD KEPT OF WHICH ACCESSION NUMBER WAS ASSIGNED TO EACH ARTICLE.

NOW, FOR EACH ARTICLE, PHASE I HAS PRODUCED A LIST OF KEY WORDS FOR FURTHER PROCESSING. THEREFORE, THERE IS ESSENTIALLY A GROUP OF KEY WORDS FOR EACH ACCESSION NUMBER. THE PURPOSE OF THIS PART OF THE SYSTEM IS TO INVERT THIS FILE. INSTEAD OF HAVING A GROUP OF KEY WORDS FOR EACH ACCESSION NUMBER,

IT IS DESIRED TO HAVE A GROUP OF ACCESSION NUMBERS FOR EACH KEY WORD USED IN THE SYSTEM.

THIS CAN BE ACCOMPLISHED THROUGH A COMPUTER PROGRAM. THE MEMORY CAN BE USED TO STORE ALL THE ACCESSION NUMBERS AND KEY WORDS AND THEN BY SCANNING PROCEDURES, SORT THEM INTO A KEYWORD FILE. THEN ANY NEW INFORMATION WITH ITS ACCESSION NUMBER AND SEVERAL KEY WORDS COULD BE INCLUDED EASILY BY THE COMPUTER. THE FILE WOULD BEST BE KEPT ON CARDS THAT COULD BE PRINTED OUT WHEN DESIRED. WHEN A CONSIDERABLE AMOUNT OF NEW INFORMATION WAS PUT IN, A NEW SET OF CARDS COULD BE PUNCHED OUT.

THE STORAGE ON CARDS WOULD BE ADVANTAGEOUS IN THAT NO PART OF THE COMPUTER WOULD BE CONTINUALLY TIED UP AND ALSO THAT THE CARDS WOULD SERVE AS SUITABLE INPUT DATA FOR A PROGRAM THAT IS CONDUCTING A SEARCH FOR KEY WORDS. IF THE COMPUTER IS EQUIPPED WITH TAPE UNITS IN THE FUTURE, IT MIGHT BE MORE ADVANTAGEOUS TO USE THESE FOR STORAGE. THIS IS DEALT WITH A LITTLE MORE FULLY IN A LATER SECTION.

3.6 A COMPUTER INDEX

IT WAS INDICATED IN SECTION (3.2) THAT AN INDEX SIMILAR TO WADEX (SEE REFERENCE 11) COULD BE ACHIEVED AS A BY-PRODUCT OF THIS SYSTEM. IT SHOULD BE POINTED OUT THAT IT IS NOT ESSENTIAL TO THE OPERATION OF THE SYSTEM AND COULD BE PUT INTO USE ONLY WHEN DESIRED.

IF EVERY TIME A NEW PIECE OF INFORMATION WERE PUT INTO THE SYSTEM, A COMPUTER CARD OR CARDS WERE PREPARED CONTAINING THE TITLE, AUTHOR AND ACCESSION NUMBER, A FILE COULD BE KEPT FOR USE WHEN THE INDEX WAS DESIRED.

WADEX is an index that picks out all the key words in a title by simply eliminating what it considers to be common words according to its list of common words. Then it lists each article once for the author and once for every key word in the title. This could be accomplished by feeding in the list of cards containing the author's name, title and accession number. The key words could be picked out of the title including the author's name and then be arranged in alphabetical order. Then for each occurance of the key word or name, the complete title and accession number could be printed out.

The suggestion here is that if a program exists that picks out key words from the text of an article, why not include the most important of them in the list of key words determined from the title and list the article for each of these selected. This would not entail much extra work and would not tie up much computer-memory. Key word cards from Phase I could be included in the input data and processed in the same way as the key words selected in the title.

It has been pointed out that WADEX and similar computer indexes are very useful to people for browsing when they may not know exactly what they want. Therefore, it is thought that the inclusion of such an index, once the system was established, would contribute a great deal more than it would cost to maintain.

3.7 REQUEST FOR INFORMATION

As indicated in section (3.2), the selection of search key words may come from two different sources. A person requesting information may simply wish to list all the key words considered pertinent. This would be a most difficult operation for the user to do by himself. More could be accomplished by having the user discuss his problem briefly with the operator of the system who would cause the user, in his discussion to be more explicit than he would normally be compiling the list alone.

The second method of selecting key words would be to have the user of the system write in about three or four short paragraphs, the information requested and complete details of why it is requested. This might have to be done with the aid of an operator but when the user became used to the system, it could possibly be eliminated. This would facilitate use of the system over a considerable distance since user and operator would not have to converse freely. The written request would be subjected to a Phase I type program so that the key words in the request could be selected. It might be possible that the relative shortness of the request as compared to a full article would necessitate that some different criteria be used for the selection of key words. It might be necessary to screen out by means of a more extensive common word list, some words that would be key words for the whole article. This is a little difficult to determine until some experimentation is done in this area. However, it is thought that suitable results can be obtained.

If the search key words have come from a Phase I type program, then it could easily be arranged to have them punched out on cards. If they only come from a list prepared by the user, then the cards will have to be manually punched. By either of these methods, we have prepared part of the input data

FOR THE SEARCHING PROGRAM.

THE REMAINDER OF THE INPUT DATA WOULD BE THE KEY WORD FILE (IN CARD FORM) AS DISCUSSED IN SECTION (3.5). THE SEARCH KEY WORDS COULD BE KEPT IN THE MEMORY. THE KEY WORD FILE COULD BE SCANNED AND WHENEVER A KEY WORD WAS ENCOUNTERED THAT WAS ONE OF THE SEARCH KEY WORDS, ALL THE ACCESSION NUMBERS COULD BE PLACED IN THE MEMORY. THIS WOULD BE REPEATED UNTIL THE COMPLETE KEY WORD FILE WAS PROCESSED. THEN A COMPARISON OF ACCESSION NUMBERS FOR EACH KEY WORD WOULD BEGIN. WHEN MATCHES OF ACCESSION NUMBERS ARE FOUND IN GROUPS OF KEY WORDS, IT IS ASSUMED THAT THE ARTICLE HAS TO DO WITH THE GROUPING OF THE KEY WORDS. THIS IS THE THEORY BEHIND CLASSICAL CO-ORDINATE INDEXING. CRITERIA WOULD HAVE TO BE SET UP AS TO HOW MANY MATCHES AMONG HOW MANY KEY WORDS WOULD CONSTITUTE A POSITIVE REFERENCE. THIS COULD BE DONE AFTER SOME EXPERIMENTATION WAS DONE WITH A TRIAL SYSTEM. THE CRITERIA MIGHT BE LEFT ADJUSTABLE TO ALLOW FOR DIFFERENCES IN NEEDS OF USERS. FOR EXAMPLE, THE DESIGNER MAY BE IN A GREAT RUSH TO GET HIS REFERENCES AND MAY BE WILLING TO TAKE MANY FALSE DROPS THAT COULD BE ELIMINATED BY EXAMINATION OF THE TITLE AND/OR ABSTRACT. ON THE OTHER HAND, A RESEARCHER MAY HAVE MORE TIME AND WOULD WANT AS ACCURATE A RESPONSE FROM THE SYSTEM AS POSSIBLE.

ONCE THE POSITIVE REFERENCES HAVE BEEN SELECTED, THE ACCESSION NUMBERS COULD BE PRINTED OUT. THIS, IN ESSENCE, IS THE END OF THE SEARCH PROCEDURE ALTHOUGH A GOOD DEAL MORE COULD BE INCLUDED AS DESIRED. THE LIST OF TITLES VS. ACCESSION NUMBERS MAY BE A MANUAL FILE OR ONE PREPARED ON CARDS. IF THE LIST WAS ON COMPUTER CARDS, THE TITLES COULD BE PRINTED OUT VERY RAPIDLY

BY A VERY SIMPLE SORTING PROGRAM. THE USE OF A MACHINE PROGRAM HERE WOULD DEPEND ON THE SIZE OF THE LIST AND THE SPEED REQUIRED. THE RETRIEVAL OF THE ABSTRACTS AND DOCUMENTS WOULD BE VERY DEPENDENT ON THE DEGREE OF AUTOMATION USED IN THEIR STORAGE. IT WAS PREVIOUSLY MENTIONED THAT A MANUAL FILE OF ABSTRACTS AND DOCUMENTS COULD BE KEPT. THE USER, SUPPLIED WITH THE ACCESSION NUMBERS, WOULD SIMPLY SEARCH THE FILE UNTIL HE FOUND THE INFORMATION THAT WAS REQUIRED. HOWEVER, DEPENDING ON HOW MUCH INFORMATION WAS BEING REQUESTED PER UNIT TIME, IT MIGHT BE NECESSARY TO AUTOMATE THIS PROCESS IN ORDER TO BE ABLE TO SATISFY ALL THE USERS. SEVERAL MICROFILM COMPANIES HAVE COMPUTER CARDS THAT CAN BE SORTED BY ACCESSION NUMBERS AND THAT HAVE MICROFILM MOUNTED IN THEM. A SYSTEM EMPLOYING SUCH CARDS WOULD BE SPEEDED UP. THE SORTING COULD BE DONE BY MACHINE AND THE USER COULD GET HIS INFORMATION OVER A MICROFILM VIEWING SCREEN OR EVEN HAVE COPIES OF THE INFORMATION MADE ON A MICROFILM PRINTER.

IT IS THE FEELING OF THE AUTHOR THAT INITIALLY A MANUAL STORAGE AND RETRIEVAL BE USED. THIS WOULD GIVE FAIRLY GOOD ACCESS TO QUITE A LARGE SYSTEM. WHEN THE SYSTEM GREW BEYOND THIS, THEN MORE AUTOMATIC METHODS COULD BE EMPLOYED AS DISCUSSED IN SECTION (3.9).

3.8 ADDITION OF A THESAURUS

IN ANSWER TO THE PROBLEM OF SYNONYMS, IT IS RECOMMENDED THAT A THESAURUS LOOK-UP BE INCORPORATED INTO THE PROGRAM. A THESAURUS SUCH AS THAT PREPARED BY THE ENGINEERS JOINT COUNCIL IN THE UNITED STATES MIGHT BE PUT INTO A TECHNICAL SYSTEM. THE EJC THESAURUS IS AVAILABLE NOW ON TAPE. SINCE

THE COMPUTER IS NOT CURRENTLY EQUIPPED FOR TAPE, IT WOULD BE NECESSARY TO HAVE A CONVERSION TO CARDS. THEN WHEN THE SEARCH KEY WORDS WERE SELECTED, IT WOULD BE POSSIBLE TO HAVE A PROGRAM THAT WOULD SCAN THE THESAURUS CARDS FOR MATCHES WITH THE KEY WORDS. WHEN A MATCH WAS FOUND, ALL THE SYNONYMS COULD BE INCLUDED IN THE SEARCH KEY WORDS. IT WOULD BE EXPECTED THAT THIS WOULD INCREASE THE NUMBER OF REFERENCES SUPPLIED CONSIDERABLY.

THE ACTUAL PREPARATION OF SUCH A PROGRAM AND A SET OF THESAURUS CARDS WOULD NOT BE VERY COMPLICATED AND WOULD NOT TAKE VERY LONG, AND THEREFORE, IT MUST BE CONSIDERED EVEN MORE WORTHWHILE. THERE WOULD, HOWEVER BE A TREMENDOUS NUMBER OF CARDS THAT WOULD REQUIRE STORING AND HANDLING UNTIL SUCH TIME AS THE COMPUTER WAS EQUIPPED TO HANDLE THE THESAURUS IN TAPE FORM. IN ADDITION, IF A REQUEST WAS ENCOUNTERED THAT WAS EXTREMELY URGENT, THEN THE THESAURUS STEP COULD BE BY-PASSED ON AN INITIAL RUN UNTIL SOME OF THE REFERENCES WERE QUICKLY FOUND. THEN THE THESAURUS STEP COULD BE INCLUDED IN ORDER TO COMPLETE THE LIST OF REFERENCES.

3.9 A FINAL PROPOSAL

THE BASIC SYSTEM IN THIS DESIGN HAS BEEN OUTLINED AND IT IS EXPECTED THAT MOST OF THESE SECTIONS WILL REMAIN INTACT. THE ONLY PART THAT IS SUBJECT TO A GREAT DEAL OF CHANGE IS THAT HAVING TO DO WITH THE STORAGE OF THE INFORMATION. THE EQUIPMENT USED IN THIS SECTION CAN BE ANYTHING FROM A MANUAL SYSTEM UP TO A COMPLETELY AUTOMATED ONE. THE METHOD OF STORAGE ACTUALLY USED WILL DEPEND ON THE SOPHISTICATION OF THE SYSTEM AND ALSO ON THE NUMBER OF PEOPLE SERVED BY IT. THE MORE PEOPLE THAT ARE SERVED, THE MORE MONEY WILL

BE AVAILABLE FOR MORE SOPHISTICATED EQUIPMENT AND THE MORE PEOPLE SERVED, THE MORE SOPHISTICATED WILL THE EQUIPMENT AND THE SYSTEM HAVE TO BE. IT IS WITH THIS EVOLUTIONARY PROGRESS OF AN INFORMATION SYSTEM IN MIND THAT THE FOLLOWING PROPOSAL FOR THE ENGINEERING FACULTY AT MCMASTER UNIVERSITY HAS BEEN MADE. IN ESSENCE, IT IS SIMPLY A PROPOSAL FOR THE CREATION OF AN AUTOMATED INFORMATION STORAGE AND RETRIEVAL SYSTEM BUT IN THREE DISTINCT LEVELS OF SOPHISTICATION.

THE FIRST LEVEL SUGGESTED IS THE SIMPLEST. THIS WILL BE USED UNTIL THE SYSTEM HAS BEEN DEVELOPED TO A STAGE WHERE IT REQUIRES MORE ADVANCED EQUIPMENT. IT IS PROPOSED, ON THIS LEVEL, THAT THE SYSTEM BE SET UP AS HAS BEEN OUTLINED IN THIS REPORT, AND THAT THE STORAGE IS DONE MANUALLY. THE FIRST STEP WOULD BE TO FORMALLY WRITE THE COMPUTER PROGRAMS FOR THE SECTIONS OMITTED HEREIN. IT IS SUPPOSED THAT NOT MUCH DIFFICULTY WILL BE ENCOUNTERED HERE EXCEPT THAT THE PROGRAMS MUST BE WRITTEN BY SOMEONE COMPLETELY FAMILIAR WITH THE MACHINE LANGUAGE. THEREFORE, IT WOULD BE BEST TO HAVE AN OPERATOR, TRAINED IN THE MACHINE LANGUAGE, WHO WOULD SET UP THE SYSTEM AND WHO COULD MAKE CHANGES TO IT AS REQUIRED LATER ON.

NOW, IT IS SUGGESTED THAT EACH DEPARTMENT OF THE ENGINEERING FACULTY COULD PARTICIPATE IN THE PROGRAM. EACH DEPARTMENT HAS ITS OWN BOOKS THAT MUST BE PROCESSED. ITIS SUGGESTED THAT ONE OR TWO STENOGRAPHERS BE HIRED AS REQUIRED TO PUNCH BOOKS, ARTICLES, ETC., ONTO CARDS FOR INPUT AND ALSO TO BE OF ASSISTANCE TO THE OPERATOR. SINCE EACH DEPARTMENT IS A SEPARATE FIELD, EACH IS EXPECTED TO HAVE A JARGON OF ITS OWN. THEREFORE, EACH COULD PREPARE its own common word list for processing its own articles. As an initial step, each department would be required to have its own storage. It is thought that this would be completely manual at this stage. There would have to be an abstract file according to accession numbers and the complete article would have to be stored under its accession number. It is possible that the print out from Phase II (the abstract) could be done in such a format, that it could be attached to the face of a card bearing its accession number. This would facilitate storing and sorting when necessary. The key word file on cards would be kept by the operator as a central file.

The scope of this first stage of development doesn't go much beyond the actual setting up of the system. This is done in this manner so as to develop and test the system without incurring additional equipment costs. The only costs here would be for an operator, stenographers as required and the time of the people doing the work in each department. It is hoped that at this stage, most of the difficulties of the system could be ironed out and after a period of approximately two years, the system would be operating smoothly.

When the system reaches such a point, it is ready for the second level of sophistication. The system would essentially remain unchanged except that storage methods would be greatly improved. Necessary to such an improvement would be the addition of tape unit machines to the IBM 7040 in use at the present time. Then tape storage would replace card storage of the key word file. The key word file could be kept on tape and scanned in that form. All

THE ABSTRACTS AND DOCUMENT LISTS COULD BE PREPARED ON TAPE AND STORED IN THAT FORM. ALL THOSE PREVIOUSLY PUNCHED ONTO CARDS COULD SIMPLY BE CONVERTED ONTO TAPE BY THE COMPUTER. THIS WOULD ENABLE MANY MORE PEOPLE TO USE THE SYSTEM FROM A GREATER DISTANCE. WHEN A REQUEST CAME IN, THE ABSTRACT AND THE WHOLE ARTICLE COULD BE PRINTED OUT AND THE USER OF THE SYSTEM WOULD NOT HAVE TO HAVE THE ORIGINAL SENT TO HIM.

THE COST OF SUCH A TAPE UNIT SHOULD NOT BE BORN COMPLETELY BY THIS PROGRAM BECAUSE IT WOULD UNDOUBTEDLY BE OF TREMENDOUS USE TO EVERYONE THAT USES THE COMPUTER. IT IS QUITE POSSIBLE THAT BEFORE THIS INFORMATION SYSTEM WAS READY FOR THE TAPE UNITS, THEY MIGHT BE INSTALLED ANYWAY FOR OTHER PURPOSES. THE TRANSITION FROM THE FIRST TO THE SECOND LEVEL OF SOPHISTICATION COULD BECOME GRADUAL. IN ANY CASE, THE USE OF SUCH TAPE UNITS WOULD PROBABLY SERVE QUITE EFFICIENTLY UNTIL THE SYSTEM GREW TO A VERY LARGE SIZE.

WHEN THE SYSTEM REACHED A LARGE SIZE, IT WOULD JUSTLY BE CALLED AN INFORMATION CENTRE. BY THIS TIME, IT WOULD BE READY FOR ITS THIRD STAGE OF SOPHISTICATION. THIS WOULD BE THE INCLUSION OF A COMPLETELY AUTOMATIC STORAGE SYSTEM. SUCH A SYSTEM HAS NOW BEEN DEVELOPED BY THE AMPEX CORPORATION IN THE UNITED STATES. THE SYSTEM IS CALLED VIDEOFILE AND USES VIDEO TAPE MUCH THE SAME AS IS USED IN TELEVISION WORK. THE ARTICLES ARE PHOTOGRAPHED WITH A TELEVISION CAMERA AND STORED ON VIDEO TAPE. THEY CLAIM THAT A QUARTER OF A MILLION DOCUMENT PAGES CAN BE STORED ON ONE FOURTEEN INCH REEL OF TAPE. WHEN THE ACCESSION NUMBER OF DESIRED INFORMATION IS KNOWN, IT IS SIMPLY PUNCHED INTO A KEYBOARD AND THE MACHINE QUICKLY FINDS THE ARTICLES AND ABSTRACTS ON THE

TAPE. THEY CAN BE PORTRAYED ON A TELEVISION MONITOR SCREEN FOR VIEWING AND WHEN REQUIRED, HARD PERMANENT COPIES CAN BE MADE. SUCH A SYSTEM IS THE VERY ULTIMATE IN MODERN STORAGE TECHNIQUES AND IS THOUGHT TO BE IDEAL FOR A WELL DEVELOPED INFORMATION SYSTEM.

AN AMPEX VIDEOFILE IS SCHEDULED TO BE INSTALLED AT THE NATIONAL AERONAUTICS AND SPACE ADMINISTRATION IN THE UNITED STATES FOR A COST OF $875,000. THE COST OF SUCH A SYSTEM OF COURSE VARIES AND THEY CLAIM A SYSTEM CAN BE SET UP FROM BETWEEN $200,000 TO $1,000,000, DEPENDING ON THE AMOUNT OF AUXILIARY EQUIPMENT REQUIRED. THEREFORE, IT IS SEEN THAT THIS TYPE OF SYSTEM COULD BE USED ONLY WHEN THE WHOLE INFORMATION SYSTEM BECAME VERY LARGE BUT HERE AGAIN THE EXTRA EQUIPMENT COULD BE USED FOR MANY OTHER THINGS AND THE TOTAL COST SHOULD NOT BE CHARGED AGAINST THIS SYSTEM.

THE EVOLUTION OF THIS SYSTEM WITH RESPECT TO ITS STORAGE EQUIPMENT HAS BEEN LAID OUT IN STEP FASHION AS A SUGGESTED PLAN OF DEVELOPMENT. IT IS SUPPOSED THAT IN THE TIME THAT IT TAKES IN GOING THROUGH EACH STAGE, NEW EQUIPMENT WILL BE DEVELOPED AND CHANGES MAY BE REQUIRED BUT THE PROCEDURE WILL NEVERTHELESS BE THE SAME.

4. DISCUSSION AND CONCLUSIONS

The system proposed is not intended to solve all of the problems of the field of information storage and retrieval. It is rather a first attempt at the formalization and design of an essentially automated system. Because of the time limitations involved, only parts of the system have been detailed but indications as to how the remainder might be programmed have been given. It is felt that any person experienced in computers, could prepare programs for the remaining sections using the outline given.

One of the main limiting factors of this system has been the lack of additional memory in the IBM 7040 computer. Each of Phase I and II, has been designed, as they are presented herein, to occupy essentially all of the memory available. It has been noted that Phase I, as it stands, can only accomodate an article using seven hundred different, non-common words, because of lack of computer memory.

A major increase in the capacity of this system could be achieved by having the program re-written in MAP, a basic symbolic language for the 7040. In Fortran, an A1 field has been used for each letter and space. After words, sentences, etc., have been identified, storage in the memory has also been in A1 fields. Now if a basic MAP language were used, it would be possible to read and identify words and sentences in A1 fields as before, but store them in A6 fields. This is generally known as a packing procedure. Each

A field can hold a maximum of six characters. When identifying letters and spaces, it is necessary to have only one letter per A field. However, once identification has been made, it would be possible to store six letters where only one was previously. At present, there is no procedure for packing in Fortran IV.

The use of the MAP language would increase the capacity of the system for two reasons. First of all, the packing, multiplies the storage capacity by a factor of approximately five. Secondly, when a program is written in MAP, there would be, according to computer experts, an increase in the computer's memory capacity of up to ten percent, which is significant. Therefore, it is recommended that any further work using this system, should have all the programs written in MAP language using a packing procedure.

It should be noted that this system was designed in pieces, largely because of memory restrictions. Additions to memory made by packing procedures would be used to increase the capacity of the system so that it could accomodate much larger articles. Even a five-fold increase by packing would allow an article to be processed that uses three thousand five hundred different, non-common words. This would indeed be an article of considerable size. If at some time it was found that for most operations, there was an excess of memory, it would be possible to combine parts or all of Phase I, Phase II and the Co-ordinate Indexing steps. Such a change is not absolutely necessary, but it would eliminate some of the transferring of cards used. Such a change might be possible if additional core storage was added to the

IBM 7040 at some time.

In time, it is expected that a few other changes will have been made in the system. For example, it is expected that the common word list will expand to a much larger size. Since the programs used here are experimental, the common word list is quite small. It is possible that if this system were used for different fields, each field might have its own list of common words. For example, the word chemical would not be expected to hold much significance in the field of chemical engineering and yet in an article on electrical or mechanical engineering, the word chemical might be very important.

Another change that might occur as more experimentation is done in this field, is the criteria for selecting key words and/or sentences. Currently a simple frequency count has been used. Luhn has suggested briefly that some words of very high frequency really hold no significance. Such a word might be the word 'machine' in mechanical engineering. Luhn has followed this up by suggesting that low and very high frequency words be eliminated from the selection of key sentences. This and other future criteria changes should all be experimented with until satisfactory operating criteria have been found.

Currently, it has been suggested that co-ordinate indexing be done using all of the key words. However, further work may show that only a portion of the more important key words should be used to aid the problem of false drops. This change is very likely, but the determination of the optimum cut-off point would have to be found by experiment.

If too many false drops were occurring for some users of the system

IT WOULD BECOME NECESSARY TO INCLUDE ROLE INDICATORS IN THE SYSTEM AS ARE NOW USED IN CLASSICAL CO-ORDINATE INDEXING. THIS WOULD MAKE THE SYSTEM CONSIDERABLY MORE COMPLEX AND WOULD SLOW DOWN A SEARCH A GREAT DEAL. THEREFORE, IF IT BECOMES NECESSARY TO INCLUDE ROLE INDICATORS, IT IS SUGGESTED THAT THEY BE INCLUDED AS AN AUXILIARY PROGRAM TO BE USED ONLY WHEN NECESSARY. SEARCHES FOR INFORMATION CONDUCTED BY DESIGNERS WOULD NOT LIKELY NEED THE USE OF THE ROLE INDICATORS.

THE MOST REMARKABLE CHANGES THAT OCCUR IN THIS FIELD TAKE PLACE IN EQUIPMENT CHANGES. RECENTLY, MUCH PREVIOUSLY UNHEARD OF EQUIPMENT IN THE FIELD OF INFORMATION STORAGE AND RETRIEVAL HAS BEEN RELEASED. EQUIPMENT SUCH AS ALL THE MICROFILM SYSTEMS AND EQUIPMENT AND NEWER DEVELOPMENTS SUCH AS THE PROPOSED VIDEO TAPE SYSTEMS, WILL PROBABLY BE OF GREAT HELP TO THE FIELD. NEW DEVELOPMENTS ARE SOON TO BE RELEASED, SUCH AS PRINT READING MACHINES THAT ARE CURRENTLY UNDER EXPERIMENTATION. AS THE INFORMATION EXPLOSION HAS BROUGHT ABOUT INTEREST IN THE FIELD OF INFORMATION STORAGE AND RETRIEVAL, SO HAS THE INTEREST BROUGHT ABOUT A MINOR EXPLOSION OF INFORMATION STORAGE EQUIPMENT. ANY EQUIPMENT SPECIFIED FOR A PARTICULAR SYSTEM TODAY WILL LIKELY BE OUT-MODED IN A VERY SHORT TIME.

IN GENERAL, IT IS FELT THAT A GIVEN SYSTEM OF INFORMATION STORAGE AND RETRIEVAL DOES NOT CHANGE NEARLY AS QUICKLY AS THE EQUIPMENT USED WITHIN IT. EQUIPMENT MAY BE CHANGED WITHIN A SYSTEM TO MAKE INFORMATION MORE READILY ACCESSIBLE OR TO SPEED UP A CERTAIN OPERATION, AND YET THE BASIC THEORETICAL SYSTEM STRUCTURE HAS BEEN UNALTERED. FOR THIS REASON, THIS DESIGN HAS BEEN

CARRIED OUT WITH THE SYSTEM HAVING PRIMARY IMPORTANCE AND THE EQUIPMENT HAVING ONLY SECONDARY IMPORTANCE.

5. REFERENCES

1. Becker, J. and R. M. Hayes– "Information Storage and Retrieval – Tools, Elements, Theories" , Wiley, 1963, 448 pp.

2. Logue, P. – "Deep Indexing Technical Reports" , J. of Chem. Doc. 2, No. 4, Oct. 1962.

3. Morse, R. – "Information Retrieval" , Chem. Eng. Progress, Vol. 57, No. 5.

4. Holm, B. E. –"Information Retrieval – A Solution" , Chem. Eng. Progress, Vol. 57, No. 6.

5. Jacobson, R. A. –"Locating Engineering Literature" , Machine Design, Vol. 36, June 18, 1964, pp. 152–158.

6. Bar–Hillel, Y. – "Some Theoretical Aspects of the Mechanization of Literature Searching" , Office of Tech. Services, Washington, 1960, 74 pp., (PB161547)

7. Perry, J. W., A. Kent and M. M. Berry – "Machine Literature Searching" , Vol. 1, Interscience, 1956, 175 pp.

8. Perry, J. W. and A. Kent –"Documentation and Information Retrieval–

An Introduction to Basic Principles and Cost Analysis", Interscience, 1957, 168 pp.

9. Luhn, H.P. -"A Statistical Approach to Mechanized Encoding and Searching of Literary Information" , IBM Journal of Res. and Dev 1, No. 4, Oct. 1957, pp. 309-317.

10. Luhn, H.P. -"The Automatic Creation of Literature Abstracts", IBM Journal of Res. and Dev. 2, No. 2, Apr. 1958, pp. 159-165.

11. Ripperger, E.A., H. Wooster and S. Juhasz -"Wadex - A New Tool in Literature Searching" , Mechinical Engineering, Vol. 86, No. 3, March 1964, pp. 45-50.

12. Vickery, B.C. -"On retrieval System Theory" Butterworth, 1961, 159 pp.

6. LIST OF ILLUSTRATIONS

FIGURE 1. USE OF KEY WORD FILE IN CO-ORDINATE INDEXING

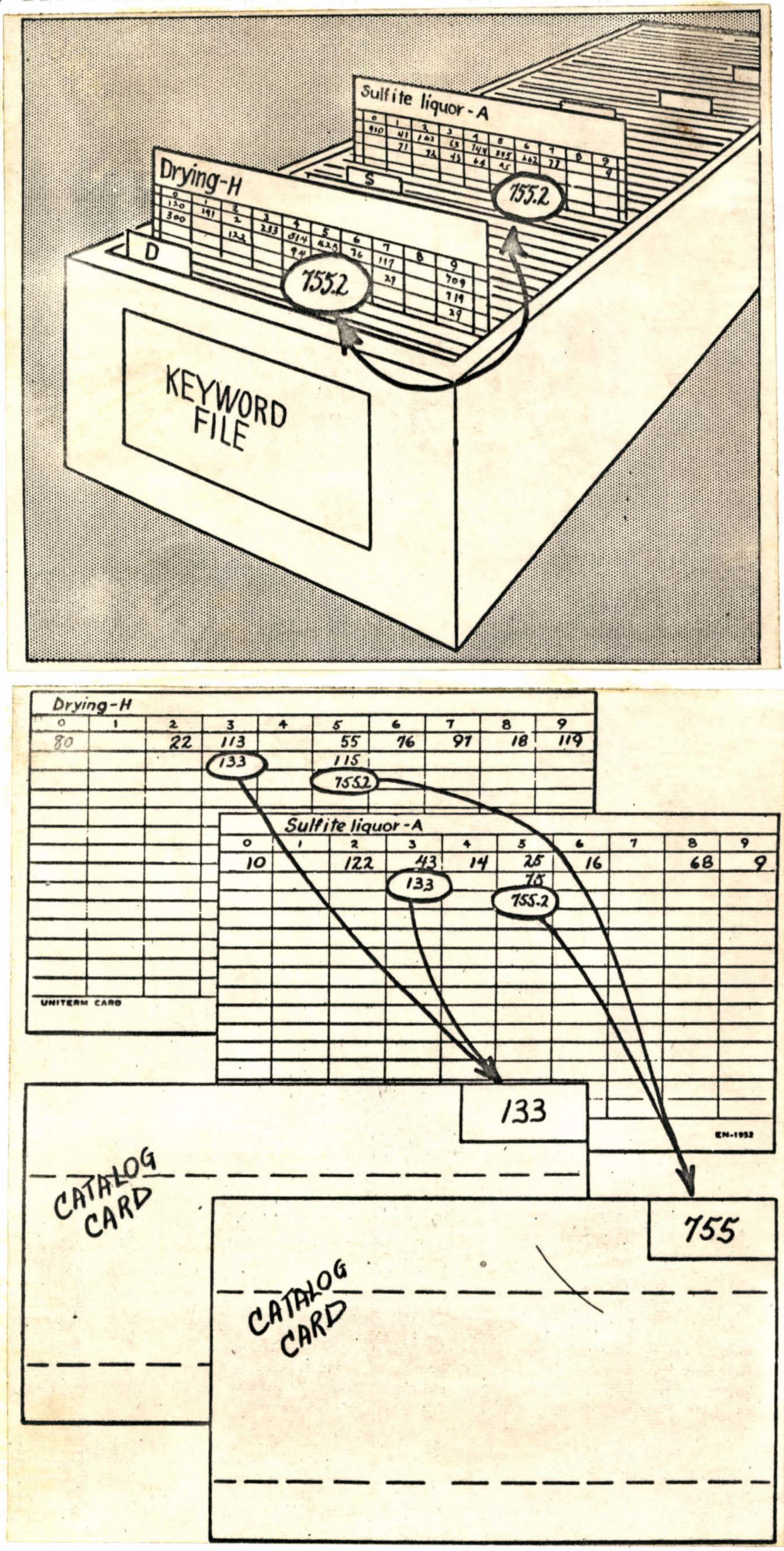


FIGURE 2. ENCODED ABSTRACT IN WESTERN RESERVE SYSTEM

```
SNØØ19Ø19ØATR/C.O.BURGESS;R.W.BISHOP
DCT/TRAMFRMAS;PPT44-16PP35YR1944¶||⌇
LOKSACNMYTLØ1KALRYRNØ1||⌇KAJKEJKOVRA
RNØ4KUJ&C&/>2.5%<3.5%& Si&/>1%<2.5%K
WVPAPR48PAPR55PAPRMYTL11⌇KAKMWTLP
ARSLYDD22KQJLALLØ1KUJ&Si,Mn,Zr&KQJM
ACHLQCTHQTTØ7/35LBKOKTAMPØ1/<28ØØ°F
⌇KAKFWRMPARSØ1KOKTAMPØ1/<26ØØ°F⌇KWJ
KOVRARNØ4KUJ& C,Si,Mn,Zr& /<<% KWVKQV
PAPR48PAPR55PAPRMYTL11⌇KUVFWMRPAR
SØ1KUVRARNØ4KUJ &&%KUVFWRMPARSØ1⌇|| ¶
```

FIGURE 3. TERMS USED IN ABSTRACT IN FIGURE (2)

Serial No.	SN
00190190	ØØ19Ø19Ø
Authors (namely)	ATR/
C. O. Burgess	C.O BURGESS/
R. W. Bishop	R.W.BISHOP
Document reference (namely)	DCT/
Trans. Am. Foundrymens' Assoc.	TRAMFRMAS
Preprint No. 44-16	PPT44-16
35pp.	PP35
(1944) (year of publication)	YR1944
Start of encoded abstract	¶
Start of sentence	||
Start of phrase	⌇
Field (to which document pertains)	LOK
Metallurgy	SACNMYTLØ1
General attributive relation	KAL
Ferrous	RYRNØ1
End of preceding phrase	⌇
End of sentence (start of next sentence)	||
Start of phrase	⌇
Starting material	KAJ
Material processed	KEJ
Properties given for	KOV
Gray iron	RARNØ4
Containing (components)	KUJ
Carbon (chemical element)	&C&
Namely, more than 2.5%, less than 3.5%	/>2.5%<3.5%
Silicon (chemical element)	&Si&
Namely, more than 1%, less than 2.5%	/>1%<2.5%
Properties	KWV
Tensile strength	PAPR48
Transverse strength	PAPR55
Chill depth	PAPRMYTLØ8
Start of phrase (end of preceding phrase)	⌇
Process	KAK
Ladle treatment	MWTLPARSLYDD22
By means of	KQJ
Alloy	LALLØ1
Components	KUJ
Silicon, manganese, zirconium (chemical elements)	&Si,Mn,Zr&

FIGURE 3. (CONTINUATION)

By means of	KQJ
Induction furnace	MACHLQCTHQTT07
Namely 35 lb.	/35LB
Condition	KOK
Temperature	TAMP01
Namely below 2800 F	/< 2800°F
Start of phrase (end of preceding phrase)	I
Process	KAK
Casting	FWRMPARS01
Condition	KOK
Temperature	TAMP01
Namely below 2600 F	/< 2600°F

Start of phrase (end of preceding phrase)	I
Product	KWJ
Properties given for	KOV
Gray iron	RARN04
Containing (components)	KUJ
Carbon, silicon, manganese, zirconium (chemical elements)	&C, Si, Mn, Zr&
Namely, in small amounts	/<< %
Properties	KWV
Improvement	KQV
Tensile strength	PAPR48
Transverse strength	PAPR55
Depth of chill	PAPRMYTL11
Start of phrase (end of preceding phrase)	I
Discussion	KUV
Ladle treatment	MWTLPARSLYDD22
Discussion	KUV
Gray iron	RARN04
Composition (chemical)	KUJ&&%
Discussion	KUV
Casting process	FWRMPARS01
End of phrase	I
End of sentence	||
End of abstract	¶

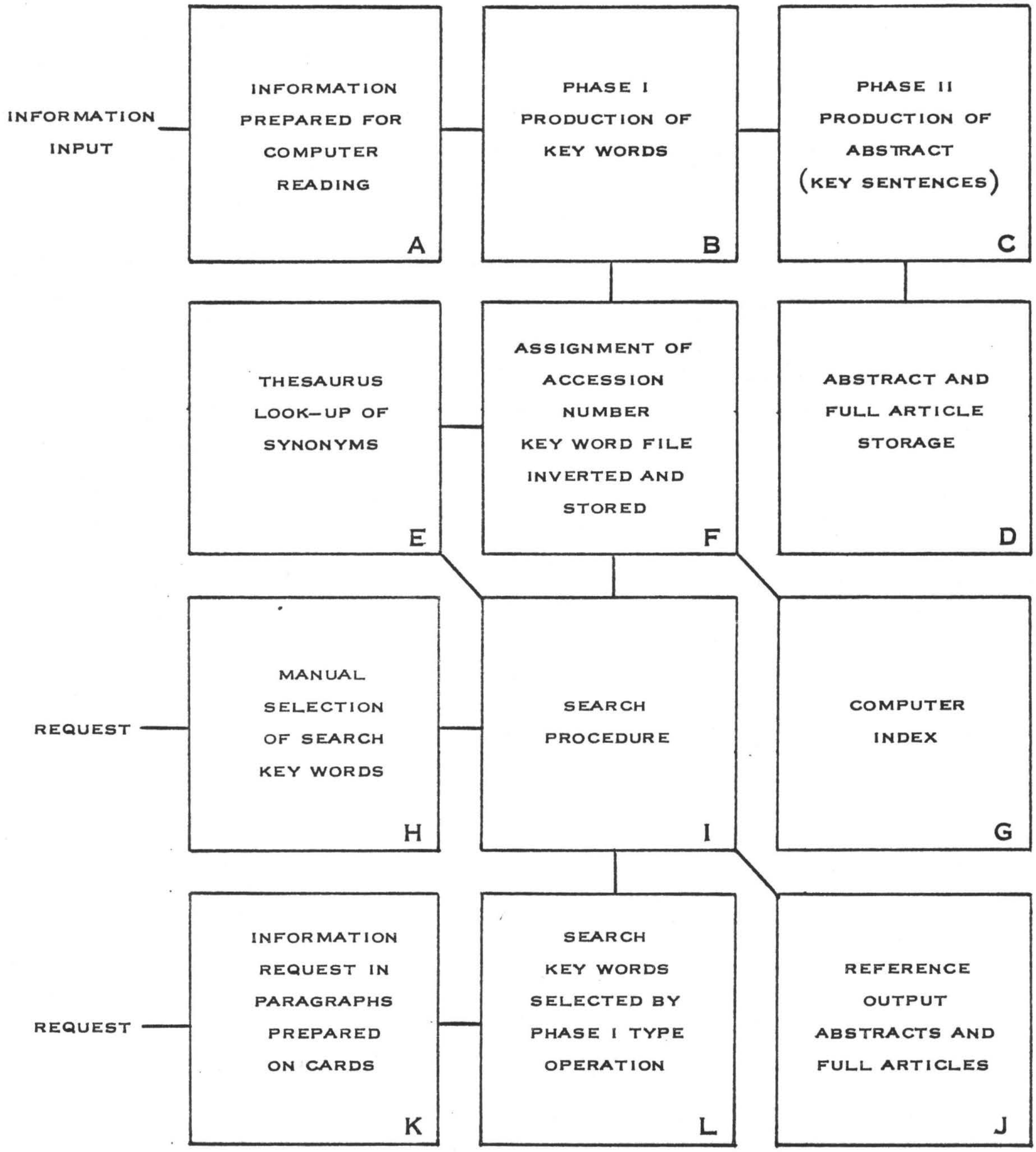
Figure 4. Example of KWIC Index

```
TERMINATION OF VERTICAL WIND DISTURBANCES                             --
 OSCILLATIONS CAUSED BY WIND GUSTS                                    --
                        WIND LOAD ON TOWERS                           --
                FORM OF WIND PROFILE IN NEAR GROUND LAYER             --
                      A WINDTUNNEL INVESTIGATION INTO THE             --
N A 45 DEGREE SWEPTBACK WING AT TRANSONIC SPEEDS                      --
                        WING PLAN FORMS FOR TRANSONIC SPE             --
STRIBUTION ON SYMMETRIC WING PROFILES IN THE CASE OF HIGH             --
TS OF AN ASPECT-RATIO-1 WING WITH FAN AT 0.354 CHORD                  --
SSURE DISTRIBUTION ON A WINGSURFACE IN A NON-UNIFORM SUPE             --
```

KWIC TYPE INDEX WITHOUT RECIRCULATION

DOTTED LINES SHOW POSITIONS OF BIBLIOGRAPHICAL REFERENCES

FIGURE 5. DESIGN SCHEMATIC



APPENDICES

APPENDIX 1

```
C       FIRST PHASE OF AUTO-ABSTRACT PROGRAM
        DIMENSION A(160),S(11),F1(5,1),F2(20,2),F3(16,3),F4(
       116,4),F5(8,5),WD(5600),NS(700),NF(700),LW(700),NFQ(7
       200)
    1 FORMAT (80A1)
      READ 1, BLANK
      READ 1, (S(I),I=1,11)
      READ 17,(F1(I,1),I=1,5)
   17 FORMAT (5(A1,1X))
      READ 18,((F2(I,J),J=1,2),I=1,20)
   18 FORMAT (20(2A1,1X))
      READ 19,((F3(I,J),J=1,3),I=1,16)
   19 FORMAT (16(3A1,1X))
      READ 20,((F4(I,J),J=1,4),I=1,16)
   20 FORMAT (16(4A1,1X))
      READ 21,((F5(I,J),J=1,5),I=1,8)
   21 FORMAT (8(5A1,1X))
      READ 90, NC
   90 FORMAT (I4)
      NNC=1
   63 KTR=0
    2 READ 1,(A(I),I=1,80)
    3 READ 1,(A(I),I=81,160)
   64 K=0
```

```
      I=1
    4 IF (A(I).EQ.BLANK) GO TO 5
      I=I+1
      GO TO 4
    5 IF ((I-K).NE.1) GO TO 6
      IF (I.EQ.80) GO TO 7
      K=I
      I=I+1
      GO TO 4
    6 J=1
      KK=K
      II=I
   12 IF(A(KK+1).NE.S(J)) GO TO 13
      KK=KK+1
   13 IF (A(II-1).NE.S(J)) GO TO 15
      II=II-1
   15 IF (J.EQ.11) GO TO 16
      J=J+1
      GO TO 12
   16 L=II-KK-1
      IF (L.GT.5) GO TO 28
      GO TO (22,23,24,25,26),L
   22 J=1
  122 IF (A(KK+1).EQ.F1(J,1)) GO TO 27
      IF (J.EQ.5) GO TO 28
      J=J+1
```

```
      GO TO 122
  23  J=1
 323  JJ=1
      KKK=KK
 223  IF (A(KKK+1).NE.F2(J,JJ)) GO TO 123
      IF (KKK.EQ.(KK+1)) GO TO 27
      KKK=KKK+1
      JJ=JJ+1
      GO TO 223
 123  IF (J.EQ.20) GO TO 28
      J=J+1
      GO TO 323
  24  J=1
 324  JJ=1
      KKK=KK
 224  IF (A(KKK+1).NE.F3(J,JJ)) GO TO 124
      IF (KKK.EQ.(KK+2)) GO TO 27
      KKK=KKK+1
      JJ=JJ+1
      GO TO 224
 124  IF (J.EQ.16) GO TO 28
      J=J+1
      GO TO 324
  25  J=1
 325  JJ=1
      KKK=KK
```

```
225 IF (A(KKK+1).NE.F4(J,JJ)) GO TO 125
    IF (KKK.EQ.(KK+3)) GO TO 27
    KKK=KKK+1
    JJ=JJ+1
    GO TO 225
125 IF (J.EQ.16) GO TO 28
    J=J+1
    GO TO 325
 26 J=1
326 JJ=1
    KKK=KK
226 IF (A(KKK+1).NE.F5(J,JJ)) GO TO 126
    IF (KKK.EQ.(KK+4)) GO TO 27
    KKK=KKK+1
    JJ=JJ+1
    GO TO 226
126 IF (J.EQ.8) GO TO 28
    J=J+1
    GO TO 326
 27 GO TO 7
 28 KKK=KK+1
    III=II-1
    IF (KTR.NE.0) GO TO 31
    DO 30 M=KKK,III
    MM=M-KKK+1
 30 WD(MM)=A(M)
```

```
      LW(1)=L
      NS(1)=1
      NF(1)=L
      NFQ(1)=1
      KTR=KTR+1
      GO TO 7
   31 NN=1
   33 IF (L.NE.LW(NN)) GO TO 32
      GO TO 34
   32 IF (NN.EQ.KTR) GO TO 37
      NN=NN+1
      IF (NN.EQ.700) GO TO 98
      GO TO 33
   34 NSI=NS(NN)
      NFI=NF(NN)
      J=0
   35 KKKJ=KKK+J
      NSIJ=NSI+J
      IF (A(KKKJ).NE.WD(NSIJ)) GO TO 32
      IF (NSIJ.EQ.NFI) GO TO 36
      J=J+1
      GO TO 35
   36 NFQ(NN)=NFQ(NN)+1
      GO TO 7
   37 KTR=KTR+1
      NKK=NF(KTR-1)+1
```

```
      LW(KTR)=L
      NS(KTR)=NKK
      NF(KTR)=NKK+L-1
      NFQ(KTR)=1
      DO 40 M=KKK,III
      MM=M-KKK+NKK
   40 WD(MM)=A(M)
      IF ((NF(KTR)+L).GT.5600) GO TO 98
      GO TO 7
    7 IF (I.GE.80) GO TO 8
      K=I
      I=I+1
      GO TO 4
    8 DO 9 J=1,80
      JJ=J+80
    9 A(J)=A(JJ)
      IF (I.EQ.80) GO TO 11
      II=I-80
      DO 10 J=1,II
   10 A(J)=BLANK
   11 IF (NC.EQ.NNC) GO TO 91
      NNC=NNC+1
      IF (NC.EQ.NNC) GO TO 64
      GO TO 3
   98 PRINT 99
   99 FORMAT (1X,14HTOO MANY CARDS)
```

```
      PRINT 97,NNC
   97 FORMAT (1X,I4)
   91 J=1
      KR=KTR
  191 IF (NFQ(J).NE.1) GO TO 192
      KR=KR-1
  192 J=J+1
      IF (J.EQ.(KTR+1)) GO TO 196
      GO TO 191
  196 PUNCH 100,KR
      J=1
   92 M=NS(J)
      N=NF(J)
      IF (NFQ(J).EQ.1) GO TO 295
      PRINT 95,(NFQ(J),LW(J),(WD(I),I=M,N))
   95 FORMAT (1X,I4,3X,I3,3X,70A1)
      PUNCH 100,(NFQ(J),LW(J))
  100 FORMAT (I4,3X,I3)
      PUNCH 1,(WD(I),I=M,N)
  295 J=J+1
      IF (J.EQ.(KTR+1)) GO TO 96
      GO TO 92
   96 STOP
      END
```

APPENDIX 2

```
C      SECOND PHASE OF AUTO-ABSTRACT PROGRAM
       DIMENSION A(480),S(11),WD(5600),NS(700),NF(700),SN(4
      180,5),KS(5)
     1 FORMAT (80A1)
       KA=0
       READ 1, BLANK
       READ 1, (S(I),I=1,11)
       M=1
       J=1
       READ 100, KR
   101 READ 100,(NFQ(J),LW(J))
   100 FORMAT (I4,3X,I3)
       N=M+LW(J)-1
       READ 1,(WD(I),I=M,N)
       NS(J)=M
       NF(J)=N
       M=N+1
       J=J+1
       IF (J.EQ.(KR+1)) GO TO 102
       GO TO 101
   102 DO 103 N=1,5
   103 KS(N)=0
       DO 112 J=1,5
   111 DO 110 N=1,480
```

```
110 SN(N,J)=BLANK
112 CONTINUE
    READ 100, NC
    NNC=0
  2 READ 1, (A(I),I=1,80)
    NNC=NNC+1
    IF (NNC.EQ.NC) GO TO 72
  3 READ 1, (A(I),I=81,160)
    NNC=NNC+1
    IF (NNC.EQ.NC) GO TO 73
  4 READ 1, (A(I),I=161,240)
    NNC=NNC+1
    IF (NNC.EQ.NC) GO TO 74
  5 READ 1, (A(I),I=241,320)
    NNC=NNC+1
    IF (NNC.EQ.NC) GO TO 75
  6 READ 1, (A(I),I=321,400)
    NNC=NNC+1
    IF (NNC.EQ.NC) GO TO 76
  7 READ 1, (A(I),I=401,480)
    NNC=NNC+1
    GO TO 33
 72 NI=81
    GO TO 77
 73 NI=161
    GO TO 77
```

```
74 NI=241
   GO TO 77
75 NI=321
   GO TO 77
76 NI=401
77 DO 78 I=NI,480
78 A(I)=BLANK
33 KSN=0
   DO 57 I=1,480
   IF (A(I).EQ.S(1)) GO TO 70
57 CONTINUE
   GO TO 60
70 K=0
   I=1
90 IF (A(KA+1).NE.BLANK) GO TO 8
   KA=KA+1
   GO TO 90
 8 IF (A(I).EQ.BLANK) GO TO 9
   I=I+1
   GO TO 8
 9 IF ((I-K).NE.1) GO TO 10
   K=I
   I=I+1
   GO TO 8
10 J=1
   KK=K
```

```
   II=I
12 IF (A(KK+1).NE.S(J)) GO TO 13
   KK=KK+1
13 IF (A(II-1).NE.S(J)) GO TO 15
   II=II-1
15 IF (J.EQ.11) GO TO 16
   J=J+1
   GO TO 12
16 L=II-KK-1
   J=1
17 IF (L.NE.LW(J)) GO TO 18
   M=NS(J)
20 KKK=KK
21 IF (A(KKK+1).NE.WD(M)) GO TO 18
   IF (KKK.EQ.(KK+L-1)) GO TO 23
   KKK=KKK+1
   M=M+1
   GO TO 21
23 KSN=KSN+NFQ(J)
   GO TO 19
18 J=J+1
   IF (J.EQ.(KR+1)) GO TO 19
   GO TO 17
19 IF (A(I-1).EQ.S(1)) GO TO 24
   K=I
   I=I+1
```

```
      GO TO 8
   24 N=1
      DO 25 NN=2,5
      IF (KS(NN).GE.KS(N)) GO TO 25
      N=NN
   25 CONTINUE
      LSN=(I-1-KA)
      IF (KSN.LE.KS(N)) GO TO 226
      KS(N)=KSN
      DO 27 J=1,LSN
      M=KA+J
      SN(J,N)=A(M)
   27 A(M)=BLANK
      LLN=LSN+1
      DO 28 J=LLN,480
   28 SN(J,N)=BLANK
      GO TO 26
  226 DO 227 J=1,LSN
      M=KA+J
  227 A(M)=BLANK
   26 DO 29 NN=1,6
      IF ((NN*80).GE.II) GO TO 30
   29 CONTINUE
   30 IF (NN.NE.1) GO TO 32
      KA=II+1
      GO TO 33
```

```
32 KA=II-((NN-1)*80)+1
   NJ=7-NN
   NNJ=NJ*80
   DO 31 N=1,NNJ
   M=(N+(NN-1)*80)
31 A(N)=A(M)
   NNJ1=NNJ+1
   DO 86 I=NNJ1,480
86 A(I)=BLANK
   NNN=NN-1
   IF (NNC.NE.NC) GO TO 85
   DO 80 IN=1,480
   IF (A(IN).NE.BLANK) GO TO 33
80 CONTINUE
   GO TO 40
85 GO TO (7,6,5,4,3),NNN
40 PRINT 43
43 FORMAT (1X,21HFREQ   W.LG.   KEYWORD//)
   J=1
41 M=NS(J)
   N=NF(J)
   PRINT 42,(NFQ(J),LW(J), (WD(I),I=M,N))
42 FORMAT (1X,I4,3X,I3,3X,70A1)
   J=J+1
   IF (J.EQ.(KR+1)) GO TO 50
   GO TO 41
```

```
50 PRINT 51
51 FORMAT (//,1X,13HKEY SENTENCES//)
   N=1
53 PRINT 62,(SN(M,N),M=1,480)
   PRINT 54,(KS(N))
54 FORMAT (1X,I4//)
   N=N+1
   IF (N.EQ.6) GO TO 55
   GO TO 53
60 PRINT 61
61 FORMAT (1X,20HSENTENCE IS TOO LONG)
   PRINT 62,(A(M),M=1,480)
62 FORMAT (1X,80A1)
55 STOP
   END
```

APPENDIX 3 EXAMPLE RUN OF AUTO-ABSTRACT PROGRAM

INPUT ARTICLE- NOISE AND VIBRATION

ENGINEERING PROBLEMS CONNECTED WITH SHOCK AND VIBRATION ARE BETTER UNDERSTOOD TODAY THAN THEY WERE A DECADE AGO, BUT A NUMBER OF CURRENT PRACTICES AND APPROACHES CAN STAND IMPROVEMENT. SHOCK RATINGS OF PARTS, FOR EXAMPLE, ARE STILL EXPRESSED IN TERMS OF GS. THESE SIMPLE RATINGS, OFTEN BASED ON ACCELERATION LEVELS OF UNSPECIFIED SHOCK PULSES, DO MUCH TO CONFUSE FUNDAMENTAL INSIGHT INTO SHOCK-INDUCED FAILURE.

REFERENCE TO A STANDARD PULSE SHAPE WOULD BE BETTER. TALKING IN TERMS OF A SPECTRUM OVER A FREQUENCY RANGE WOULD BE BETTER STILL- EVEN IF THE SPECTRUM HAS TO BE INFERRED SOMEWHAT CRUDELY FROM PULSE-SHAPE DATA. THE LATTER APPROACH WOULD BE ESPECIALLY BENEFICIAL IF IT GAVE SOME IDEA OF THOSE FREQUENCY BANDS IN WHICH AN ITEM IS PARTICULARLY SENSITIVE BY VIRTUE OF ITS RESONANCES.

A RANDOM VIBRATION IS MORE COMPLETELY DESCRIBED WHEN PROBABILITY DISTRIBUTIONS ARE CONSIDERED ALONG WITH POWER SPECTRA. DEVIATIONS FROM GAUSSIAN DISTRIBUTION HAVE CREATED SOME INTEREST, BUT TOO MUCH OF THIS HAS BEEN CONCERNED WITH MOUNTING OR EXCITATION POINTS. TOO LITTLE ATTENTION IS FOCUSED ON GAUSSIAN DEVIATIONS AT THE FAILURE POINT CAUSED BY NON-LINEARITIES ASSOCIATED WITH THE FAILURE MECHANISM. A NON-GAUSSIAN DISTRIBUTION IS A DELICATE THING, READILY ALTERED BY FILTERING AND PHASE SHIFT, SO THAT DEVIATIONS AT EXCITATION POINTS ARE PROBABLY LESS SIGNIFICANT THAN IRREGULARITIES OF SPECTRUM WHICH, FOR PRACTICAL REASONS, MUST BE IGNORED.

THEORETICAL TREATMENTS HAVE CONCENTRATED ON EFFECTS OF GENTLE NON-LINEARITIES. THESE ARE WORTHWHILE ADVANCES, BUT MORE CRITICAL PROBLEMS ARE ASSOCIATED WITH ABRUPT NON-LINEARITIES AT THE EXTREME TAILS OF A DISTRIBUTION.

MORE PERSPECTIVE SHOULD BE USED IN THE REDUCTION OF DATA, AND PERSPECTIVE SHOULD BE GUIDED MORE BY EXPECTED UTILITY. METICULOUS DATA REDUCTION IS BEING CARRIED OUT ON LOW-LEVEL VIBRATIONS WITH OVERALL RMS ACCELERATIONS LESS THAN 1G. ESPECIALLY IN THIS PERIOD OF COST SENSITIVITY, THERE ARE DANGERS IN TOO MUCH REDUCED DATA, EXTREMES OF IT DIVERT ATTENTION FROM THE MORE IMPORTANT PROBLEMS AND MAY SERVE ONLY AS PUBLISHED EVIDENCE THAT EFFORT HAS BEEN EXPENDED WITHOUT EXPECTATION OF BENEFICIAL EFFECT ON A PROGRAM.

RECENTLY, ENGINEERS HAVE BECOME MORE AWARE OF NOISE AS WELL AS VIBRATION. THERE HAS BEEN TOO MUCH TENDENCY TO TREAT THEM AS INDEPENDENT ENVIRONMENTS AND SIMULATE BOTH AS A ROUTINE MATTER OF POLICY. CRUDELY SPEAKING, IN MISSILES AND SPACE VEHICLES, NOISE AND VIBRATION ARE PROPORTIONAL. THE CONCLUSION THAT BOTH MUST BE SIMULATED IS NOT NECESSARILY VALID AND SHOULD BE APPROACHED WITH CAUTION.

THE STATE OF THE ART OF VIBRATION TESTING IS STILL SUCH THAT A CHANGE OF TEST FIXTURE WILL OFTEN RESULT IN SIGNIFICANTLY DIFFERENT FAILURES. BUT IT DOES NOT NECESSARILY FOLLOW THAT BOTH FIXTURES MUST BE USED. ANY DIFFERENCE IN FAILURE FROM NOISE VS VIBRATION EXCITATION MAY BE FROM PROBLEMS ANALOGOUS TO THAT OF A DIFFERENCE IN FIXTURES. APART FROM THE SUBTLETIES OF TEST APPARATUS, THERE ARE INHERENT UNCERTAINTIES IN THE REALISM OF SIMULATION OF

ENVIRONMENT AT THE PARTS LEVEL. DISTINCTIONS ARE IMPORTANT BETWEEN WHAT IS SCIENTIFICALLY PROVED AND WHAT IS CONVENTION OR STANDARDIZATION. WAYS IN WHICH TENSILE STRENGTHS AND FATIGUE LIMITS ARE MEASURED ARE CONVENTION, NOT A MATTER OF PROCEDURES THAT HAVE BEEN SCIENTIFICALLY PROVED TO BE BEST.

THE PIONEERS WHO ESTABLISHED THE PROCEDURES DID SOMETHING THAT WAS NECESSARY BEFORE ANY EFFECTIVE APPROACH TO DESIGN COULD BE POSSIBLE AND BEFORE ANY FURTHER PROGRESS COULD BE MADE. THEY DID WHAT ALL GOOD ENGINEERS SHOULD BE ABLE TO DO ON OCCASION– USE THEIR BEST JUDGMENT AND GET ON WITH THE JOB. THE FUNDAMENTALS OF FATIGUE AND FAILURE ARE UNDER CONTINUING STUDY. THE CURRENT DESIGN PROCEDURES WILL, HOWEVER, CONTINUE TO BE USED, PERHAPS EVENTUALLY WITH SLIGHT MODIFICATION.

PRESENT USE OF SPECIFICATIONS IS A CONVENTION WHICH HAS BEEN FOUND EMPIRICALLY TO BE USEFUL. THE WAY IN WHICH SPECIFICATIONS ARE USED IS CONVENTION. THE IDEA THAT THE TEST EXCITATION SHOULD SIMULATE THE ENVIRONMENT IS CONVENTION.

IT MAY SEEM OBVIOUS THAT SIMULATION IS A NECESSARY OBJECTIVE, BUT THE OBVIOUS SHOULD NOT BE TAKEN AS ENTIRELY AXIOMATIC. GREAT SCIENTIFIC ADVANCES HAVE COME FROM QUESTIONING THE OBVIOUS, AND SCIENCE HAS REPEATEDLY DISCLOSED AREAS IN WHICH THE OBVIOUS IS NOT VALID. SPECIFICATIONS SERVE ULTIMATELY AS A MOTIVATION TO PEOPLE WHO INFLUENCE DESIGN. THEIR PROBABLE EFFECT ON DESIGN IS POTENTIALLY AS FRUITFUL A FIELD FOR STUDY AS THE PROBLEM OF SIMULATION.

IN THE FUTURE, ADVANCES SHOULD BE MORE SIGNIFICANT THAN THEY HAVE BEEN IN THE PAST DECADE. THIS WILL BE EASIER IF PROBLEM AREAS THAT NEED MORE

ATTENTION ARE RECOGNIZED.

SHOCK AND VIBRATION ENGINEERS HAVE GENERATED AN EXTENSIVE LORE OF QUALITATIVE DESIGN CONSIDERATION, BUT THIS HAS NOT BEEN SATISFACTORILY GATHERED IN ONE PLACE. SUCH AN EFFORT WOULD HELP AVOID THE MORE OBVIOUS DESIGN ERRORS.

THE INTERACTION OF SPECIFICATIONS WITH EQUIPMENT DYNAMICS INDIRECTLY BEING CONTROLLED DESERVES MORE INVESTIGATION. INSTRUMENTATION FOR DYNAMICAL ANALYSIS OF EQUIPMENT IS NEEDED. THE BULK OF THE ACCELEROMETERS MANUFACTURED ARE NOT DESIGNED FOR INVESTIGATION OF DYNAMICS INSIDE EQUIPMENT. THIS DEFICIENCY OF INSTRUMENTATION AND ASSOCIATED TECHNIQUES IS PROBABLY THE GREATEST IMMEDIATE OBSTACLE TO PROGRESS.

ANOTHER NEED IS FOR A TABULATION OF RESONANCE FREQUENCIES OF STANDARD PARTS. IN DESIGN FOR RELIABILITY UNDER VIBRATION, MORE INFORMATION SHOULD BE DEMANDED THAN JUST THE SIMPLE STANDARD TEST CONDITION SURVIVED. DATA SHOULD BE ACCUMULATED ON TYPICAL EQUIPMENT STRUCTURES.

A MORE COMPREHENSIVE DISPLAY OF OBJECTIVES IS GIVEN IN TABLE 1. THEY ARE NOT MUTUALLY INDEPENDENT, AND ONE MUST MAKE TRADEOFFS BETWEEN THE DIFFERENT OBJECTIVES. SHOCK AND VIBRATION ENGINEERING IS AS MUCH A MATTER OF MANAGEMENT-TYPE TECHNIQUES AS IT IS OF CLASSICAL TECHNICAL DISCIPLINES. IN TIME, OPERATIONS RESEARCH MAY MAKE ITS CONTRIBUTION IN CONJUNCTION WITH THOSE OF THE MORE PURELY TECHNICAL DISCIPLINES THAT ARE APPLIED NOW.

OUTPUT- FROM PHASE II

FREQ	W.LG.	KEYWORD	FREQ	W.LG.	KEYWORD
2	11	ENGINEERING	3	9	ENGINEERS
4	8	PROBLEMS	3	5	NOISE
5	5	SHOCK	2	11	INDEPENDENT
9	9	VIBRATION	2	8	SIMULATE
3	6	BETTER	3	4	BOTH
2	6	DECADE	3	6	MATTER
2	7	CURRENT	2	11	NECESSARILY
2	7	RATINGS	2	5	VALID
3	5	PARTS	2	4	SUCH
3	5	STILL	4	4	TEST
2	5	TERMS	2	9	DIFFERENT
2	6	SIMPLE	2	8	FIXTURES
2	5	OFTEN	2	10	DIFFERENCE
5	4	MUCH	3	10	SIMULATION
5	7	FAILURE	2	11	ENVIRONMENT
3	8	STANDARD	2	7	BETWEEN
3	8	SPECTRUM	2	14	SCIENTIFICALLY
2	9	FREQUENCY	2	6	PROVED
2	7	CRUDELY	5	10	CONVENTION
5	4	DATA	2	7	FATIGUE
2	8	APPROACH	3	10	PROCEDURES
2	10	ESPECIALLY	2	4	BEST
2	10	BENEFICIAL	2	3	WHO
2	4	SOME	2	3	DID
2	4	IDEA	2	9	NECESSARY
3	10	DEVIATIONS	2	6	BEFORE
2	8	GAUSSIAN	7	6	DESIGN
3	12	DISTRIBUTION	2	8	PROGRESS
7	4	BEEN	2	3	USE
4	10	EXCITATION	2	5	THEIR
2	6	POINTS	2	5	UNDER
3	9	ATTENTION	2	5	STUDY
3	15	NON-LINEARITIES	4	14	SPECIFICATIONS
3	10	ASSOCIATED	5	7	OBVIOUS
2	8	PROBABLY	2	5	AREAS
2	4	LESS	2	7	PROBLEM
2	11	SIGNIFICANT	2	4	NEED
3	8	ADVANCES	2	3	ONE
2	11	PERSPECTIVE	4	9	EQUIPMENT
9	6	SHOULD	2	8	DYNAMICS
4	4	USED	2	13	INVESTIGATION
2	9	REDUCTION	2	15	INSTRUMENTATION
2	5	BEING	2	10	TECHNIQUES
2	9	IMPORTANT	2	10	OBJECTIVES
2	5	SERVE	2	4	MAKE
2	6	EFFORT	2	9	TECHNICAL
2	6	EFFECT	2	11	DISCIPLINES

OUTPUT (CONTINUED)

KEY SENTENCES

SHOCK AND VIBRATION ENGINEERS HAVE GENERATED AN EXTENSIVE LORE OF QUALITATIVE DESIGN CONSIDERATION, BUT THIS HAS NOT BEEN SATISFACTORILY GATHERED IN ONE PLACE.
33

ANY DIFFERENCE IN FAILURE FROM NOISE VS VIBRATION EXCITATION MAY BE FROM PROBLEMS ANALOGOUS TO THAT OF A DIFFERENCE IN FIXTURES.
31

MORE PERSPECTIVE SHOULD BE USED IN THE REDUCTION OF DATA, AND PERSPECTIVE SHOULD BE GUIDED MORE BY EXPECTED UTILITY.
33

IN DESIGN FOR RELIABILITY UNDER VIBRATION, MORE INFORMATION SHOULD BE DEMANDED THAN JUST THE SIMPLE STANDARD TEST CONDITION SURVIVED.
36

ESPECIALLY IN THIS PERIOD OF COST SENSITIVITY, THERE ARE DANGERS IN TOO MUCH REDUCED DATA, EXTREMES OF IT DIVERT ATTENTION FROM THE MORE IMPORTANT PROBLEMS AND MAY SERVE ONLY AS PUBLISHED EVIDENCE THAT EFFORT HAS BEEN EXPENDED WITHOUT EXPECTATION OF BENEFICIAL EFFECT ON A PROGRAM.
36

BIBLIOGRAPHY

A-1. JACKSON, LUCILLE (EDITOR) –"TECHNICAL LIBRARIES, THEIR ORGANIZATION AND MANAGEMENT", SPECIAL LIBRARIES ASSOCIATION, 31 EAST 10TH ST., NEW YORK 3, N. Y., 202 PP., 1951.

A-2. DOSS, M. P. (EDITOR) –"INFORMATION PROCESSING EQUIPMENT", REINHOLD, 1955.

A-3. LEWIS, C. M. AND W. H. OFFENHAUSER JR. –"MICRORECORDING – INDUSTRIAL AND LIBRARY APPLICATIONS", INTERSCIENCE, 1959, 468PP.

A-4. PERRY, J. W., A. KENT AND M. M. BERRY –"MACHINE LITERATURE SEARCHING", VOL. 1, INTERSCIENCE, 1956, 175 PP.

A-5. PERRY, J. W. AND A. KENT –"DOCUMENTATION AND INFORMATION RETRIEVAL– AN INTRODUCTION TO BASIC PRINCIPLES AND COST ANALYSIS", INTERSCIENCE, 1957.

A-6. COBLANS, H –"NEW METHODS AND TECHNIQUES FOR THE COMMUNICATION OF KNOWLEDGE", UNESCO BULLETIN FOR LIBRARIES 11; 154–75, JULY, 1957.

A-7. SHERA, J. H., A. KENT AND J. W. PERRY –"INFORMATION SYSTEMS IN DOCUMENTATION", INTERSCIENCE, 1957, 659 PP.

A-8. GUNNING, MRS. R. T. –"AUTOMATION IN THE LIBRARY", SPECIAL LIBRARIES 48; 332–5, SEPTEMBER, 1957.

A-9. LIBRARY ASSOCIATION RECORD –"INFORMATION STUDY CONFERENCE ON CLASSIFICATION FOR INFORMATION RETRIEVAL, 1957", DORKING, ENGLAND, (SUMMER REPORT), LIBRARY ASSOCIATION RECORD 59; 304–6, SEPTEMBER, 1957.

A-10. RAKOV, B. M. AND V. P. CERENIN –"MACHINES FOR RETRIEVING INFORMATION IN THE U. S. S. R.", UNESCO BULLETIN FOR LIBRARIES, 11; 192–7, SEPTEMBER, 1957.

A-11. WARHEIT, I. A. –"MACHINES AND SYSTEMS FOR THE MODERN LIBRARY", SPECIAL LIBRARIES 48; 357–63, OCTOBER, 1957.

A-12. CASEY, R. S., J. W. PERRY, A. KENT AND M. BERRY –"PUNCHED CARDS", 2ND ED., REINHOLD, 1958, 707 PP.

A-13. COILE, R. AND B. FOSTER –"USE OF COMPUTERS FOR MECHANIZED LITERATURE SEARCHING IN OPERATIONS-RESEARCH LIBRARIES", OP. RESEARCH, 6; 434–8, MAY 1958

A-14. SHERA, J. H., A. KENT AND J. W. PERRY - "INFORMATION RESOURCES, A CHALLENGE TO AMERICAN SCIENCE AND INDUSTRY", INTERSCIENCE, 1958, 230 PP.

A-15. KENT, A., AND J. W. PERRY - "CENTRALIZED INFORMATION SERVICES - OPPORTUNITIES AND PROBLEMS", INTERSCIENCE, 1958, 164 PP.

A-16. PERRY, J. W. AND A. KENT - "TOOLS FOR MACHINE LITERATURE SEARCHING - SEMANTIC CODE DICTIONARY - EQUIPMENT - PROCEDURE", INTERSCIENCE, 1958, 992 PP.

A-17. BOURNE, C. P. - "BIBLIOGRAPHY ON THE MECHANIZATION OF INFORMATION RETRIEVAL", STANFORD RESEARCH INSTITUTE, DIV. OF ENG. RESEARCH, MESLO PARK, CALIFORNIA, 1958, 22 PP.

A-18. FADEN, B. R. - "INFORMATION RETRIEVAL", SPECIAL LIBRARIES., 50; 162-5, APRIL, 1959, 50; 197-200 MAY, 1959, 50; 244-9 JULY, 1959, 50; 392-7 OCTOBER, 1959.

A-19. HYSLOP, M. R. - "ANSWERING METALLURGICAL QUESTIONS BY MACHINE - A PROGRESS REPORT", SPECIAL LIBRARIES., 50; 435-42, NOVEMBER, 1959.

A-20. VICKERY, B. C. - "MACHINES AND INDEXES", BIBLIOGRAPHY, UNESCO BULLETIN FOR LIBRARIES, 13; 249-52, NOVEMBER, 1959.

A-21. WORSLEY, P. K. AND OTHERS - "A STUDY OF THE FUNDAMENTALS OF INFORMATION STORAGE AND RETRIEVAL", PREPARED FOR OFFICE OF NAVAL RESEARCH, BEHNSON-LEHNER CORP., SANTA MONICA, CALIFORNIA, 1959, 97 PP.

A-22. U. S. LIBRARY OF CONGRESS, SCIENCE AND TECHNOLOGY DIVISION - " A LIST OF SCIENTIFIC AND TECHNICAL SERIALS CURRENTLY RECEIVED BY THE LIBRARY OF CONGRESS", LIBRARY OF CONGRESS, 1960.

A-23. ENGINEERING SOCIETIES LIBRARY - "BIBLIOGRAPHY ON FILING, CLASSIFICATION AND INDEXING SYSTEMS FOR ENGINEERING OFFICES AND LIBRARIES", ENGINEERING SOCIETIES LIBRARY, 1960.

A-24. FISHER, E. L. - "A CHECKLIST FOR THE ORGANIZATION, OPERATION AND EVALUATION OF A COMPANY LIBRARY", SPECIAL LIBRARIES, N. Y., 1960, 60 PP.

A-25. KENT A. (EDITOR) - "INFORMATION RETRIEVAL AND MACHINE TRANSLATION", INTERSCIENCE, 1960.

A-26. FROME, J. - "SEMI-AUTOMATIC INDEXING AND ENCODING", JOURNAL OF THE FRANKLIN INSTITUTE, V. 270, N. 1, JULY 1960, PP. 3-11, 26.

A-27. BAR-HILLEL - "SOME THEORETICAL ASPECTS OF THE MECHANIZATION OF LITERATURE SEARCHING", OFFICE OF TECH. SERV, WASHINGTON, 1960, 74 PP., (PB161547)

A-28. MARON, M. E. -"AUTOMATIC INDEXING - AN EXPERIMENTAL ENQUIRY", RAND CORP., SANTA MONICA, CALIFORNIA, 1960, 35 PP. (RAND CORP., U. S. A. F. PROJECT RAND RESEARCH MEMORANDUM RM-2601).

A-29. ANDREWS, D. D. -"STORING NUMERICAL VALUES IN INVERTED FILE FOR MECHANIZED INFORMATION RETRIEVAL", JOURNAL OF FRANKLIN INSTITUTE, 270; 34-41, JULY, 1960.

A-30. GARDIN, J. C. -"DOCUMENT ANALYSIS AND INFORMATION RETRIEVAL", UNESCO BULLETIN FOR LIBRARIES, 14; 2-5, JANUARY, 1960.

A-31. KENT A. -"MACHINE LITERATURE SEARCHING IN SCIENCE", JOURNAL OF THE FRANKLIN INSTITUTE, 270; 42-50, JULY, 1960.

A-32. MAIZELL, R. E. -"INTERNATIONAL CONFERENCE FOR STANDARDS ON A COMMON LANGUAGE FOR MACHINE SEARCHING AND TRANSLATION", COLLEGE AND RESEARCH LIBRARIES, 21; 125-6, MARCH, 1960.

A-33. MELCHER, D. -"PRIMER IN MACHINE INFORMATION STORAGE AND RETRIEVAL", THE LIBRARY JOURNAL, 85; 909-12, MARCH 1, 1960.

A-34. SCHULTZ, C. K. AND J. J. O'CONNOR -"DESIGNING MORE EFFICIENT INDEXES", UNESCO BULLETIN FOR LIBRARIES 14; 160-3, JULY 1960.

A-35. WARHEIT, I. A. -"PRINCIPLES OF COMPUTER OPERATION", SPECIAL LIBRARIES, 51; 485-92, NOVEMBER, 1960.

A-36. AVCO CORP. -"TECHNICAL INVESTIGATION OF ELEMENTS OF A MECHANICAL LIBRARY SYSTEM", FINAL REPORT NO. EW-6680, JAN. 11, 1960, AVCO CORP. CROSLEY DIV., ELECTRONICS RESEARCH LABORATORIES, BOSTON, 1960, 110 PP.

A-37. CASE INSTITUTE OF TECHNOLOGY, OPERATIONS RESEARCH GROUP -"AN OPERATIONS RESEARCH STUDY OF THE DISSEMINATION AND USE OF RECORDED SCIENTIFIC INFORMATION", U. S. OFFICE OF TECH. SERVICES, 1960, 63 PP.

A-38. COUGHLAN, R. E. -"DOCUMENTATION FOR SCIENCE AND ENGINEERING", BURLINGTON PUBLISHING CO., BURLINGTON, CONN., 1960, 177 PP.

A-39. MACHINE DESIGN -"PEEK-A-BOO CARDS DIAL NBS INFORMATION DRUM", MACHINE DESIGN, JULY 6, 1961, 8 PP.

A-40. ISACSON, S. -"INDUSTRY FACES THE INFORMATION CHALLENGE", ARS PAPER NO. 2200-61, 1961, 4 PP.

A-41. NATIONAL SCIENCE FOUNDATION -"SPECIALIZED SCIENCE INFORMATION SERVICES IN THE UNITED STATES, A DIRECTORY OF SELECTED SPECIALIZED SERVICES IN THE PHYSICAL AND BIOLOGICAL SCIENCES", NAT. SCIENCE FOUNDATION, WASHINGTON, 1961, 528 PP.

A-42. VICKERY, B. C. –"ON RETRIEVAL SYSTEM THEORY", BUTTERWORTH, 1961, 159 PP.

A-43. WESTERN RESERVE UNIVERSITY, CLEVELAND, CENTER FOR DOCUMENTATION AND COMMUNICATION RESEARCH –"DOCUMENTATION AND INFORMATION RETRIEVAL, A SELECTED BIBLIOGRAPHY", WESTERN RESERVE UNIVERSITY, CENTER FOR DOCUMENTATION AND COMMUNICATION RESEARCH, SCHOOL OF LIBRARY SCIENCE, 1961, 8 PP.

A-44. FAIRTHORNE, R. A. –"TOWARD INFORMATION RETRIEVAL", BUTTERWORTHS, 1961, 211 PP.

A-45. BARDEN, B. R. AND B. DENISON –"GUIDE TO THE SLA LOAN COLLECTION OF CLASSIFICATION SCHEMES AND SUBJECT HEADING LISTS OF WESTERN RESERVE UNIVERSITY AS OF MARCH 20, 1961", SPECIAL LIBRARIES ASSOC., N. Y., 1961, 97 PP.

A-46. MELTON, J. S. –"MACHINE INFORMATION RETRIEVAL – BOON OR BUST IN SOLVING THE COMMUNICATION PROBLEM", JOURNAL OF ENGINEERING EDUCATION, APR. 1962, 6 PP.

A-47. BONN, G. S. –"HOW TO USE NATIONAL INFORMATION CENTERS", PAPER NO. 52–MD–7, PROCEEDINGS OF THE 1962 ASME DESIGN ENGINEERING CONFERENCE, AMERICAN INC. OF MECH. ENGINEERS.

A-48. KENT, A. –"TEXTBOOK ON MECHANIZED INFORMATION RETRIEVAL", INTERSCIENCE, 1962, 268 PP.

A-49. "LITERARY MARKET PLACE – THE BUSINESS DIRECTORY OF AMERICAN BOOK PUBLISHING"; LATEST EDITION, R. R. BOWKER CO., NEW YORK.

A-50. BECKER J. AND R. M. HAYES –"INFORMATION STORAGE AND RETRIEVAL – TOOLS, ELEMENTS, THEORIES", WILEY, 1963, 448 PP.

A-51. LUHN, H. P. –"A STATISTICAL APPROACH TO MECHANIZED ENCODING AND ABSTRACTING OF LITERARY INFORMATION", IBM J. OF RES. AND DEV. 1, NO. 4, OCT. 1957, 8 PP.

A-52. LUHN, H. P. –"THE AUTOMATIC CREATION OF LITERATURE ABSTRACTS", IBM J. OF RES. AND DEV. 2, NO. 2, APR. 1958, 6 PP.

A-53. LUHN, H. P. –"POTENTIALITIES OF AUTO-ENCODING OF SCIENTIFIC LITERATURE", RESEARCH REPORT RC–101, MAY 15, 1959, IBM RESEARCH CENTRE, YORKTOWN HEIGHTS N. Y.

A-54. LUHN, H. P. –"KEYWORD IN CONTEXT INDEX FOR TECHNICAL LITERATURE", ASDD REPORT RC–127, AUG. 31, 1959, IBM ADVANCED SYSTEMS DEVELOPMENT DIV., YORKTOWN HEIGHTS, N. Y.

A-55. I. B. M. –"GENERAL INFORMATION MANUAL – STORAGE RETRIEVAL AND DISSEMINATION OF INFORMATION", INTERNATIONAL BUSINESS MACHINES CORP., 1962

A-56. I. B. M. -"REFERENCE MANUAL - INDEX ORGANIZATION FOR INFORMATION RETRIEVAL", INTERNATIONAL BUSINESS MACHINES CORP., 1961.

A-57. HOLM, B. E. -"INFORMATION RETRIEVAL - THE PROBLEM, CO-ORDINATE INDEXING - A SOLUTION", N. Y., ENGINEERS JOINT COUNCIL, 1963.

A-58. MARON, M. E. -"ON CYBERNETICS, INFORMATION PROCESSING AND THINKING", SANTA MONICA, RAND CORP., 1964.