UNITED ARAB REPUBLIC

THE INSTITUTE OF NATIONAL PLANNING

Memo No. 559

Information Retrieval; And Its Potential For Progress Of Research In The U.A.R.

By

Engineer Mohamed A. Mongy
Mrs. Aida M. Elewa
Abdel Hamid M. Dawash
Mahmoud I. Hendi

Supervised by
Dr. Ahmad Badr

Operations Research Center
April, 1965

Information Retrieval

And Its potential for Progress of Research In the U.A.R.

By

Engineer Mohamed A. Mongy

Mrs. Aida M. Elewa

Operations Research Center

and

Abdel Hamid M. Dawash

Mahmoud I. Hendi

National Information and Documentation Center

Programs prepared by: Mrs. Mary N. Youssef

Collator work by : Ramadan Abdel Mouti

Supervised by

Dr. Ahmad Badr

National Information and Documentation Center

March, 1965.

Acknowledgements

The group wishes to express his appreciation and gratitude to Dr. Salah Hamed , Director of the Operations Research Center and to Dr. Ahmed Kabesh, Director of the National Information and Documentation Center for their initiation . in calling for this joint study between the two institutions as well as for their enthusiasm and continued encouragement of the group.

Table of Contents

Pages

Introduction

This is the first report prepared jointly between the Operations Research Center, Institute of National Planning, and the National Information and Documentation Center, Ministry of Scientific Research, as a result of the effort carried out by the group.

The purpose of this study has two dimensions. The first is to acquaint and train a group of Scientists in the field of documentation and information retrieval about the present state of information and documentation techniques from the theoretical points of view as well as from the equally important applications, examplified in this first case with the equipment available at the Operations Research Center.

The second dimension is to stimulate a deeper consciousness among governmental, academic and industrial institutions about potentialities of information, information retrieval techniques and their important role in the development process and particularly in the progress of science and technology which is the bulwark of the national economy.

The joint cooperation between both institutions in the field of information retrieval is a pioneering and rewarding experience. This example should be continued between the two institutions and also should be followed among other institutions interested in the information retrieval field.

Ahmed Bedr

Chapter I

The Need For Information Retrieval

1.1. The Value of Information:

The value of information as an essential prerequisite for progress and development has become an undisputed fact in both advanced and developing countries. If countries have realized early the value of education in the process of development they have now mobilized a great part of their efforts towards the maximum accessibility of information because of its great role in developing their economies based principally on science and technology.

The extent to which the research worker depends upon the work of others has been clearly stated by one of the greatest of all scientists, the atomic physicist, Ernest Rutherford. As quoted by James Newmanin a recent issue of the "Scientific American", Lord Rutherford said:

"I have also tried to show you that it is not in the nature of things for any one man to make a sudden violent discovery; science goes step by step, and every man depends on the work of his predecessors. When you hear of a sudden unexpected discovery a bolt from the blue as it were - you can always be sure that it has grown up by the influence of man on another, and it is his mutual influence which makes the enormous possibility of scientific advance.

Scientists are not dependent on the ideas of a single man, but on the combined wisdom of thousands of men, all thinking the same problem, and each doing his little bit to add to the great structure of knowledge which is gradually being erected"(1).

As a commodity, information has two characteristics which can give it monitary value:

i. Relevence.

ii. Timeliness.

If either of these characteristics is missing, the information

(1) U.S., Congress, Senate, Committee on Government Operation, Science Program, Report No. 120, 86TH Congress, First Sess., 1959, p. 120.

is useless.

Information centers are operated in order to provide information to those who need it promptly and comprehensively from the world - wide literature.

The uses of such information are:-

a. It makes possible the completion of some projects which otherwise would have to be delayed or abandoned for lack of knowhow.

b. It saves valuable time for research and other decision - making personnel.

c. It helps develop knowledgeable and effective management at all levels of an organization.

d. It avoids undesireable duplication of research or other efforts and thus saves money which can be alloted to other productive purposes.

1.2. The Information "Problem" and the Information "Explosion".

Nowadays published and unpublished literature is increasing in quantity and complexity as a consequence of the expansion of human intellectual activities which resulted in a flood of literature and recorded knowledge unmatched in previous history.

It is estimated that about 55,000 periodicals are published annually containing about 1,200,000 articles of potential interest. Published literature includes also about 60,000 books as well as about 100,000 research reports and pamphlets..etc. Within this vast body of world wide research, published and unpublished, in different languages lie the information that research workers need to perform their work.(1)

More information has been produced than can be stored and retrieved. At the same time the need for effective exploitation of the literature is increasing because of the increasing

(١) أحمد بدر ، التعاون الدولى فى ميدان التوثيق العلمى ، مجلة المكتبة العربية : المجلد ١ عدد ١ ، ص ٥٠ - ٥١ ، سنة ١٩٦٣

pressures for avoidance of repetition and duplication of research. This is a common problem in both developed and less developed countries. The problem may be more acute in the developing nations because they cannot afford to duplicate already done research.

Abilities of individual libraries, documentation centers & organizations to cope with this flood of literature are decreasing. The traditional library methods in storage & retrieval of this information are no longer adequate to respond to the complex demands and information requests of engineers, scientists, managers, decision makers... etc..

Thus the trend and the need to develop traditional library & documentation techniques and to use mechanical methods & devices is pressing in order to collect, store, organize, dessiminate and retrieve the information.

1.3. Traditional versus Non-Conventional Information Retrieval Systems:

Traditional library methods and techniques in analysis, organization and storage of information are valid only to meet limited requirements of research - workers in location and identification of needed information. But the complex and diversified needs and requests of research-workes nowadays have proven to be only met by advanced methods and techniques in which both the machine and the humanbeing constitute a functional information retrieval system.

It has to be asserted that the long developed and adopted library techniques & procedures especially cataloguing, classification, circulation as well as reference services constitute the basis from which different information retrieval systems have emanated as evolutionary (and may be revolutionary) to respond to the previous mentioned complex & diversified needs of research workers.

The greater degree of depth of analysis of information contained in documents is one of the main characteristics of non-conventional systems and which make them superior to traditional ones.

In a comparison between traditional and non-conventional systems, it can be said that analysis of information occurs by the following three methods:

Classification - Abstracting - Indexing.

1.3.1 Classification :

It is a systematic logical arrangement of index enteries usually in a hierarchical or tree pattern. The standard library classification systems, such as Dewey, Decimal, Bliss, Cutter, Library of Congress and Universal Decimal, all try to be hierarchical systems. The terms are arranged so that they proceed from the most general to the most specific.(1)

The traditional library classification (e.g. Dewey Decimal Classification or Library of Congress) is a rigid classification which is often called a "pigeonhole" classification and it involves the characterization of each record from a single point of view. When a record is to be stored, and only a single copy is available, a pigeonhole, or a single physical location, must be provided for this record.

Most classifications are artificial or synthetic and when these classifications are developed still represent, individually, a single rigid approach to a subject. This does not coincide with the needs and viewpoint of the searcher because records which contain information are generally multidimensional in nature. Hence the physical classification of records often tends to assume the characteristics of a rigid classification.

Information retrieval systems stress the utilization of multidimensional approach in classification in order to respond to the multidimentional requests of researchers and also in

(1) IBM, Reference Manual, Index Organization for Information Retrieval, p.9.

order to avoid the problems of the expansion of notation of classification.

1.3.2. Abstracting:

An abstract is a summary of a publication or article accompanied by an adequate bibliographical description to enable the research worker to trace the publication or the article. Three types of abstracts can be identified:-

a. Traditional abstract: which may be either descriptive or informative. The former kind embodies a general statement of the scope of a document but this abstract does not substitute reading the original document. The informative abstract is written in order to provide a concise but comprehensive summary of the significant items of the document and this kind of abstracts may serve as a substitute for reading the original document.

b. Extract: and this is analogous to an abstract in that it represents what is considered by an analyst to be the important subject matter of a graphic record and this extract is selected from the original words of the documents.

Extracts may be selected by human analysts, or by the application of machine techniques to produce the "auto-abstract"(1)

The techniques used by humans to prepare extracts are subjective, and involve the exercise of judgement by an analyst in order to determine which portion of a document is of sufficient potential significance to warrant recording. When a machine is used for extracting, the entire text of a record is converted to machine - readable form.

---

(1) H.P. Luhn, "The automatic creation of Literature. Abstracts". IBM Journal of Research and Development, 2, No. 2 (April 1959), pp. 159 - 165.

It is then scanned by a digital computer. It is assumed when applying these methods that the frequency and distribution of keywords in the text can be used as the basis for determining the relative significant sentences in the text. Following this assumption, the sentences that are highest in "significance" (as determined by their high content of keywords) are printed out to produce an extract or "auto - abstract".

c. Telegraphic abstract:(1)

A telegraphic abstract is a detailed index to a graphic record. It is composed of:

(i) Significant words selected from the document,
(ii) Code symbols called role indicators which supply a context for the selected words, and
(iii) Punctuation symbols which separate and group the words and role indicators into various units in somewhat the same fashion as conventional punctuation does.

A file of telegraphic abstracts, though it is not used in the same way, serves the same purpose as a conventional index or card catalog - that is, to locate literature on a given subject.

The finished product of telegraphic abstract is a reel of tape with the information which the abstracters have partly furnished, translated into a computer code ready for searching by machine.

The purpose of the telegraphic abstracts, then is to provide "input" to the machine in a consistent and predictable form so that the machine can be programmed to search for certain predictable forms of information within this input. Details of application of telegraphic abstract is discussed in chapter (3) about the Western Reserve University information retrieval system.

(1) For definition of different types of abstracts see: kent, Allen, Text book on Mechanized Information Retrieval, pp.100-105.

1.3.3. Indexing:

Indexing involves the generation of index terms and the systematic ordering of these terms into an index. An index term is a device used to characterize the information content of a document.

The group of index terms used in the information system constitutes the list or dictionary of index terms.

Indexing may be performed in two different ways. One is to derive appropriate index terms (derived index) directly from incoming documents by lifting certain words or phrases from the text. The second way is to assign appropriate terms to documents from a pre-established index (assigned index).(1)

The form of indexing that is simplest to apply is the one that assumes on the part of the indexer the least amount of subject - matter background and the least amount of technical skill in indexing. It is in this type of indexing that a machine can perform with precision.

Indexing involves the traditional library techniques in subject analysis, namely the use of subject headings, which can be chosen from a subject authority list. In information retrieval systems-both manual and mechanical-the same principle of subject analysis is applied, but more depth of analysis is the main objective of these systems.

If we call the individual terms as subject headings in traditional methods we refer to these terms in nonconventional systems as descriptors, keywords, key-terms, discriminators, identifiers, or uniterms.

The two types of indexing can also be mentioned as follows:

A) Word indexing,

B) Controlled indexing

Word indexing, which involves the selection of words from a document, and their use as index entries. Whereas controlled

(1) IBM, general Information Manual; Storage, Retrieval and Dissemination of Information, p.6.

indexing implies a careful selection of terminology, for storage in the index, in order to avoid as far as possible, the scattering of related subjects under different headings. The control may be imposed by limiting the indexing in (a) the subjects that may be chosen.

(a) the subjects that may be chosen.
(b) the number of aspects that may be chosen,
(c) the language used to express the results of analysis.

Examples of two famous word indexing systems can be mentioned.

(i) Key - Word - In - Context (KWIC) or Permutation index:

This is an auto - encoding system where a computer recognizes significant words (keywords) and these words would be permuted to a predetermined position for alphabetization. When this process is applied to a title of an article or its abstract, all nonsignificant or "common" words are ignored. The following example is an illustration of a KWIC index prepared by IBM company about the titles of articles published on information retrieval and machine translation. (page 9)

(ii) Uniterm Indexing:

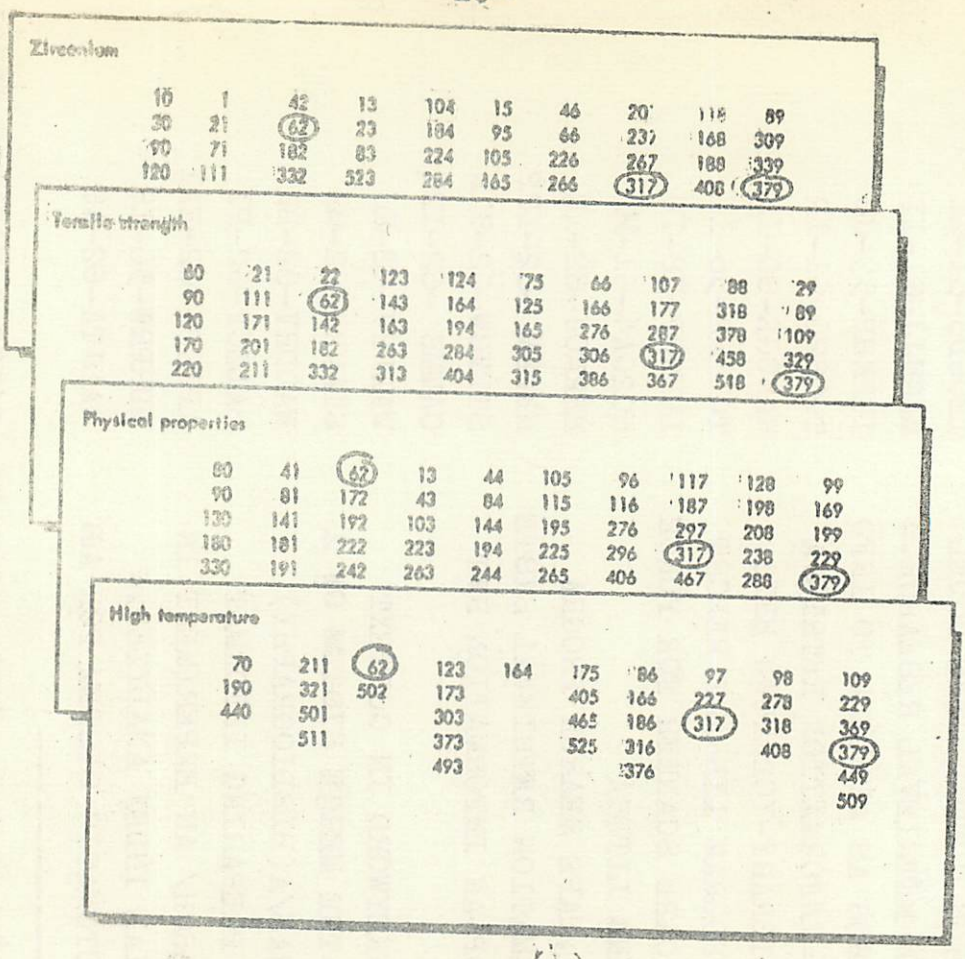
This is an index which involves the analysis of the contents of graphic records in terms of single descriptors called "uniterms" to define a document and to facilitate the manual coordination of these descriptors.

For example, using the following subject headings Zirconium - Physical properties - Tensile strength - High temperature, a card would be prepared for each descriptor used (term card) and the document numbers, referring to the document that contain this information, punched into these cards. When references are wanted covering this complex subject, the appropriate term cards are pulled and matched. All document numbers which appear on all four cards will contain information on the high temperature strength of zirconium. If one is searching for the more general topic of physical properties of Zirconium then a match of two term cards Zirconium and Physical Properties will also retrieve these documents.

KWIC INDEX

ING SYSTEM BASED ON THE	IBM COLLATOR.=AN INFORMATION SEARCH	LANCER-56-IS
SYSTEM PROPOSAL FOR AN	IBM DOCUMENTATION CENTER.=THE MODEL	TE06D6-61-MI
	IBM INFORMATION RETRIEVAL RESEARCH.=	MAGNJJ-61-II
OW SCANNING SYSTEMS FOR	IBM PUNCHED CAROS AS APPLIED TO INFO	IUHNHP-59-RS
AGEMENT CENTER WITH THE	IBM RAMAC 305.=INFORMATION RETRIEVA	T00826-61-IR
HTS.=	IBM RESEARCH LIBRARY-YORKTOWN HEIG	RANOGE-60-IR
=	IBM TYPE 650 PROGRAM LIBRARY REPORT	TOLLA2-58-II
SEARCHING SYSTEMS.=THE	IBM UNIVERSAL CARD SCANNER FOR PUNCH	LUHNHP-58-IU
TURE SEARCHING WITH THE	IBM 101.=MACHINE LITERA	N020A4-61-ML
FORMATION RETRIEVAL FOR	IBM 1401,,ASK,,TAPE SEARCH PROGRA	HERMOJ-62-IR
INVERTED CARD FILE -	IBM 1401 INFORMATION RETRIEVAL SYSTE	NEWKM -62-ICP
IEVAL TECHNIQUE FOR THE	IBM 1401.=SNOOP-AN INFORMATION RET	STEVCW-62-SIR
FORMATION RETRIEVAL.=/	IBM 1620/ IN	CAMMS -62-IRI
(KWIC)INDEXING ON THE	IBM 7090 OPS.=KEYWORD IN CONTEXT	WADDRV-62-KC
AND INDEXING USING THE	IBM 7090 OPS.=THE MERGE SYSTEM OF A	STANRH-62-MS
EXT(KWIC)INDEXER.=/	IBM 7090 PROGRAM/A(BIBLIOGRAPHY)	WADDRV-62-BK
ION FOR KWIC INDEXING (	IBM 7090).=ON PREPARING INFORMAT	BALZCF-62-PI
DOCUMENT RETRIEVAL.=/	IBM-705/ / FREHCH/ AN EXPERIMENT IN	APICOG-62-EAI
THE	IBM 9900 SPECIAL INDEX ANALYZED.=	MURPRW-58-IS
TECHNICAL DOCUMENTS.=	IBM 7090/AN AUTOMATIC SYSTEM FOR AB	GALITA-62-AS

The user is expected to scan down the center of the page to locate a desired index word (in the above example "IBM"), and to note left and right modifying words. If the index word is of interest in its original context then the source document (LANCER-56-IS) is examined as shown in the first line.



Uniterm Cards - Manual matching (1)

The matching of terms to find information is not efficient in manual systems, but lends itself to mechanized and semi - mechanized procedures. The actual visual matching of numbers is a fatiguing process and mechanized systems are performing this matching efficiently as it is applied in the CROSS Index (computer Rearrangement Of Subject Specialities) as shown from the following example taken from the Biological Abstracts (BA).

THE CROSS INDEX:

ARRANGEMENT The subject headings correspond with the titles of the sections and sub - sections of Biological Abstracts. These headings are arranged in the order of their appearances in Biological Abstracts. Under each heading, arranged in ten vertical columns, are the numbers of all those abstracts which

(1) IBM, Reference Manual ....., p.22.

pertain to that heading (BA section or sub - section), those printed under the heading, and those cross - referenced to it. The ten vertical columns are based upon the terminal digits of the abstract number.

USE: The primary use is the determination of all those abstracts which are pertinent to a particular section or sub - section of Biological Abstracts, The second and in many cases the more important use of the CROSS Index is for the determination of those abstracts which have interests common to two or more categories in Biological Abstracts. Abstracts which have common pertinence to two headings may be compared with as many other headings as will be usefull in a progressive narrowing of the search for the exact combination of information attributes wanted.(1)

0	1	2	3	4	5	6	7	8	9
GENERAL BIOLOGY-PHILOSOPHY								44628 44948	44949
	44641	44642		49084					
GENERAL BIOLOGY-TAXONOMY, NOMENCLATURE, TERMINOLOGY									44629
44630 47480		48302		48894					
GENERAL BIOLOGY-EXPLORATIONS, EXPEDITIONS									
	44631	48003							
48990								49018	
GENERAL BIOLOGY-INSTITUTIONS, ADMINISTRATION									
		44632	44633	44634	44635	44636	44637	44638	
					44665			44668	
									44699
									44759

(1) The previous description of the CROSS Index is quoted from the Biological Abstracts Index for 1963.

The following is a summary of the drawbacks of the traditional information systems:

1. Can be considered useful and sufficient only when applied in a small information center.
2. It takes a considerable long time for searching.
3. Multi-searching is not possible at the same time.
4. It cannot go so deep in specifying a certain document, e.g. in a library the main subject headings of a document are only considered.
5. Searching is not so relevant.
6. The results of searching in this system are always unaccountable.
7. The necessity of bringing the vocabularies of both the index and searcher into coincidence.

To conclude we can say that information retrieval systems have increased the capabilities of the librarian, the documentalist & the information scientist to analyse, organize & disseminate information to meet with the multi - dimensional and complex requests of research - workers, scientists and engineers.

Information retrieval systems have adopted many library techniques, but developed them to cope with the present explosion of information and to control the flood of world literature.

Chapter II

Mechanized Information Retrieval Systems

2.1. An approach:

Mechanized information searching systems are those that are characterized by some device, manipulative technique, recording medium, or similar non-conventional feature that distinguishes them from the traditional reference tools of the library.

Although a "mechanized system" usually conjures up the picture of a computer, it is more than a single physical device and it can be considered a system of interacting parts. Some of these parts may be devices, some may be humanbeings, some may be procedures.

2.2. The Choice and Design of an Information Retrieval System:(1)

2.2.1. Objectives:

An information retrieval system can be judged as good only by the use to which it is put. Its objectives may constitute maximization of profits and receipts or fulfillment of an obligation to promote the free flow of information among members of an organization. The objective may be also providing new or attractive services in a competitive situation. Before all, the objective of an information retrieval system is the provision of information or documents, in response to requests with such quality, scope, speed and economy that there is a scientific, economic, military, or other important justification for its establishment & maintenance.

2.2.2. Criteria to be considered:

(i) Clientele: Identification of Clientele or the group of users of the retrieval system is quite important and this identification is relatively feasible. However,

(1) See Kent, Allen,..... p. 197 - 201, 209-17.

after a new system has been developed, it can be expected that the provision of more effective service will attract a larger clientele than the already identified.

(ii) Subject interests and depth of analysis of subject matter:

Surveys of the clientele will provide some guidance in this regard. However, interests may change or develop with the continuous developements and complexity of subject matter and the continuous increase of published and unpublished literature. In any case, subject coverage which can be determined by a delineation of the sources of documents as well as the editorial policy are required.

Degree of detail of analysis of documents has often been described in the literature by specifying an "average" number of descriptors, analytics, clues, or other reference points provided for source documents. This may or not be a valid way of measuring degree of detail of analysis, since the mere proliferation of reference points does not ensure greater or more effective penetration into the subject matter of a document.

(iii) Precision of Service:

The precision of an information retrieval system may be discussed in terms of the amount of documents analysed and subsequentely the number of documents to be provided in response to a question.

The precision of the service is also determined by the pertinency or relevency of documents that may be supplied in response to a question. This requirement is perhaps the most difficult to satisfy, since it has not yet been possible to provide an operational definition

of pertinency upon which both designer and users of a retrieval system agree. Precision of service is finally determined by the speed of providing information to interested users. Speed in this respect constitute promtness with which unit operations of the retrieval system (acquisition of documents, analysis, recording, storage, search, and delivery of results) are carried out.

(iv) Costs

Three types of costs are involved. The first may be the capital costs for acquisition and analysis of records and for purchase of a search mechanism. Secondly, operational costs for conducting the services or for provision of services. And finally, costs involved in not providing suitable information services.

Costs and expenditures of information retrieval system are affected by the following considerations.

- Equipment: When mechanized information retrieval systems, are designed, the desired speed of searching may dictate the choice of equipment. However, [illegible], in many organizations equipment already being used for other purposes may be more or less suited for literature searching. But we should not forget that the information retrieval field is subjected, nowadays to intense research activity and thus it is to be expected that newer tools and equipment would be provided and it is the management decision to select, choose, or substitute existing equipment which may become obsolete.
- Form of Input and Output: The design of a retrieval system must take into account the costs involved in the storage of data or documents in the form of microfilms, punched cards, magnetic tape, paper tape, disks, etc., which itself is determined by the choice of a

particular retrieval system. Again output must recieve due consideration because the customer may require a particular size, shape, or microform of output. Furthermore, if the output is to undergo subsequent processing by the customer, care must be devoted to assuring that it can be processed by equipment available to him, or within his financial means.

- Compatibility with other systems:

The design of a system may be influenced strongly by the need to coordinate with other existing systems or with older systems. The motivation for achieving compatibility is to avoid the expense of reprocessing material, or of using more than one system in conducting a search.

- Personnel Skills:

Availability of personnel trained in information retrieval system design and operation is necessary. Experience is not only needed to "run" the system, but it is also needed to develop the system and acquaintance with continued advances in this field.

(v) Society Response:

It may be a human tradition to resist mechanization of already established manual techniques because of imaginative (or sometimes real) fear of getting rid of personnel. The fact is that machines and mechanized systems will increase human capabilities to control the flood of literature and there would be no dispensing with trained personnel. Only these personnel have to be retrained on the new techniques and new jobs would be created by the introduction of mechanized systems.

Two groups are to be addressed with the introduction of mechanized information retrieval systems:

a) Managerial groups especially in industrial and governmental organizations who are responsible for taking decisions with respect to application of mechanized systems in order to maximalize profits or gains.

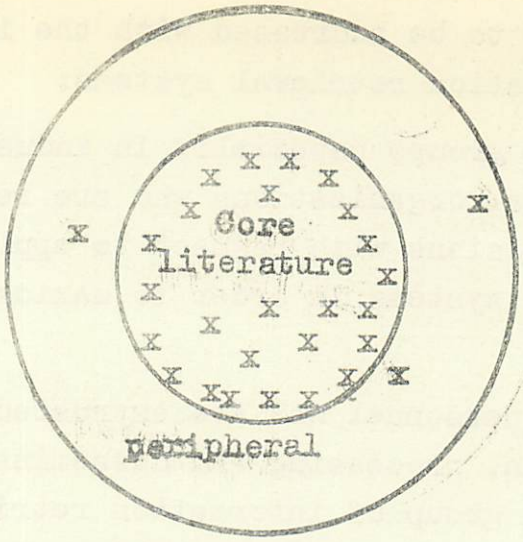
b) Technical personnel who are entrusted with the acquisition, processing and dissemination of information. A group of information retrieval specialists have to be created from professional librarians as well as from other sectors of scientists and engineers who are trained in this field. This group is supposed not only to run information retrieval systems but also to be able to understand and apply advanced techniques and to be able also to disseminate the idea of utilization of mechanized information retrieval systems.

2.3. Unit operations:

Any literature - searching system whether based on the most traditional library methods or the most modern computers, involves a series of functions or steps or unit operations that must be carried through. These unit operations are:

2.3.1 Acquisition of source documents:

Achievement of acquiring "all" source documents pertinent to a special subject is practically improbable. The most that can be attained is to acquire the maximum amount of literature covering this subject (core literature) as well as selection of items of interest to this subject from "peripheral" literature.



2.3.2. Analysis of source documents.

This is conducted in order to infer the intentions of the author as well as anticipations of interests of potential readers. Indexing, classifing, or abstracting constitute the analytical work. But it is not possible to predict every variation of subject content that may interest some potential users.

2.3.3. Control of terminology and subject headings: This control is an essential operation which may be achieved by means of:

a) Subject authority lists.

b) Role indicators or role directors.

c) Modifications to index entries.

d) Cross references or thesaural associations.

e) By use of codes to record the terms chosen in any of the above indexing procedures.

Terminology control has been traditionally accomplished by: (i) Having analysts (indexers, Cataloguers, Classifiers) attempt to remember the headings or classes available for the various aspects of subject matter chosen from source documents. (ii) Providing analysts with authority lists, scope notes, or indexes to classification systems, to avoid placing too much

reliance on human memory.

It is now possible with modern data processing equipment, to impose, control over raw indexing or classifying product of analysts by means of dictionaries, "thesauri", or indexes which are consulted mechanically or electronically.

2.3.4. Recording or results of analysis on a searchable medium:

Media suitable for recording may include:

a) Cards (such as catalog cards).

b) Marginal-hole, hand-sorted punched cards.

c) Machine-sorted punched cards.

d) Punched paper tape.

e) Magnetic tape, cards, cores, or drums.

f) Films (unit, strip, or continuous), and others.

2.3.5. Storage of source documents:

The product of analysis may be an abstract, an extract, and/or a bibliographic references. And it may be stored as the full text.

For document systems, it is sometimes economically feasable to store full size copies of bibliographic references, abstracts, and brief extracts together with "index" data. In aspect systems, the very nature of the searching principle (i.e. matching of document serial numbers) mitigates against storing a copy of a document together with the "index".

2.3.6. Analysis of questions and development of search strategy:

The analysis of questions is conducted in order to transform the language of the inquirer into that of the information-retrieval system and also to ensure that documents relevant to the

the question are identified without too many non-pertinent documents. The search strategy is established in order to exploit in an optimum manner the peculiar capabilities (of human-beings and machine) of the retrieval system used.

2.3.7. Conducting of search:

Search can be conducted by humanbeings only if results of analysis are recorded on catelogue cards for example. Search can also be conducted by humanbeings & machines together according to the techniques used as in case of marginal-hole or machine-sorted-punched cards. Search is conducted only by appropriate equipment when machine-sorted-punched cards are used or when punched cards converted to punched paper tape or to magnetic tape are used.

2.3.8. Delivery of Results of Search:

If the documents are stored separately from index data then the delivery of search results involves the following processes:-

a) The search produces serial or accession numbers of documents indentified as being pertiment to a question.

b) A file of documents arranged in appropriate order is consulted in order to locate the documents identified during the search.

c) The documents may then be consulted, with-drawn, or copied.

If the documents (or extracts, abstracts, bibliographic references, etc.) are stored together with index data, then the search actually provides the source document by physically selecting it as with a marginal-hole punched card or with a film record; or by printing out, as with a magnetic tape system.

2.3.9. Utilization of Results of Search:

Every machine literature-searching system involves a man-machine partnership. Even if every step of the system is mechanized, at least the final one, utilization, remains a human task.

2.4. Applications of Information Retrieval Systems:

Information storage & retrieval systems and techniques go far beyond the handling of scientific and technical documents, as most of these techniques are also employed for all types of documents, for examples-

a) Engineering drawings.

b) Photographs.

c) Maps.

d) Licenses.

e) Insurance policies.

f) Correspondence.

g) Books, reports, and other publications.

The same techniques are also used for information that is not formally recorded in any document as for examples

a) Personnel information.

b) Programmatic information and like.

Chapter III

Examples of Information Retrieval Systems In Action

In this Chapter we will discuss two information retrieval systems. The first is a manual system and the second is the Western Reserve University system which is a fully mechanized information retrieval one.

3.1. A Manual Information Retrieval System:

Marginal-hole punched cards are used in this system and three types of coding[1] are applied to record the results of analysis of a certain subject (Fertilizers in our example).

3.1.1 Direct Code: where a seperate meaning is assinged to each individual part of the search medium in which a notation may be recorded. In our example direct recording of broad subjects of "Fertilizers" are punched at left edge. Notching appropriate holes indicates that the meaning attributed to that hole is true for the record being analized. This direct code is illustrated in Master Card No. (1) as shown below (page 25)

---

(1) A Code is any system of symbols in the communication process, particularly a system that achieves some other desireable advantages over common language or numerical expression; (Documentation In Action, N.Y., 1956, p. 19)

3.1.2. Indirect Code:

This code may also be called "combination code" and is used when there is need to provide notation for more subjects or aspects than there are individual positions available on the search medium. In our example the indirect code is used for alphabetization of author's name mentioned in the document analysed. The three first letters of the author's name are recorded by punching the appropriate number of holes in the above edge of the punched card according to a system of numeration of the alphabet. This indirect code is illustrated in Master card No. 2 shown below (p.26).

The three first letters (PER) of the author (PERRIN) are notched in the selector code in the above edge.

3.1.3. Superimposed Code:-(1)

The information specialist will select indexing headings to analyse his documents from a subject authority list or compiled by himself. Every one of these specialities (index headings) will be given a notation in the form of randum numbers (randum numbers can be chosen from an already prepared dictionary). Two numbers are punched from the randum numbers and punching of these numbers will occur on the lower edge of the card. All the 29 holes in the lower edge would be used for the randum superimposed coding. In order to determine the number of notations that may be superimposed in the same field we must determine the number of combinations that can be recorded in a field of known number of holes. The mathematical equation that gives this number is:

$$C = \frac{H!}{Y!(H-Y)\ !}$$

Where:

C denotes the number of combination of H things taken Y at a time.

(1) See Kent, Allen, ... p. 183 - 5.

In our case, if we wish to attach a meaning to a combination of two holes in a field of 29 holes, then we can indicate any one of 406 subjects in that field. We determine this by substituting in the above equation:

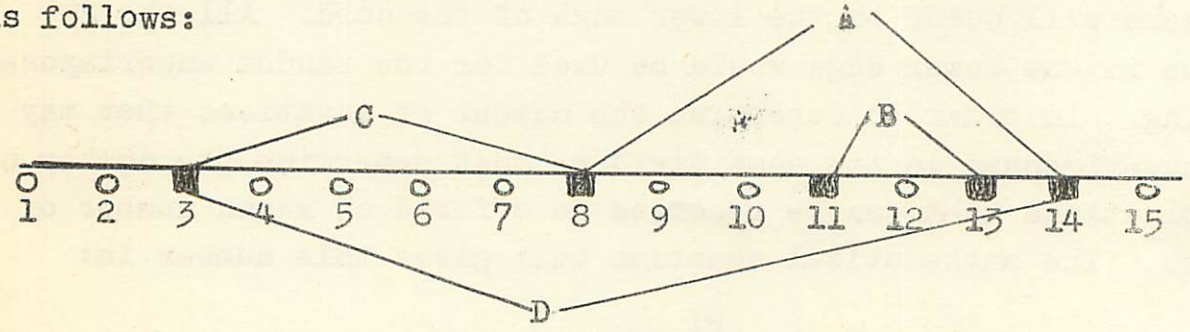
$$C = \frac{29!^{(1)}}{2!\,(29-2)!^{(1)}} = \frac{1 \times 2 \times 3 \ldots\ldots\ldots \times 29}{(1\times2)\,(1\times2\times3 \ldots\ldots\ldots \times 27)} = 406$$

It is obvious that in using notations for more than 406 subjects, using randum numbers as notation, some unwanted combinations will be generated, we should add that in this manual information retrieval system only needles and a punc er are used. Master Card No. 3 illustrates the randum superimposed coding (p.27)

As an example for superimposed coding we can say that if we have four subjects for which are given the following randum numbers

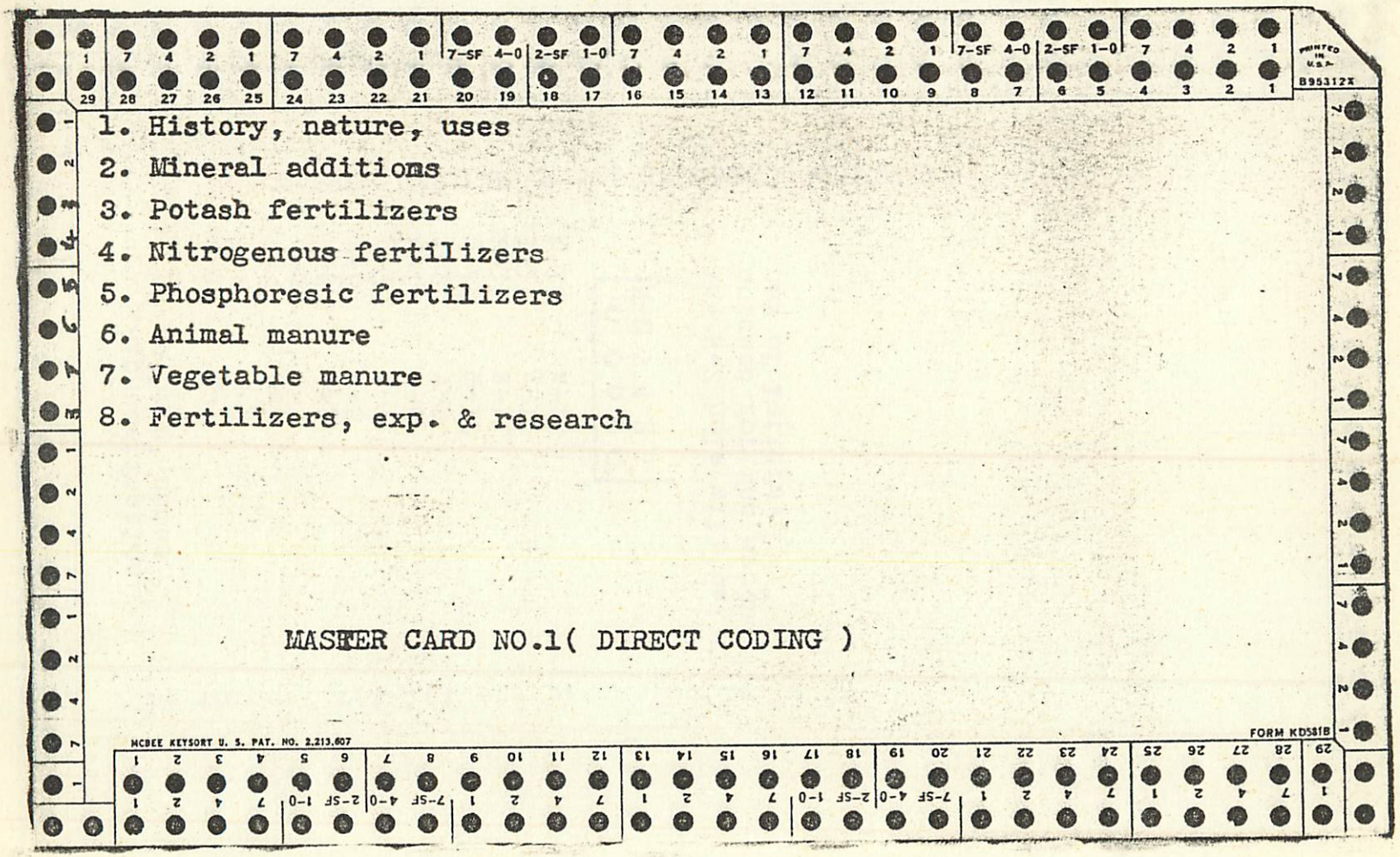
Subject	Randum Numbers
A	8 - 14
B	11 - 13
C	3 - 8
D	3 - 14

These subjects are recorded on the lower edge of the card as follows:



---

(1) The (!) sign is used in mathematics to indicate the factorial of a number, i.e. the continued product of numbers from 1 upward. If H = 3, then H! mean 1 x 2 x 3 = 6


1. History, nature, uses
2. Mineral additions
3. Potash fertilizers
4. Nitrogenous fertilizers
5. Phosphoresic fertilizers
6. Animal manure
7. Vegetable manure
8. Fertilizers, exp. & research
MASTER CARD NO.1(DIRECT CODING)
FORM KD581B
McBEE KEYSORT U. S. PAT. NO. 2,213,607

MASTER CARD NO.2 (INDIRECT CODING)

The three first letters of the auther's name

Alphabetization
of authers' names

a-1	n-1
b-2	o-2
c-3	p-3
d-4	q-4
e-5	r-5
f-6	s-6
g-7	t-7
h-8	u-8
i-9	v-9
j-10	w-10
k-11	x-11
l-12	y-12
m-13	z-13

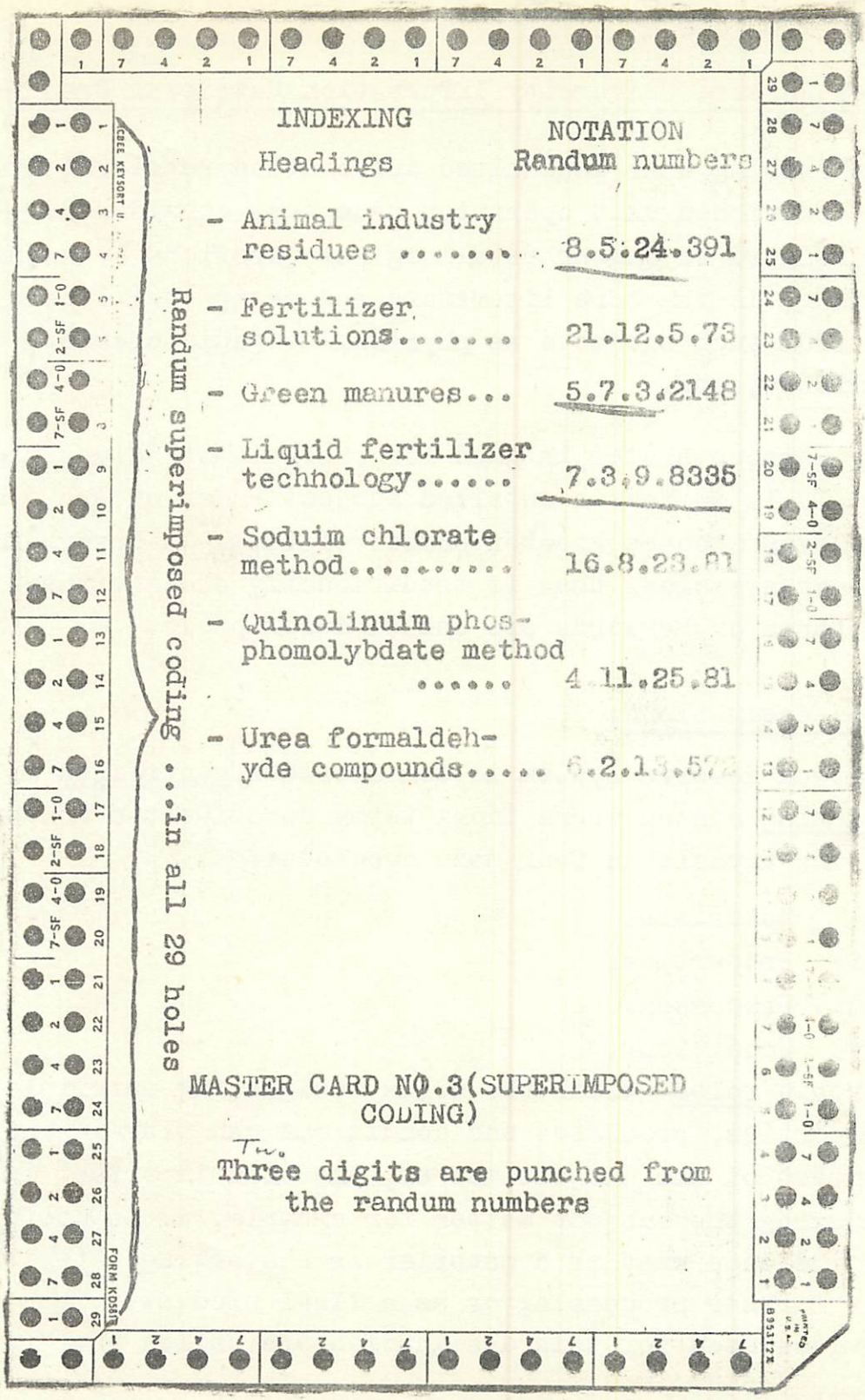
O	O	O	O	O
N-Z	7	4	2	1

The N-Z hole will be notched for the second row of letters.

McBEE KEYSORT U. S. PAT. NO. 2,213,807

FORM KD5618

PRINTED IN U.S.A. B95312X



INDEXING Headings	NOTATION Randum numbers
- Animal industry residues	8.5.24.391
- Fertilizer solutions.....	21.12.5.73
- Green manures...	5.7.3.2148
- Liquid fertilizer technology.....	7.3.9.8335
- Soduim chlorate method.....	16.8.23.81
- Quinolinuim phosphomolybdate method	4.11.25.81
- Urea formaldehyde compounds.....	6.2.13.572

Randum superimposed coding ...in all 29 holes

MASTER CARD NO.3(SUPERIMPOSED CODING)

The Three digits are punched from the randum numbers

3.2. Western Reserve University Information Retrieval System:(1)

In establishing this mechanized information retrieval system, the previous mentioned unit operation have been considered. The system was designed to answer metallurgical questions by machine, i.e. metallurgy is the core literature and peripheral literature is selected from such subjects as physics, inorganic chemistry, electro-chemistry, etc.

Western Reserve system is used for analysis of scientific literature. It is in fact a detailed subject index of articles where the indexer chooses greater number of terms in describing the contents of articles, than is traditionally chosen, these significant terms or keywords are called descriptors.

3.2.1. Basic Assumptions:

a) Western Reserve system is considered a controlled indexing system where index terms describe the contents of the article in four main concepts:

   (i) Materials.
   (ii) Properties.
   (iii) Processes.
   (iv) Conditions.

b) Certain roles which the words designating materials, properties, processes and conditions can play in the context of the subject matter, are also important in indexing the subject matter for example, it is useful to indicate whether a material is a starting material for further processing or as a final product. The devices used for this designation are called the "role indicators".

(1) Hyslop, M.R. "Answering Metallurgical Questions by Machines," Special Libraries, 50, 435-42 Nov. 1959.

3.2.2. OUTLINE OF PROGRAMMING PROCEDURE (1)

a. Compile a preliminary list of English words to be used as search criteria.

   i. Discuss with the questioner modifications in the question that will provide either a more usable question or a more satisfactory set of answers.

      The question is unusable if some of its ideas are not represented in the code dictionary.

   ii. List the words, with a many synonyms as come to mind, that must occur in the telegraphic abstract of any article that will answer the question.

b. Compile the final list of words and codes to be used as search criteria.

   i. Using the word-to-code dictionary, look up the codes for all the words on the preliminary list.

   ii. Using the code-to-word dictionary, select related terms to be added to the list.

   iii. Using the code-to-word dictionary, determine the shortest and least specific grouping of semantic codes that can be used for each word or set of related words.

      Be sure that a proposed grouping of semantic codes does not include codes for unwanted words.

   iv. Select role indicators required to express relationships among the terms.

C. Construct the program.

   i. Write the formula using English words and conventional Boolean-algebraic symbols.

   ii. Substitute codes for words.

---

(1) Pamphlet with limited distribution by the Center for Documentation and Communication Research, WRU, 1962.

iii. Substitute machine symbols for Boolean-algebraic symbols.

3.2.3. Example of a Question/Answer by WRU System:(1)

An actual example of the above mentioned system can be shown:

An American company sent to the Center of Documentation and Communication Research asking to have retrospective search (covering 1962 - 1963) about what is published on Bainite. In the question the company was particularly interested in (a) mechanical properties (b) austenite to bainite transformation and decomposition. From the question the following indexing terms as well as the corresponding semantic code can be mentioned.

Terms	Semantic Code
A. Bainite	LALL 1002 RYNG 1013 RYRN 3002
B. Decomposition	DATR 1002
C. Mechanical properties	PARR 1087
D. Transformation	CUNG 1019

It should be noted that the semantic code includes all other synonyms, e.g.:

Bainite	LALL 1002 RYNG 1013 RYRN 3002
Martensite	LALL 1002 RYNG 1013 RYRN 3001
Troostite	LALL 1002 RYNG 1013 RYRN 3003
Anstenite	LALL 1002 RYNG 1013 RYRN 3025

(1) A practical Question/Answer conducted by Ahmad Badr at western Reserve university, 1962.

In our case two programs are established, the first one may be more specific in which the different particulars of the questions are included in the program. The formula for this program may appear as follows:

$$6\ \left[2(A)2\ .\ (B + C + D)\right]\ 6$$

In this formula the numbers indicate that the search would be conducted on level 6 which means that all information required should be found in one sentence and grouped between the two main brackets. Also number 2 means that the search is conducted on the level of one or more[1] words.

If we substitute the letters in the formula with their codes, the formula may appear as follows:

6%2%L-LL. 1002≠R-NG. 1013≠R-RN □ 2≠%D-TR. 1002 &

P-PR. 1087 & C-NG. 1019 □ □ 6

Explanation of the substitution of the letters as well as brackets and other machine symbols can be elucidated by the two tables of the Boolean Polynomial and the Machine Symbols mentioned at the end of this chapter.

---

(1) The words are grouped into sequences of different complexity called "syntactic" grouping. Nine degrees of complexity are provided for. The largest grouping is the abstract, called a grouping of "level 8". Level 6" is the grouping of the sentence. A grouping of "level 2" consists of one or more words. A grouping of "level 4" consists of one or more role indicators & one or more 2 level groupings.

The program in machine language is then punched on IBM cards. This program is fed into the computer (GE 225) which has a memory in which all literature are previously stored on a magnetic tape. When matching of information required (indicated by holes in the punched card) occurs with that information stored into the magnetic tape, the output (numbers of documents that contain pertinent information) will be automatically listed by the machine typewriter.

The output (on punched cards or lists) is checked against original abstracts in the "Review of Metal Literature" to select the pertinent answers of the required question.

3.2.4. THE BOOLEAN POLYNOMIAL

Boolean notation is a method for expressing the following combinations of terms:

Combination	Notation
A must be found with B	A . B
Either A or B must be found	A + B
A must not be found with B	A - B
A must be found with either B or C	A . (B + C)
A must not be found with either B or C	A - (B + C)

Levels on which combinations may occur are expressed by enclosing the combination with numbered brackets.

Examples:

A must be found with B on the 2 level	$2\left[A \cdot B \right] 2$
A must be found with B or C on the 4 level	$4\left[A \cdot (B + C) \right] 4$

A must be found with B on the 2 level, with C or D on the 4 level, and with E on the 6 level

$$6\left[4\left[2\left[A \cdot B\right]2 \cdot (C+D)\right]4 \cdot E\right]6$$

3.2.5. MACHINE SYMBOLS

Level Two and Above

1. Parenthesis or bracket open %
2. Parenthesis or bracket closed □
3. dot indicating logical and ¤
4. Plus sign indicating logical or &
5. Minus sign indicating logical not is indicated by an asterisk followed by a space followed by NOT followed by the set of terms to be negated, which set must be enclosed in parentheses. ¢
6. universal character —
7. space ≠

Examples:

1, 2, 3, & 4. 4[A • (B+C)]4 becomes 4%A¤%B&C □□4

5. ¤≠NOT%A+B □
6. C-NS
7. LALL¤RERN¤SGA/.H≠≠≠

Level One

1. Logical and .(Period)
2. Logical or /

Examples:

1. CUNSO1025 becomes CUNS. 1025
2. CUNSO1025 + CUNSO1027 becomes CUNS.1025/1027

Chapter IV

Some Actual and Potential Information Retrieval Systems in the U.A.R.

4.1. An approach:

World development in the area of computers as well as information retrieval systems both manual and mechanized has reached a highly advanced stage and is being developed all the time. The U.A.R. specially in its social revolutionary stage where science is its main tool, has not stood far from this world development. On the contrary it has contributed actively in this field. The work being done in the Institute of National Planning, Operations Research Center, as well as in the National Information and Documentation Center of the Ministry of Scientific Research in the area of documentation, information retrieval and applications of mechanical tools, including computers, may be considered as a nucleus of the future trend in this field.

4.2. The superimposable system of the National Information and Documentation Center:(1)

An illustration of a manual information retrieval system is the one applied at the National Information & Documentation Center. Periodicals at the Center have reached about 3000 periodicals from all-over the world. Thus multidimensional types of questions about periodicals became increasingly difficult. Such questions as "what Russian periodicals in your library are written in English in the area of nuclear physics?" or "what German periodicals dealing with popular biology?" The center has developed a system to answer all these questions efficiently. This system is called the system of superimposable punched cards. The following are the features of the system.

(1) Garrids, "Superimposable punched cards.....see references at the end of the report.

a. Each periodical is given a serial number which is at the same time its position on the shelves.

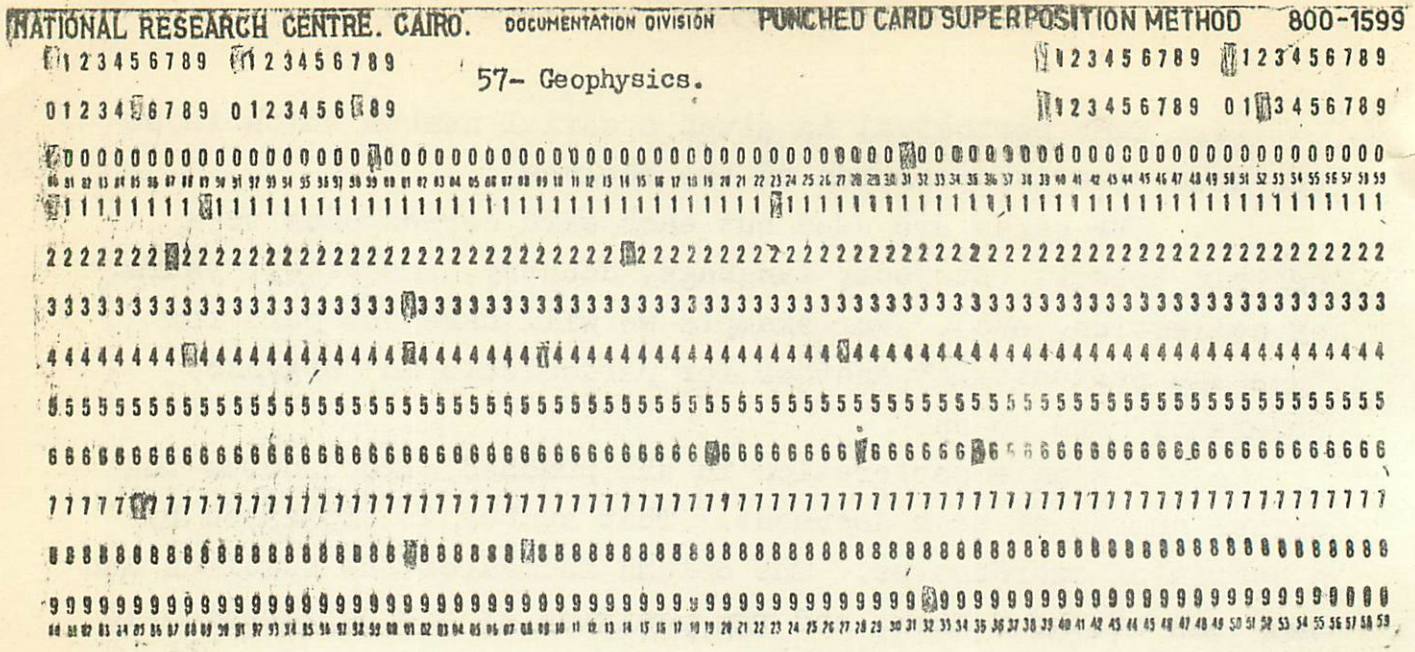
b. IBM cards are used but each card corresponds to a certain aspect (subject, language, country, frequency, years of collection, etc). For example we will have one card for "Russian periodicals", another for periodicals in "organic chemistry" and so on.

c. A single perforation in the punched card denotes the serial number of each document. That number is expressed by a system of coordinates. The column indicates the hundreds and the tens of the number, and the position of the perforation in the column indicates the units. There is one drawback, however, a card cannot take more than 800 periodicals and thus another deck of cards must be started for each batch of 800 periodicals.

d. A total of 275 different aspects have been selected. Each card represents one aspect or feature.

e. Let us now suppose that we want to select periodicals combining two different aspects we take the two cards corresponding to these two aspects, and by placing one on top of the other we obtain the necessary information, since the squares corresponding to periodicals possessing both aspects will have been perforated on both cards.

g. The production of the cards is an easy matter with the use of IBM machines. The first step is to make out a set of cards, with one card for each periodical, on each card we record, in code, all information relating to the corresponding periodical, i.e. all the distinctive aspects it displays. By the use of IBM sorting apparatus, we then pick out all the cards which have a common aspect and thus discover which numbers should appear on the card corresponding to that aspect. This process means that we can use the IBM cards as a document system as well as the aspect system. An example of the card used is shown in the figure below.



4.3. Information Retrieval by IBM Equipment in the Operations Research Center:

Two experiments have been conducted in order to indicate the potentiality of some available devices and machines in analysis, organization and retrieval of information.

4.3.1. Use of the IBM Collator 077:

The objective of this experiment is to show the possibility of greater depth of analysis of contents of books (or documents) than traditionally practised in the library. A number of books have been chosen, some of them have a common subject. Keywords selected from an authority list prepared by specialists in the field are assigned to describe in a considerable depth the contents of these books as shown in list No.1(p.41) for one of the books in which keywords describing its contents and corresponding pages appear. These keywords are abbreviated to a feasable degree. For example "LP" indicates linear Programming, "Eqn" for Equation. Keywords resulting from analysis of the book are punched on IBM cards. The card has the design shown in Fig(1)

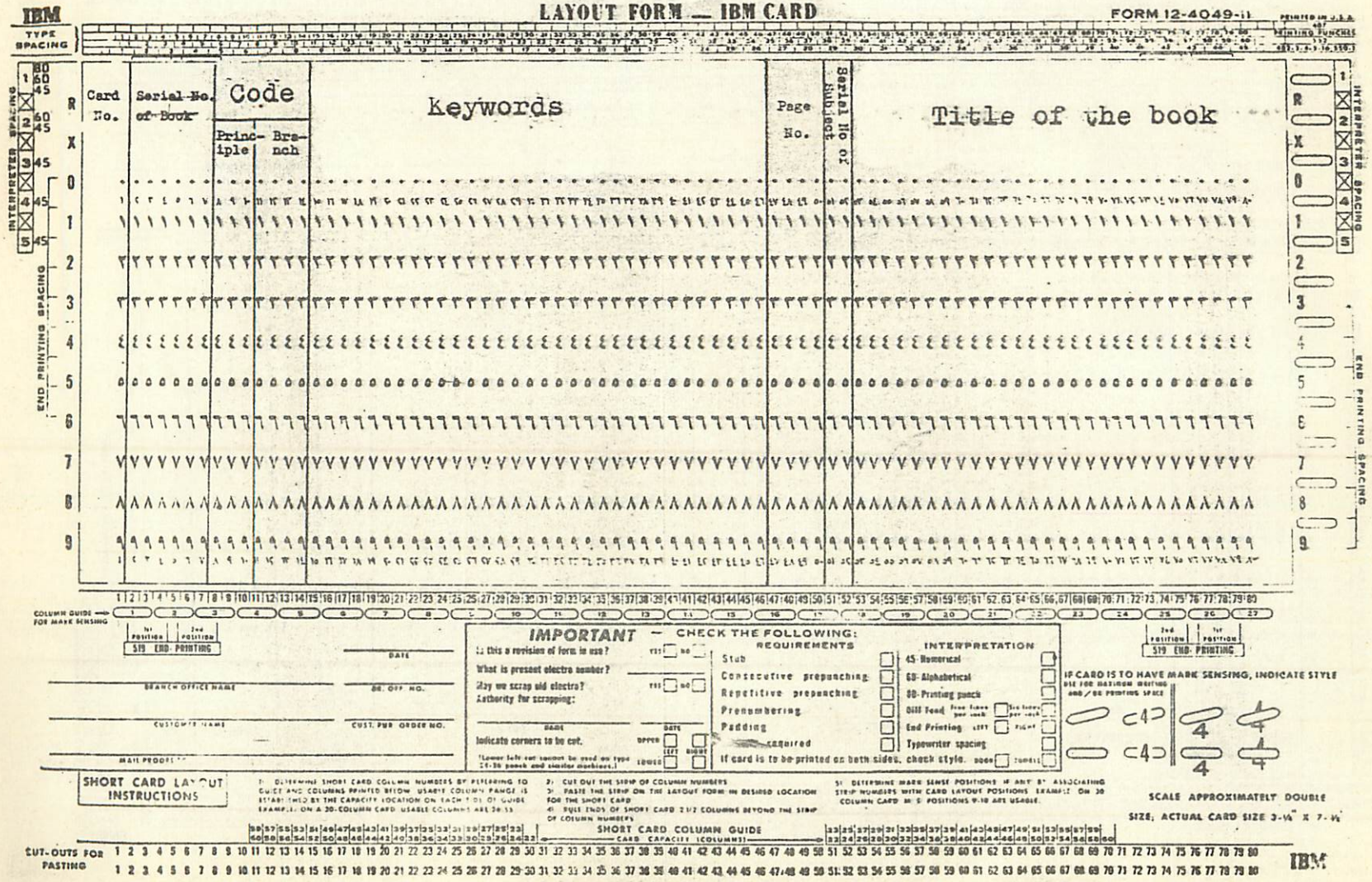
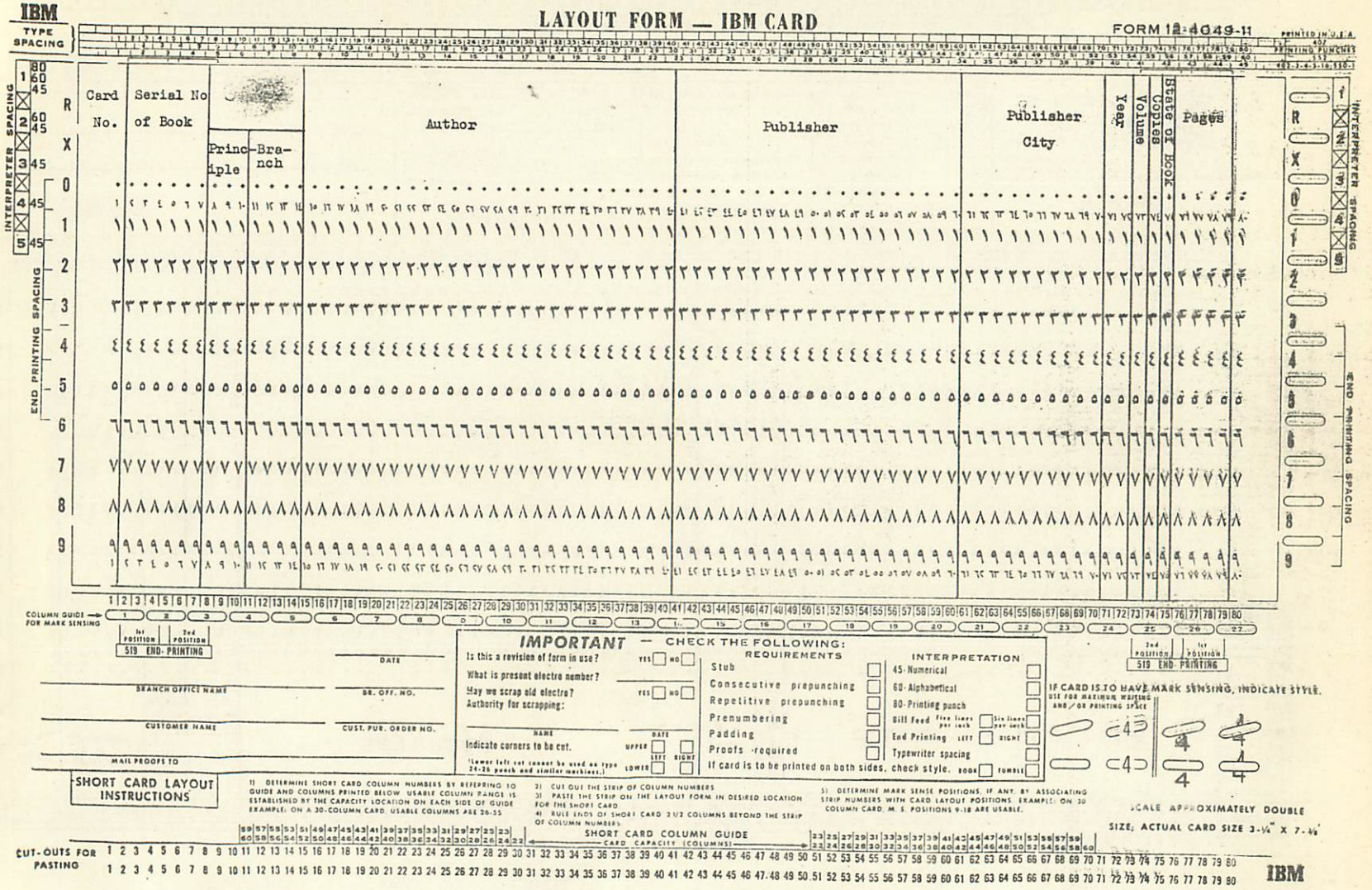

IBM
LAYOUT FORM — IBM CARD
FORM 12-4049-11
TYPE SPACING
INTERPRETER SPACING
END PRINTING SPACING
Card No.
Serial No. of Book
Code
Princ-iple
Bra-nch
Keywords
Page No.
Serial No. of Subject
Title of the book
COLUMN GUIDE FOR MARK SENSING
IMPORTANT — CHECK THE FOLLOWING:
REQUIREMENTS
INTERPRETATION
Stub
Consecutive prepunching
Repetitive prepunching
Prenumbering
Padding
required
45- Numerical
69- Alphabetical
88- Printing punch
Skip
End Printing
Typewriter spacing
If card is to be printed on both sides, check style.
IF CARD IS TO HAVE MARK SENSING, INDICATE STYLE
519 END PRINTING
SHORT CARD LAYOUT INSTRUCTIONS
SHORT CARD COLUMN GUIDE
CARD CAPACITY (COLUMNS)
CUT-OUTS FOR PASTING
SCALE APPROXIMATELY DOUBLE SIZE; ACTUAL CARD SIZE 3-1/4" X 7-3/8"

Fig (1)


IBM
LAYOUT FORM — IBM CARD
FORM 12-4049-11
PRINTED IN U.S.A.
TYPE SPACING
INTERPRETER SPACING
END PRINTING SPACING
Card No.
Serial No of Book
Princ-iple
Bra-nch
Author
Publisher
Publisher City
Year
Volume
Copies
State of Book
Pages
519 END-PRINTING
DATE
BRANCH OFFICE NAME
BR. OFF. NO.
CUSTOMER NAME
CUST. PUR. ORDER NO.
MAIL PROOFS TO
IMPORTANT — CHECK THE FOLLOWING:
Is this a revision of form in use?
What is present electro number?
May we scrap old electro?
Authority for scrapping:
Indicate corners to be cut.
REQUIREMENTS
Stub
Consecutive prepunching
Repetitive prepunching
Prenumbering
Padding
Proofs required
INTERPRETATION
45-Numerical
60-Alphabetical
80-Printing punch
Typewriter spacing
If card is to be printed on both sides, check style.
IF CARD IS TO HAVE MARK SENSING, INDICATE STYLE.
SHORT CARD LAYOUT INSTRUCTIONS
SHORT CARD COLUMN GUIDE
CARD CAPACITY (COLUMNS)
CUT-OUTS FOR PASTING
SCALE APPROXIMATELY DOUBLE SIZE; ACTUAL CARD SIZE 3-¼" X 7-⅜"
IBM

which contains beside keywords, the accession No, code, title of the book, page number. Another card is prepared for every book containing all other bibiliographic citations not included in the first ones, such as: auther, publisher, year of publication, place of publication, volume, copies, state of the book and pages, as in Fig (2).

An example of the analysis of three books appears on list No.2.(p.43). From which it can be seen that the deck of cards of the first book are arranged serially and the information given is according to the design of Master card Fig(1).

List No. 1

Analysis of one of the books

Natanson, I.P.

Theory of functions of a real variable by I.P. Natanson, translated by L.F. Boron. N.Y. 1955.

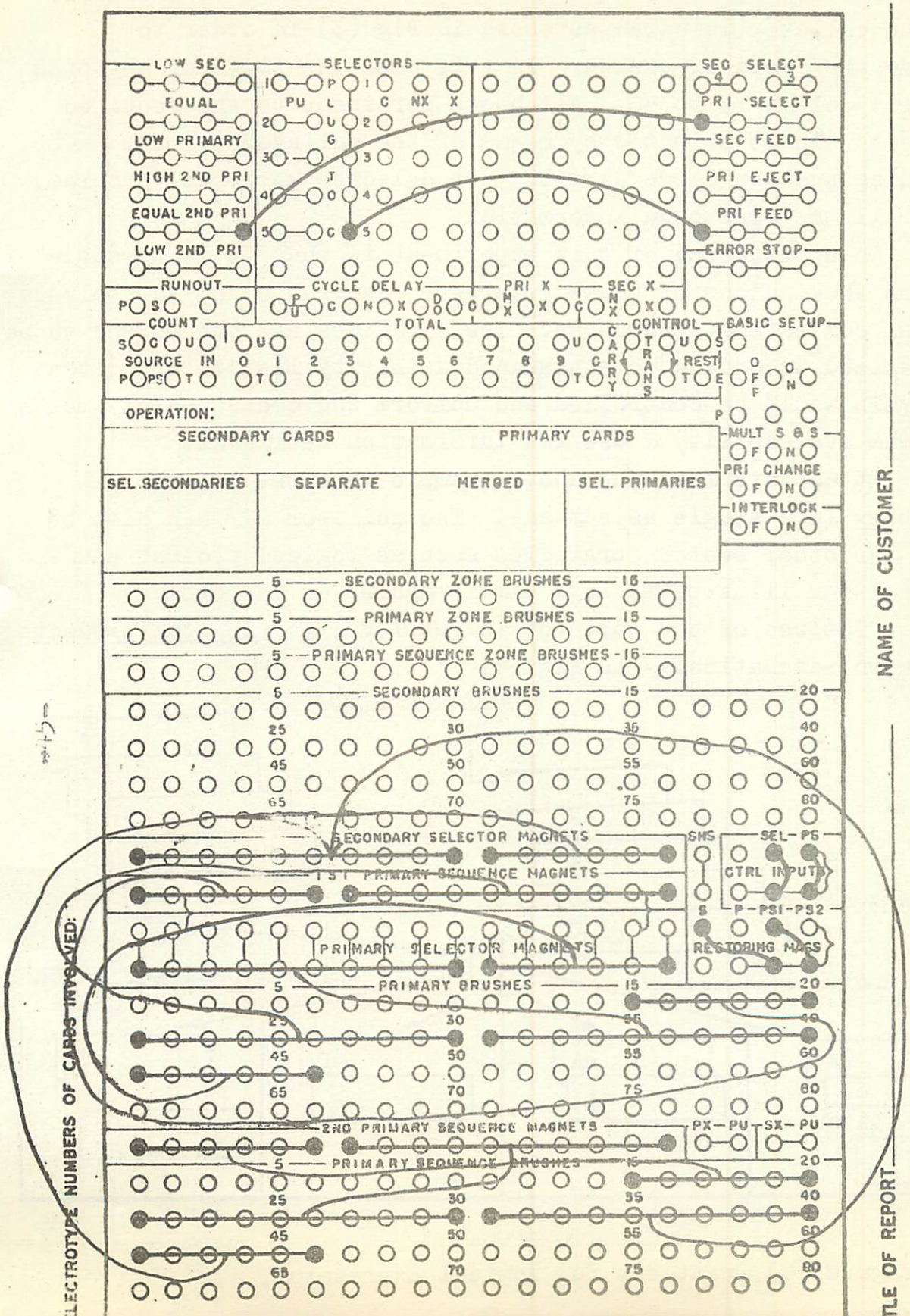
TERM	Page
Infinite sets	11
Correspondence One-to-One	15
Denumerable sets	17
Comparison Power	27
Limit Point	34
Closed Sets	36
Interior Point	41
Open Sets	41
Distance	44
Separation	44
Bounded Open Set Structure	47
" Closed " "	
Condensation Point	50
Closed Set Power	50

LIST NO. 2

Acc.No.	Code	Subject heading	Page	Ser. No.	Title
3005945	5199200	PIVØTING	0173	01	LINEAR PROGRAMMING AND EXTENTIØN
3005945	5199200	METRICES	0183	02	LINEAR PROGRAMMING AND EXTENTIØN
3005945	5199200	GAMES	0277	03	LINEAR PROGRAMMING AND EXTENTIØN
3005945	5199200	UNCERTANTY	0499	04	LINEAR PROGRAMMING AND EXTENTIØN
3005945	5199200	INFLUENCE	0012	05	LINEAR PROGRAMMING AND EXTENTIØN
3005945	5199200	ØRIGIN	0012	06	LINEAR PROGRAMMING AND EXTENTIØN
3005945	5199200	DUALITYTHERØM	0128	07	LINEAR PROGRAMMING AND EXTENTIØN
3005945	5199200	SPACEVECTØR	0177	08	LINEAR PROGRAMMING AND EXTENTIØN
3005945	5199200	INNERSPACE	0254	09	LINEAR PROGRAMMING AND EXTENTIØN
3005945	5199200	CLASSICTRANSPØRT	0299	10	LINEAR PROGRAMMING AND EXTENTIØN
3005945	5199200	FIØWMAX	0385	11	LINEAR PROGRAMMING AND EXTENTIØN
3005945	5199200	ASSIGNTØPTIMAL	0316	12	LINEAR PROGRAMMING AND EXTENTIØN
3005945	5199200	BØUNDUPPERVAR	0368	13	LINEAR PROGRAMMING AND EXTENTIØN
3005945	5199200	GEØMLP	0147	14	LINEAR PROGRAMMING AND EXTENTIØN
3000596	5199200	CØNCEPTS	0021	15	LINEAR PROGRAMMING METHØDS
3000596	5199200	CØMPARISØN	0021	16	LINEAR PROGRAMMING METHØDS
3000596	5199200	MAPS	0232	17	LINEAR PROGRAMMING METHØDS
3000596	5199200	MATRICES	0378	18	LINEAR PROGRAMMING METHØDS
3000596	5199200	GAMES	0499	19	LINEAR PROGRAMMING METHØDS
3000596	5199200	METHØDSIMPLE	0053	20	LINEAR PROGRAMMING METHØDS
3000596	5199200	CUTSSHØRT	0151	21	LINEAR PROGRAMMING METHØDS
3000596	5199200	CHECKSCØMPUTATN	0151	22	LINEAR PROGRAMMING METHØDS

Acc.No.	Code	Subject heading	Page	Ser. No.	Title
3000596	5199200	CØDINGDATA	0308	23	LINEAR PROGRAMMING METHØDS
3000596	5199200	MØDTRANSP	0332	24	LINEAR PROGRAMMING METHØDS
3000596	5199200	CØMPUTELECTRØNIC	0308	25	LINEAR PROGRAMMING METHØDS
3000596	5199200	ANALYSISINØUT	0474	26	LINEAR PROGRAMMING METHØDS
3000596	5199200	MEASUR	0132	27	LINEAR PROGRAMMING METHØDS
3002548	5100000	DETERMINANT	0013	28	ENGINEERING MATHEMATICS
3002548	5100000	CØMPLEXNØ	0062	29	ENGINEERING MATHEMATICS
3002548	5100000	ALGEBEQN	0087	30	ENGINEERING MATHEMATICS
3002548	5100000	DIFFEQN	0155	31	ENGINEERING MATHEMATICS
3002548	5100000	BESSELFN	0197	32	ENGINEERING MATHEMATICS
3002548	5100000	CALCSVECTØR	0225	33	ENGINEERING MATHEMATICS
3002548	5100000	DIAGROUND	0236	34	ENGINEERING MATHEMATICS
3002548	5100000	FØRMULAINTERPØL	0001	35	ENGINEERING MATHEMATICS
3002548	5100000	ANALYSISDIMENL	0042	36	ENGINEERING MATHEMATICS
3002548	5100000	FNHYPERB	0062	37	ENGINEERING MATHEMATICS
3002548	5100000	FØURIERSERIES	0132	38	ENGINEERING MATHEMATICS
3002548	5100000	GAMMAFN	0197	39	ENGINEERING MATHEMATICS
3002548	5100000	ALGEBVECTØR	0214	40	ENGINEERING MATHEMATICS
3002548	5100000	STRETCHSTRING	0236	41	ENGINEERING MATHEMATICS
3002548	5100000	EFFECTSKIN	0252	42	ENGINEERING MATHEMATICS

INTERNATIONAL BUSINESS MACHINES CORPORATION
TYPE 077 COLLATOR CONTROL PANEL



If some body asks about books in a certain subject key-words pertaining to the question would be selected. The board of the collator is wired as shown in Fig (3) in order to define the field of search. In this case the field of keywords is from column 15 to 46. All books pertinant to the required subject can be selected by means of the collator through matching punched cards carrying the selected keywords (program) with all the prestored information.

An application of this experiment is that it is possible to use the indices of books (wherever available) as an analysis of the contents of the books. These indices are understood to be normalized according to a standard authority list in order that analysis would be controlled and uniform and consequentely makes maximum availability & use for information retrieval.

It appears from the above example that the search strategy is a single aspect one. The collator 077 can also be used for other search strategies such as logical product and logical sum illustrated in a simple manner as follows:

The use of the collator in retrieval for <u>logical product</u> is shown schematically in Fig (4)

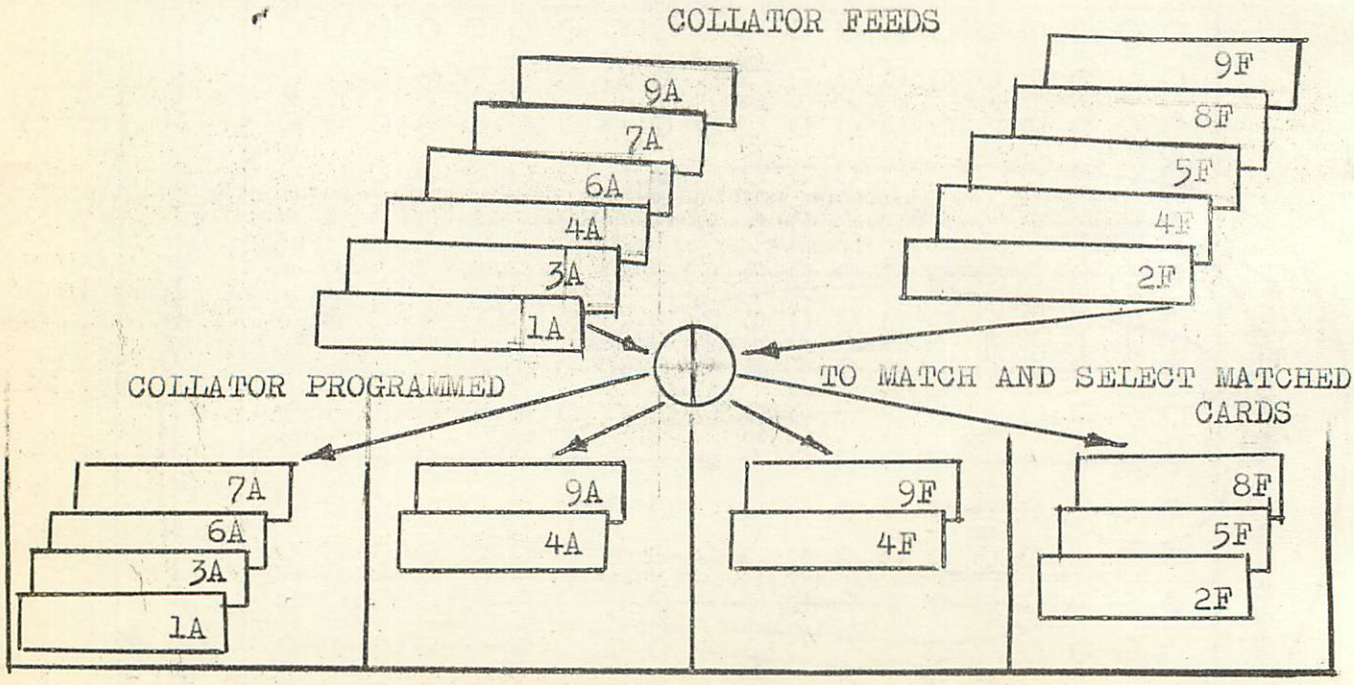


Figure (4) Treatment for logical Expression, A.F

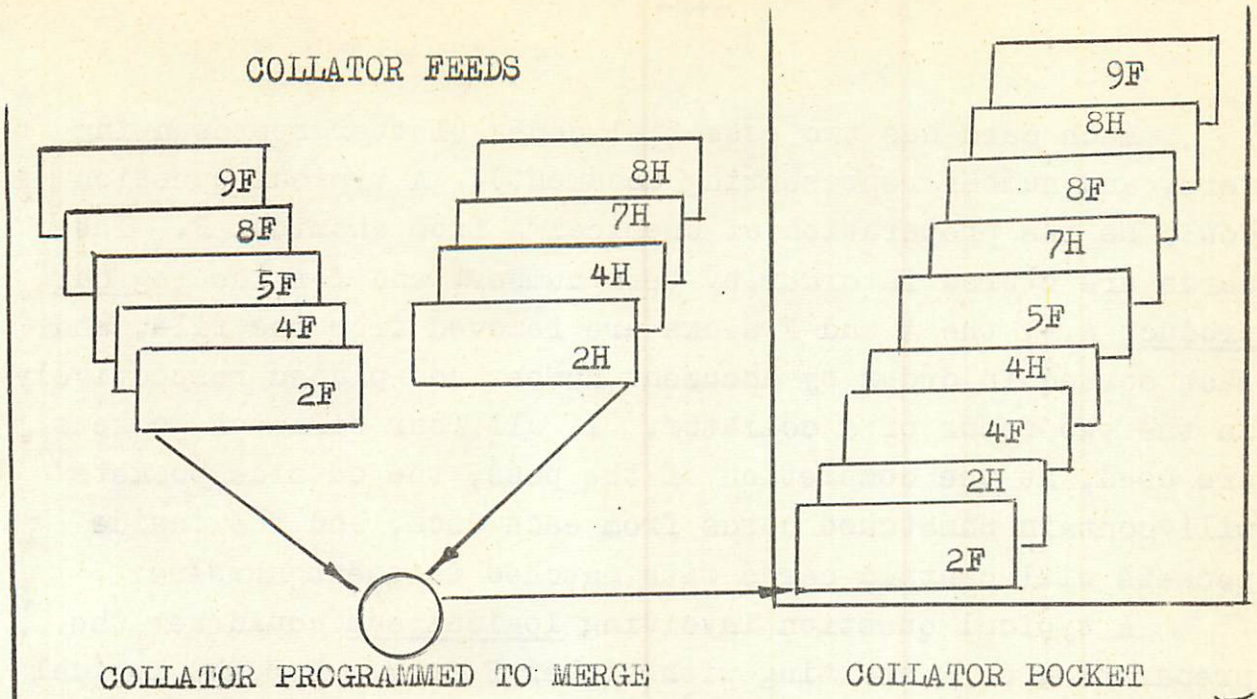


Figure (5) Treatment for logical Expression, A(F+H), Part 1: Logical Sum, (F+H)

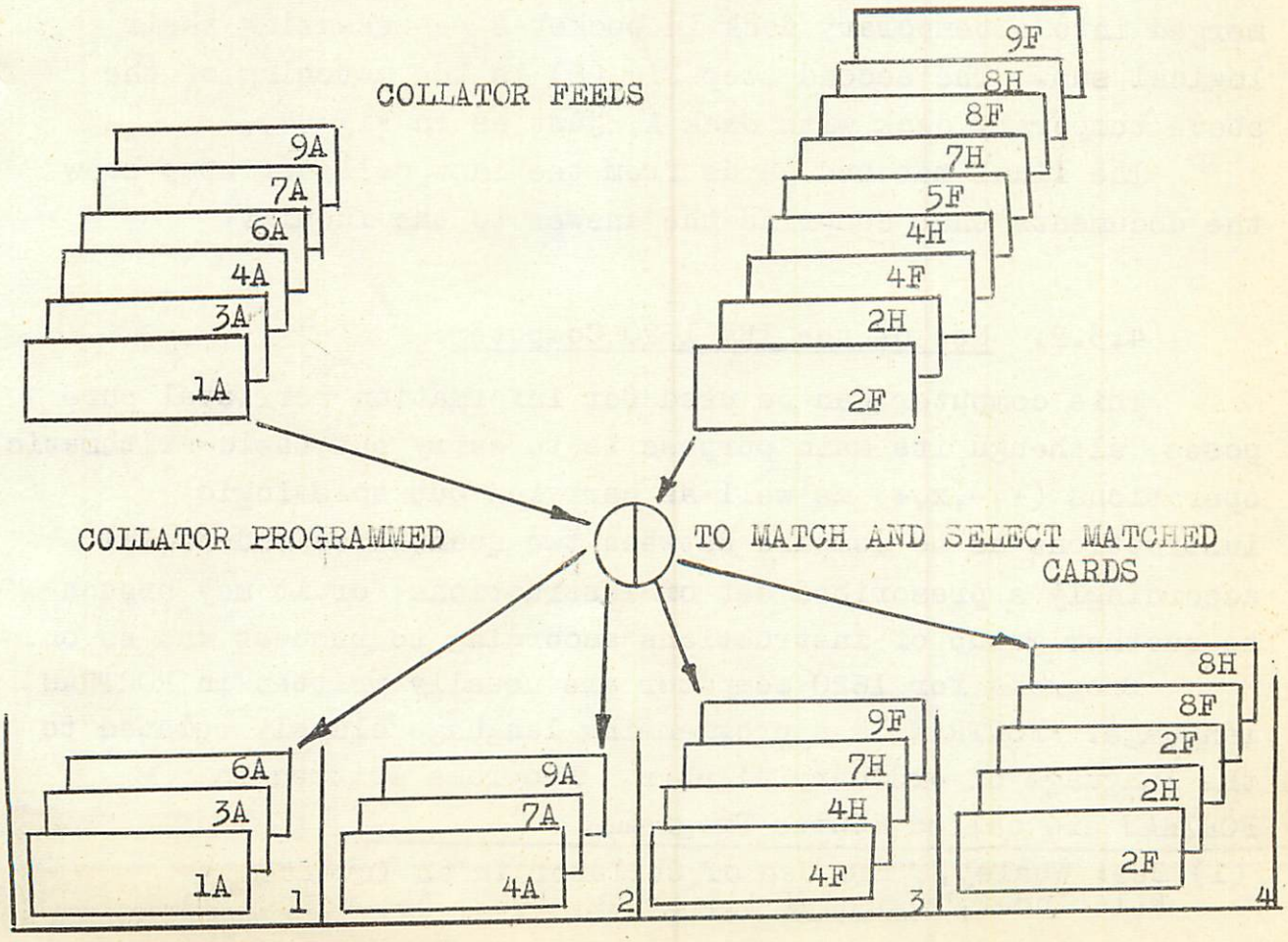


Figure (6) Treatment for Logical Expression, A(F+H), Part 11:

Each card has two essential codes (letter representing term, and number representing document). A typical question could be the preparation of chemical A from chemical F. The cards are stored in order, by term number, and for the logical product A.F, the A and F decks are removed from the file, and each sorted in order by document number, and placed respectively in the two feeds of a collator. If all four collator pockets are used, at the completion of the pass, the outside pockets will contain unmatched cards from each deck, and the inside pockets will contain cards with matched document numbers.

A typical question involving logical sum could be: the preparation of A starting with either F or H. For the logical expression A (F + H), the first step is to obtain the logical sum of F + H as shown in Fig (5). Here the two decks are merged into a temporary deck in pocket 2 representing their logical sum. The second step Fig (6) is the matching of the above temporary deck with deck A, just as in Fig (4).

The final matched cards from the last collator step show the documents that comprise the answer to the inquiry.(1)

4.3.2. Use of the IBM 1620 Computer:

This computer can be used for information retrieval purposes, although its main purpose is to carry out basic arithmatic operations (+,-,x,÷) as well as carrying out some logic instructions as to compare between two quantities and follow accordingly a prescribed set of instructions; or it may branch to another group of instructions according to request and so on.

Programs for 1620 computer are usually written in FORTRAN language. FORTRAN is a programming language closely related to the language of ordinary algebra. Programs written in FORTRAN are called Source Programs.

(1) See: Whaley, "The Use of Collator in an Inverted File Index," Special Libraries, 1962, 53, No.2 p.70-2.

Since the machine cannot accept instructions except when written in machine language, the source program has to be translated to its corresponding machine language. This is done automatically by the machine through a translator, called the processor or the compilor.

Programming in FORTRAN does not require at all any previous knowledge about computer. This enables scientists & engineers to program on the machine without any necessary knowledge about it.

There are six basic operations associated with the FORTRAN language. Each operation is represented by a specific symbol as follows:

Operation	symbol	example
Addition	+	A + B
Subtraction	-	A - B
Multiplication	*	A * B
Division	/	A / B
Exponentiation	* *	A * * B=(A^B)
Equality	=	A = B

The following is a simple illustration of FORTRAN programming.[1] Consider the quadratic equation

$$3x^2 + 1.7x - 31.92 = 0$$

The algebric representation for one of the two roots is

$$x = \frac{-B + \sqrt{B^2 - 4AC}}{2A}$$

A FORTRAN program that describes this calculation is as follows:

```
A = 3.
B = 1.7
C = - 31.92
Root = (- B + (B * * 2 - 4. * A * C) * * .5)/(2.*A)
```

(1) An example quoted from Lesson, D., Basic Programming

The first statement means : " assign the value 3. to the variable A." The next two statements have a similar meaning. The fourth statement means: " Evaluate the expression on the right side and assign the result to the variable ROOT." Of course, the program also could be written as follows:-

```
ROOT = (- 1.7 + (1.7 ** 2 - 4. * 3. * (-31.92)) ** .5) / (2. * 3.)
```

It is clear that the computer carries on its operations at a very high speed(1) and also it carries on logical processes which mainly distinguish it from the desk calculator.

The following is a standard program prepared by IBM company for information retrieval on the IBM 1620 computer:-

---

(1) The IBM 1620 computer adds 10 digits numbers at a rate of 1000 additions per second and it can execute logical operations at the rate of 5000 operations per second.

A Demonstrative FORTRAN Program For Information Retrieval:

```
      DIMENSION      IN(10),KM(10),KO(10),KN(10),KWC(10),K(10)
C     INFORMATION RETRIEVAL DEMONSTRATION-1620
C     J W BURGESON    I B M AKRON
      JAND=41554400
      JOR=56590O
      JBNT=42556300
    1 TYPE 7800
 7800 FORMAT (47H THE 1620 IS NOW GOING TO DEMONSTRATE TO YOU AN/
 7800 1         34H INFORMATION RETRIEVAL APPLICATION//
 7800 2        44H YOUR OPERATOR HAS A DECK OF CARDS WHICH ARE/
 7800 3        44H REPRESENTATIVE OF A LIBRARY FILE. EACH CARD)
      TYPE 7801
 7801 FORMAT (44H CONTAINS INFORMATION ON ONE LIBRARY ARTICLE/
 7801 1         31H THIS INFORMATION CONSISTS OF -/
 7801 2         42H  TITLE     VOL NO.   PAGE NO.  6 KEYWORDS//
 7801 3         46H YOUR OPERATOR WILL TELL THE 1620 WHAT LOGICAL)
      TYPE 7802
 7802 FORMAT (40H COMBINATION OF KEYWORDS TO SEARCH UPON.)
      IN(9)=0
 4568 TYPE 9432
 9432 FORMAT(11H SEARCH FOR)
    3 ACCEPT 7777,(K(I),IN(I),I=1,8),K(9)
 7777 FORMAT (17A5)
      KNT=0
      NKM=0
      NKO=0
      NKN=0
      DO 4 I=1,8
      IF(IN(I))45,601,4
  601 INDEX=I-1
      GO TO 602
    4 CONTINUE
      INDEX=8
  602 I=1
    5 KT=K(I)
      IF(IN(I)-JAND)20,10,20
   10 NKM=NKM+1
      KM(NKM)=KT
   15 I=I+1
      GO TO 5
   20 IF(IN(I)-JOR) 50,25,50
```

```
   25 NKO=NKO+1
      KO(NKO)=KT
      NKO=NKO+1
      I=I+1
      KO(NKO)=K(I)
      IF(IN(I)-JBNT)30,70,30
   30 IF(IN(I)-JAND) 35,15,35
   35 IF(IN(I))45,100,45
   45 TYPE 7805
 7805 FORMAT (39H PLEASE RETYPE THE SEARCH PATTERN. YOUR/
 78051         37H TYPED LINE ABOVE DOES NOT CONFORM TO/
 78052         18H MY SPECIFICATIONS/)
      GO TO 4568
   50 IF(IN(I)-JBNT) 55,65,55
   55 IF(IN(I)) 45,60,45
   60 NKM=NKM+1
      KM(NKM)=KT
      GO TO 100
   65 NKM=NKM+1
      KM(NKM)=KT
   70 I=I+1
      NKN=NKN+1
      KN(NKN)=K(I)
      IF(IN(I)-JOR) 75,70,75
   75 IF(IN(I))45,100,45
C     ANALYSIS OF SEARCH PATTERN COMPLETED
  100 IF(INDEX)45,605,610
  605 TYPE 7810,K(1)
      GO TO 701
  610 TYPE 7810,(K(I),IN(I),I=1,INDEX),K(INDEX+1)
      TYPE 7870
 7870 FORMAT (/)
  701 IF(SENSE SWITCH 9) 101,101
 7810 FORMAT (17H VERIFY SEARCH AS/9A5/8A5)
  101 IF(SENSE SWITCH 9) 500,102
  102 IF(SENSE SWITCH 1) 500,105
  105 READ 4444,(KWC(I),I=1,6)
 4444 FORMAT (33H                                    ,6(1X,A5))
      IF(NKM) 45,131,110
  110 DO 130 J=1,NKM
      DO 120 I=1,6
      IF(KWC(I)-KM(J))120,130,120
  120 CONTINUE
      GO TO 101
  130 CONTINUE
C     PASSED MUST TEST
  131 IF(NKO) 45,155,132
  132 DO 150 J=1,NKO,2
      DO 140 I=1,6
      IF(KWC(I)-KO(J))135,150,135
```

```
  135 IF(KWC(I)-KO(J+1))140,150,140
  140 CONTINUE
      GO TO 101
  150 CONTINUE
C     PASSED OR TEST
  155 IF(NKN)45,200,160
  160 DO 165 J=1,NKN
      DO 165 I=1,6
      IF(KWC(I)-KN(J))165,101,165
  165 CONTINUE
C     PASSED NOT TEST
  200 TYPE 4444
      KNT=KNT+1
      GO TO 101
  500 TYPE 7850,KNT
      PAUSE
      TYPE 7860
      GO TO 4568
 7850 FORMAT(//20H END OF THIS SEARCH.I5,6H FOUND//)
 7860 FORMAT(/25H TRY A NEW SEARCH PATTERN//)
      END
```

In the following page, sample input data is given.

```
PHYSICAL REVIEW          VOL  9 PAGE 24 INDIA FORCE STATE LARGE TEXAS PLANT
PHYSICAL REVIEW          VOL 31 PAGE 57 PHONO CRUDE METAL SOLID AUDIT TREAD
PHYSICAL REVIEW          VOL 69 PAGE 34 MAJOR AUDIO OPTIC SOLID TURBO TABLE
PHYSICAL REVIEW          VOL 53 PAGE 56 CURES SOLID MAINE SOLID TABLE PORUS
PHYSICAL REVIEW          VOL 28 PAGE 90 COSTS BUTYL MAINE QUEUE OPTIC AKRON
PHYSICAL REVIEW          VOL 19 PAGE 58 MAJOR LATEX AKRON CRUDE INDIA GROUP
PHYSICAL REVIEW          VOL 24 PAGE 49 FORCE RAYON RADIO LATEX INDIA METAL
EINSGANGENSCHAFT J.      VOL 97 PAGE 67 TREAD SALES PECOS METAL STEEL BUTYL
EINSGANGENSCHAFT J.      VOL 20 PAGE 91 PLANT TYREX SALES CURES COSTS NYLON
PROC BRITISH C. E.       VOL 71 PAGE 81 GUIDE PHONO ATMOS LARGE CURES INDIA
PROC BRITISH C. E.       VOL 85 PAGE 11 AUDIT MAINE AUDIT LARGE AKRON SALES
PROC BRITISH C. E.       VOL 97 PAGE 94 TREAD COVER TREAD LARGE MAJOR TURBO
PROC BRITISH C. E.       VOL 50 PAGE 13 OZONE TABLE NYLON TIRES QUEUE AUDIO
PROC BRITISH C. E.       VOL 75 PAGE 70 LARGE PLANT IMAGE SOLID TYREX CRUDE
PROC BRITISH C. E.       VOL 51 PAGE 49 CYCLE RAYON NYLON SOLID GUIDE STATE
RUBBER CHEMISTRY         VOL 64 PAGE 46 STUDY AKRON TABLE SALES TIRES GROUP
RUBBER CHEMISTRY         VOL 72 PAGE 30 GROUP OPTIC LATEX LATEX RAYON TUBES
RUBBER CHEMISTRY         VOL 39 PAGE 59 IMAGE INDIA OZONE CYCLE STEEL CURVE
RUBBER CHEMISTRY         VOL 83 PAGE 50 VIDEO OZONE AUDIT LARGE OPTIC PECOS
RUBBER CHEMISTRY         VOL 49 PAGE 51 RAYON CYCLE STUDY QUEUE GUIDE IMAGE
RUBBER CHEMISTRY         VOL  9 PAGE 35 INDIA AUDIT CURVE LARGE TYREX MAINE
RUBBER CHEMISTRY         VOL 43 PAGE 96 ATMOS AKRON NYLON CRUDE OPTIC TUBES
RUBBER CHEMISTRY         VOL 87 PAGE 18 RECAP QUEUE RECAP LATEX SPEED ATMOS
J. AM. CHEM. SOC.        VOL 82 PAGE 60 RADIO IDAHO VIDEO METAL LATEX PECOS
J. AM. CHEM. SOC.        VOL 62 PAGE 27 TURBO CURVE OZONE METAL TABLE COVER
J. AM. CHEM. SOC.        VOL 72 PAGE 18 GROUP QUEUE SMALL CURES SALES IDAHO
J. AM. CHEM. SOC.        VOL 73 PAGE 91 PECOS TYREX SMALL METAL STEEL STUDY
J. AM. CHEM. SOC.        VOL 51 PAGE 61 CYCLE MAINE MAJOR CYCLE TREAD AUDIO
J. AM. CHEM. SOC.        VOL 33 PAGE 93 VIDEO ATMOS BLIMP TIRES LATEX IMAGE
J. AM. CHEM. SOC.        VOL 33 PAGE  9 VIDEO INDIA TUBES SMALL VIDEO BUTYL
J. AM. CHEM. SOC.        VOL 35 PAGE 39 AUDIT EMAGE SPEED SALES FORCE SOLID
J. AM. CHEM. SOC.        VOL 98 PAGE 35 NYLON AUDIT LARGE LATEX LATEX OPTIC
J. AM. CHEM. SOC.        VOL 54 PAGE  8 TUBES LATEX CYCLE METAL PORUS PHONO
J. AM. CHEM. SOC.        VOL 85 PAGE 48 AUDIT NYLON TYREX LARGE TABLE PHONO
J. AM. CHEM. SOC.        VOL 45 PAGE 15 TEXAS STEEL CURVE TIRES TIRES TREAD
INTERNAL DOC. FILE       VOL 35 PAGE  4 AUDIT TUBES INDEX TIRES FORCE SALES
INTERNAL DOC. FILE       VOL  3 PAGE 83 CURES VIDEO TYREX TIRES SPEED COVER
INTERNAL DOC. FILE       VOL 96 PAGE 86 AKRON BLIMP AUDIT TUBES STUDY TURBO
INTERNAL DOC. FILE       VOL 43 PAGE 13 ATMOS TABLE GROUP CYCLE IMAGE COSTS
INTERNAL DOC. FILE       VOL 19 PAGE 79 MAJOR PORUS AUDIO LARGE INDEX BLIMP
INTERNAL DOC. FILE       VOL 40 PAGE 56 BUTYL SOLID SOLID CURES GROUP RADIO
INTERNAL DOC. FILE       VOL 81 PAGE 39 PHONO IMAGE TUBES TUBES MAJOR RECAP
```

Sample Input Data

ILLUSTRATIVE SOLVED EXAMPLE BY THE COMPUTER

THE 1620 IS NOW GOING TO DEMONSTRATE TO YOU AN
INFORMATION RETRIEVAL APPLICATION

YOUR OPERATOR HAS A DECK OF CARDS WHICH ARE
REPRESENTATIVE OF A LIBRARY FILE. EACH CARD
CONTAINS INFORMATION ON ONE LIBRARY ARTICLE
THIS INFORMATION CONSISTS OF -
TITLE VOL NO. PAGE NO. 6 KEYWORDS

YOUR OPERATOR WILL TELL THE 1620 WHAT LOGICAL
COMBINATION OF KEYWORDS TO SEARCH UPON.
SEARCH FOR
INDIA OR PECOS AND CURVE AND TYREX OR GLOPP AND LARGE BNT QUEUE
VERIFY SEARCH AS
INDIA OR PECOS AND CURVE AND TYREX OR GLOPP
AND LARGE BNT QUEUE

RUBBER CHEMISTRY VOL 9 PAGE 35

END OF THIS SEARCH. 1 FOUND

Keyed By The Operator

TRY A NEW SEARCH PATTERN

SEARCH FOR
GIBBERISH
PLEASE RETYPE THE SEARCH PATTERN. YOUR
TYPED LINE ABOVE DOES NOT CONFORM TO
MY SPECIFICATIONS

SEARCH FOR
RECAP OR TYREX BNT RAYON
VERIFY SEARCH AS
RECAP OR TYREX BNT RAYON

EINSGANGENSCHAFT J. VOL 20 PAGE 91
PROC BRITISH C. E. VOL 75 PAGE 70
RUBBER CHEMISTRY VOL 9 PAGE 35
RUBBER CHEMISTRY VOL 87 PAGE 18
J. AM. CHEM. SOC. VOL 73 PAGE 91
J. AM. CHEM. SOC. VOL 85 PAGE 48
INTERNAL DOC. FILE VOL 3 PAGE 83
INTERNAL DOC. FILE VOL 81 PAGE 39

A Demonstrative Symbolic Program for Information Retrieval:

```
GO      RCTY
        WATY COM1
        RCTY
        WATY COM9
        H
        BNC4 STT
        RCTY
        WATY COM2
        RCTY
        WATY COM3
        RCTY
        WATY COM4
        RCTY
        WATY COM5
        RCTY
        WATY COM6
STT     RCTY
        WATY COM7
BEG     RCTY
        WATY COMN
        RCTY
```

```
00402 34 00000 00102
00454 30 01687 00100
00426 34 00000 00102
00438 39 01783 00100
00450 48 00000 00000
00462 47 00594 00400
00474 34 00000 00102
00486 39 01837 00100
00498 34 00000 00102
00510 39 01915 00100
00522 34 00000 00102
00534 39 01977 00100
00546 34 00000 00102
00558 39 02039 00100
00570 34 00000 00102
00582 39 02125 00100
00594 34 00000 00102
00606 39 02187 00100
00618 34 00000 00102
00630 39 02255 00100
00642 34 00000 00102
```

```
        TFM   E
        RATY  INPUT
        SF    INPUT-1
        TFM   C,1
RD      BLC   NAV
        BC1   MC
        RACD  DATA
        SF    DATA-1
        SF    DATA+13
        SF    DATA+27
        SF    DATA+59
        SF    DATA+67
        SF    DATA+71
        BC2   TIT
        C     INPUT+30,DATA+58
        BNE   RD
        RCTY
        WATY  DATA
        AM    E,1,10
        BNLC  RD
        RCTY
        WATY  REP
        H
        B     BEG
TIT     C     INPUT+86,DATA+158
        BNE   RD
```

```
00654 16 03030 00000
00666 37 02277 00100
00678 32 02276 00000
00690 16 02372 00001
00702 46 01590 00900
00714 46 01290 00100
00726 37 02375 00500
00738 32 02374 00000
00750 32 02388 00000
00762 32 02402 00000
00774 32 02434 00000
00786 32 02442 00000
00798 32 02446 00000
00810 46 00942 00200
00822 24 02307 02433
00834 47 00702 01200
00846 34 00000 00102
00858 39 02375 00100
00870 11 03030 00001
00882 47 00702 00900
00894 34 00000 00102
00906 39 02947 00100
00918 48 00000 00000
00930 49 00618 00000
00942 24 02363 02533
00954 47 00702 01200
```

```
        CM    C,1
        BE    SUB
        C     CODE,DATA+12
        BNE   SUB
        C     CALL,DATA+26
        BNE   SUB
        RCTY
        TBTY
        TDM   DATA+67
        TD    DATA+68,D
        WATY  DATA+28
        BNLC  RD
        RCTY
        WATY  REP
        H
        B     BEG
SUB     RCTY
        WATY  DATA
        AM    E,1,10
        TF    CODE,DATA+12
        TF    CALL,DATA+26
        AM    C,1,10
        BNLC  RD
```

```
00966 14 02372 00001
00978 46 01158 01200
00990 24 02551 02387
01002 47 01158 01200
01014 24 02565 02401
01026 47 01158 01200
01038 34 00000 00102
01050 34 00000 00108
01062 15 02442 00000
01074 25 02443 02537
01086 39 02403 00100
01098 47 00702 00900
01110 34 00000 00102
01122 39 02947 00100
01134 48 00000 00000
01146 49 00618 00000
01158 34 00000 00102
01170 39 02375 00100
01182 11 03030 00001
01194 26 02551 02387
01206 26 02565 02401
01218 11 02372 00001
01230 47 00702 00900
```

```
        RCTY
        WATY  REP
        H
        B     BEG
MC      RACD  MSTR
        RACD  CON
        BC3   AUTH
        SF    CON+27
        C     CON+114,INPUT+86
        BNE   RD
        BE    WR
AUTH    SF    MSTR+27
        C     MSTR+78,INPUT+50
        BNE   RD
WR      TDM   MSTR+157
        TD    MSTR+158,D
        RCTY
        WATY  MSTR+2
        RCTY
        TBTY
        TDM   CON+115
        TD    CON+116,D
        WATY  CON+28
        AM    E,1,10
        BNLC  RD
```

```
01242 34 00000 00102
01254 39 02947 00100
01266 48 00000 00000
01278 49 00618 00000
01290 37 02567 00500
01302 37 02729 00500
01314 46 01374 00300
01326 32 02756 00000
01338 24 02843 02363
01350 47 00702 01200
01362 46 01410 01200
01374 32 02594 00000
01386 24 02645 02327
01398 47 00702 01200
01410 15 02724 00000
01422 25 02725 02537
01434 34 00000 00102
01446 39 02569 00100
01458 34 00000 00102
01470 34 00000 00108
01482 15 02844 00000
01494 25 02845 02537
01506 39 02757 00100
01518 11 03030 00001
01530 47 00702 00900
```

```
        RCTY
        WATY REP
        H
        B    BEG
NAV     RCTY
        CM   E,,10
        BNE  M
        WATY COM8
        RCTY
M       WATY REP
        H
        B    BEG
COM1    DAC  48,TO KNOW ABOUT SW. SETTING TURN SW.4 ON AND PUSH@
COM9    DAC  27,START OTHERWISE PUSH START@
COM2    DAC  39,SW.NO.   WHAT YOU KNOW   WHAT YOU NEED@
COM3    DAC  31,ALL OFF  SUBJECT          REFF.@
COM4    DAC  31,1.ON     TITLE            REFF.@
COM5    DAC  43,2.ON     TITLE            CONTENTS OF REFF.@
COM6    DAC  31,1.,3.ON  AUTHER           REFF.@
COM7    DAC  34,THE MACHINE IS READY FOR YOUR JOB@
COMN    DAC  11,SEARCH FOR@
INPUT   DAS  1
        DAS  45
C       DS   5
DATA    DAS  1
```

```
01542 34 00000 00102
01554 39 02947 00100
01566 48 00000 00000
01578 49 00618 00000
01590 34 00000 00102
01602 14 03030 00000
01614 47 01650 01200
01626 39 02891 00100
01638 34 00000 00102
01650 39 02947 00100
01662 48 00000 00000
01674 49 00618 00000
01687 00048
01783 00027
01837 00039
01915 00031
01977 00031
02039 00043
02125 00031
02187 00034
02255 00011
02277 00001
02279 00045
02372 00005
02375 00001
```

```
          DAS  80                                    02377 00080
D         DAC  1,@                                   02537 00001
CODE      DS   14                                    02551 00014
CALL      DS   14                                    02565 00014
MSTR      DAS  1                                     02567 00001
          DAS  80                                    02569 00080
CON       DAS  1                                     02729 00001
          DAS  80                                    02731 00080
COM8      DAC  28,REFFERENCE IS NOT AVAILABLE@       02891 00028
REP       DAC  40,YOUR JOB IS OVER FOR ANOTHER PUSH START@
                                                     02947 00040
E         DS   5                                     03030 00005
          DEND GO                                    00402

END OF PASSII
00402 GO      00594 STT     00618 BEG     00702 RD      00942
   TIT
01158 SUB     01290 MC      01374 AUTH    01410 WR      01590
   NAV
01650 M       01687 COM1    01783 COM9    01837 COM2    01915
   COM3
01977 COM4    02039 COM5    02125 COM6    02187 COM7    02255
   COMN
02277 INPUT   02372 C       02375 DATA    02537 D       02551
   CODE
02565 CALL    02567 MSTR    02729 CON     02891 COM8    02947
   REP
03030 E
```

Sample Input Data

```
  596 519.92CONCEPT                    21 1LINEAR PROGRAMMING METHODS
  596 519.92COMPARISON                  21 2LINEAR PROGRAMMING METHODS
  596 519.92SIMPLEX                     53 3LINEAR PROGRAMMING METHODS
  596 519.92METHOD                      53 4LINEAR PROGRAMMING METHODS
  596 519.92SHORT CUT                  151 5LINEAR PROGRAMMING METHODS
  596 519.92CHECK                      151 6LINEAR PROGRAMMING METHODS
  596 519.92COMPUTATION                151 7LINEAR PROGRAMMING METHODS
  596 519.92VARIABLE                   232 8LINEAR PROGRAMMING METHODS
  596 519.92RESOURCE                   232 9LINEAR PROGRAMMING METHODS
  596 519.92RESTRICTION               2321 0LINEAR PROGRAMMING METHODS
  596 519.92RESOURCE                  2321 1LINEAR PROGRAMMING METHODS
  596 519.92MAP                       2321 2LINEAR PROGRAMMING METHODS
  596 519.92DATA                      3081 3LINEAR PROGRAMMING METHODS
  596 519.92COMPUTER                  3081 4LINEAR PROGRAMMING METHODS
  596 519.92TRANSPORTATION            3321 5LINEAR PROGRAMMING METHODS
  596 519.92MODEL                     3321 6LINEAR PROGRAMMING METHODS
  596 519.92MATRIX                    3781 7LINEAR PROGRAMMING METHODS
  596 519.92INPUT                     4741 8LINEAR PROGRAMMING METHODS
  596 519.92OUTPUT                    4741 9LINEAR PROGRAMMING METHODS
  596 519.92ANALYSIS                  4742 0LINEAR PROGRAMMING METHODS
  596 519.92GAME                      4992 1LINEAR PROGRAMMING METHODS
  596 519.92PROGRAM                   5282 2LINEAR PROGRAMMING METHODS
  596 519.92VARIABLE                  5282 3LINEAR PROGRAMMING METHODS
  596 519.92INPUT                     5282 4LINEAR PROGRAMMING METHODS
  596 519.92COEFF                     5282 5LINEAR PROGRAMMING METHODS
  596 519.92NONLINEAR                 5542 6LINEAR PROGRAMMING METHODS
  596 519.92RISK                      5542 7LINEAR PROGRAMMING METHODS
  596 519.92PROGRAM                   5542 8LINEAR PROGRAMMING METHODS
 5945 519.92UPPER                     3682 8LINEAR PROGRAMMING AND EXTENTION
 5945 519.92GAME                      2772 5LINEAR PROGRAMMING AND EXTENTION
 5945 519.92PIVOTING                   173 1LINEAR PROGRAMMING AND EXTENTION
 5945 519.92VECTOR                     177 2LINEAR PROGRAMMING AND EXTENTION
 5945 519.92SPACE                      177 3LINEAR PROGRAMMING AND EXTENTION
 5945 519.92MATRIX                     183 4LINEAR PROGRAMMING AND EXTENTION
 5945 519.92INVERSE                    189 5LINEAR PROGRAMMING AND EXTENTION
 5945 519.92BOUND                     3682 9LINEAR PROGRAMMING AND EXTENTION
 5945 519.92CONCEPT                     1 7LINEAR PROGRAMMING AND EXTENTION
 5945 519.92ORIGIN                     12 8LINEAR PROGRAMMING AND EXTENTION
 5945 519.92INFLVENCE                  12 9LINEAR PROGRAMMING AND EXTENTION
 5945 519.92LINEAR                    691 0LINEAR PROGRAMMING AND EXTENTION
 5945 519.92L.P.                        1 6LINEAR PROGRAMMING AND EXTENTION
 5945 519.92EQN                       691 1LINEAR PROGRAMMING AND EXTENTION
 5945 519.92INEQUALITY                691 2LINEAR PROGRAMMING AND EXTENTION
 5945 519.92SYSTEM                    691 3LINEAR PROGRAMMING AND EXTENTION
 5945 519.92SIMPLEX                   941 4LINEAR PROGRAMMING AND EXTENTION
 5945 519.92METHOD                    941 5LINEAR PROGRAMMING AND EXTENTION
 5945 519.92PROOF                    1201 6LINEAR PROGRAMMING AND EXTENTION
 5945 519.92SIMPLEX                  1201 7LINEAR PROGRAMMING AND EXTENTION
 5945 519.92ALGORITHM                1201 8LINEAR PROGRAMMING AND EXTENTION
```

```
5945 519.92ALGORITHM          12018LINEAR PROGRAMMING AND EXTENTION
5945 519.92DUALITY            12019LINEAR PROGRAMMING AND EXTENTION
5945 519.92THEOREM            12020LINEAR PROGRAMMING AND EXTENTION
5945 519.92MULTIPLIER         21121LINEAR PROGRAMMING AND EXTENTION
5945 519.92VARIANT            24122LINEAR PROGRAMMING AND EXTENTION
5945 519.92SIMPLEX            24123LINEAR PROGRAMMING AND EXTENTION
5945 519.92ALGORITHM          24124LINEAR PROGRAMMING AND EXTENTION
5945 519.92TRANSPORTATION     29926LINEAR PROGRAMMING AND EXTENTION
5945 519.92VARIABLE           36827LINEAR PROGRAMMING AND EXTENTION
1696 519.92SHORT CUT           1 1LINEAR PROGRAMMING METHODS
1696 519.92COMPUTATION        25 2LINEAR PROGRAMMING METHODS
1696 519.92TRANSPORTATION     38 3LINEAR PROGRAMMING METHODS
1696 519.92GAME               49 4LINEAR PROGRAMMING METHODS
1696 519.92MAP                57 5LINEAR PROGRAMMING METHODS
1696 519.92GAMMAFN            90 6LINEAR PROGRAMMING METHODS
1696 519.92MATRIX            112 7LINEAR PROGRAMMING METHODS
1696 519.92ALGEBVECTOR       143 8LINEAR PROGRAMMING METHODS
1696 519.92SIMULATION        157 9LINEAR PROGRAMMING METHODS
1696 519.92VECTOR            1991 0LINEAR PROGRAMMING METHODS
```

```
5945 519.92DANTZIG                    PRINCETON UNV PRESSPRINCETON 63 13 6251
5945 519.92LINEAR PROGRAMMING AND EXTENTION
1696 519.92AMER                       MISR PRESS          CAIRO    62 25 200
1696 519.92LINEAR PROGRAMMING METHODS
2959 517.6 STEFFENSEN                 IOWA UNIV           IOWA     60 32 250
2959 517.6 INTERPOLATION
1101 517.5 KNOPP                      LONDON UNIV         LONDON   59 11 230
1101 517.5 THEORY OF FUNCTIONS
6164 517.5 FARRELL                    MACMILLAN CO
6164 517.5 GAMMA AND BETA FUNCTIONS
2225 517.5 THIELMAN                   LONDON UNIV         LONDON   53 51 520
2225 517.5 THEORY OF FUNCTIONS OF REAL VARIABLES
3182 517.6 TODD                       LIBRARY OF CONGRESSN.Y.      62 23 520
3182 517.6 SURVEY OF NUMERICAL ANALYSIS
3170 517.6 KNOPP                      LONDON UNIV         LONDON   62 14 370
3170 517.6 PRACTICAL ANALYSIS
 596 519.92HEADY                      IOWA COLLEGE PRESS  IOWA     58 12 597
 596 519.92LINEAR PROGRAMMING METHODS
```

Illustrative Solved Examples By The Computer

```
 KNOW ABOUT SW. SETTING TURN SW.4 ON AND PUSH
ART OTHERWISE PUSH START
.NO.   WHAT YOU KNOW    WHAT YOU NEED
L OFF  SUBJECT          REFF.
ON     TITLE            REFF.
ON     TITLE            CONTENTS OF REFF.
,3.ON  AUTHER           REFF.
E MACHINE IS READY FOR YOUR JOB
ARCH FOR

    (x)
MERS
 5945  51992GAME                 27725LINEAR PROGRAMMAING AND EXTENTION

  596 519.92GAME                 49921LINEAR PROGRAMMING METHODS

 1696 519.92GAME                  49 4LINEAR PROGRAMMING METHODS

UR JOB IS OVER FOR ANOTHER PUSH START
EARCH FOR

                    (x)
 NEAR PROGRAMMING METHODSRS
 3182 517.6 TODD                LIBRARY OF CONGRESS  N.Y.     62 23 520

    LINEAR PROGRAMMING METHODS
UR JOB IS OVER FOR ANOTHER PUSH START


EARCH FOR

                    (x)
NDUSTRIAL ENGINEERINGRS
EFFERENCE IS NOT AVAILABLE
OUR JOB IS OVER FOR ANOTHER PUSH START


EARCH FOR
      (x)
NASSERRS
EFFERENCE IS NOT AVAILABLE
OUR JOB IS OVER FOR ANOTHER PUSH START


SEARCH FOR
KNOPPRS (x)
  1101 517.5 KNOPP              LONDON UNIV      LONDON    59 11 230

          THEORY OF FUNCTIONS
 3170 517.6 KNOPP               LONDON UNIV      LONDON    62 14 370
```

```
SEARCH FOR
LINEAR PROGRAMMING METHODSRS(x)

    596 519.92CONCEPT                21 1LINEAR PROGRAMMING METHODS

                COMPARISON          21
                SIMPLEX             53
                METHOD              53
                SHORT CUT          151
                CHECK              151
                COMPUTATION        151
                VARIABLE           232
                RESOURCE           232
                RESTRICTION        232
                RESOURCE           232
                MAP                232
                DATA               308
                COMPUTER           308
                TRANSPORTATION     332
                MODEL              332
                MATRIX             378
                INPUT              474
                OUTPUT             474
                ANALYSIS           474
                GAME               499
                PROGRAM            528
                VARIABLE           528
                INPUT              528
                COEFF              528
                NONLINEAR          554
                RISK               554
                PROGRAM            554
    1696 519.92SHORT CUT            10 1LINEAR PROGRAMMING METHODS

                COMPUTATION         25
                TRANSPORTATION      38
                GAME                49
                MAP                 57
                GAMMAFN             90
                MATRIX             112
                ALGEBVECTOR        143
                SIMULATION         157
                VECTOR             199
YOUR JOB IS OVER FOR ANOTHER PUSH START


SEARCH FOR                                  (x)
(LET S FOOL THE 1620) NOTHING. . .RS
REFFERENCE IS NOT AVAILABLE
YOUR JOB IS OVER FOR ANOTHER PUSH START
```

(x) Keyed by the operator.

4.4. Other Potential Equipment in the U.A.R.:

Information Retrieval, specially for scientific and technical information, is a matter of national concern and thus equipment available anywhere in the U.A.R. should be utilized in a coordinated and cooperative network.

This introduction is of importance because of the increasing equipment that is, and would be, available in the U.A.R. Faculty of Engineering, Alexandria University has an IBM 1620 and other equipments are in their way to other governmental institutions. For example IBM 1401 computer would be available in the "Central Mail Organization" at Cairo.

The 1401 is very much used in information retrieval and this can be illustrated as follows:-

The 1401 is a system that is available in a wide range of machine configuration. The basic card system consists of the 1401 processing unit the 1402 card read punch and the 1403 printer. The 1401 has a magnetic core memory that is used for storing data and the machine steps required to carry out the desired logic functions (program.) In one card pass the 1401, because of its large memory, can make test & decisions that would require multiple passes through machines with control panel logic. Instructions stored in core memory are executed much faster than those stored in a wire panel. Steps controlling the card reader, card punch, and printer are also stored in memory.

An IBM 1401 can be programmed to perform selection. A program has been prepared for distribution that will permit several options of query response:

a) answers can be printed.

b) " " " punched.

c) frequency counts of the criteria desired can be prepared.

Control cards are used to specify the search arguments and criteria.

Index data may be stored in large - capacity random access disk files connected to an IBM 1401 data processing system.

Access to any of the keyword records is reduced to a fraction of a second. Automatic reference to data on titles, authers and sources permits printing a complete bibliography quickly.

Automatic procedure for indexing new documents save time and drudgery. A system can edit all new entries against the stored dictionary. Documents numbers are automatically listed under appropriate keywords, and new reference data is entered in the bibliography file as a part of the same operations.

Paper and magnetic tapes are additional input and output media that may be used with a 1401 system. Magnetic tape, offers the fastest way to read & write data. Either operation may be done at a rate of many thousands of characters per second.

These 1401 procedure & system usage are applicable to large computers, whose extensive memories permit processing greater segments of files, with additional program logic, at faster speed.

4.4.1. Some new applications of the 1401 computer:

a. The system devised by Newkirk(1) ;

The system performs literature searches on punched cards decks representing library information or document collections which have been encoded by coordinate indexing techniques. It used the inverted file organization with eighteen document numbers per card. A maintenance program produces new and updates old keyword cards automatically. There are two prints programs.

---

(1) IBM Corporation, 1026 Quarier Str.
Chaleston west virginia, U.S.A. (1401-10.3.006)

One will print one document number per line and the other will print selected information from a bibliography card file. Boolean operatives-and, and not, or - are used in document number comparisons. A 4K 1401 with the high, low, equal compare feature is required. Modifications necessary to use other sizes of core and index registers are discussed.

b. The system devised by Donald Herman(1):

The program will select records from tape that satisfy ranges, and, and not, or, or not conditions. Free form inquiry format is used and multiple inquiries may be processed with one pass of the master tape. Significant features are the ability to make multiple inquiries, the use of free form inquiry, and the omission of any necessity for control cards 4K 1401 two tape drives advanced programming package HI - Low - Equal compare feature 1402 card reader 1403 printer.

c. The system devised by Luke and Pittard (2) :

This system is a general one designed for use on an IBM 1401 with 4K memory and 2-3 magnetic tape derives. Each file item is organized to permit storage of up to 6000 characters of textual information. Inform-2 (the name of this system) employs 9 character keyword match logic together with the optional use of other specific criteria - auther source, date range and subject. A file item may be indexed using up to 54 unlinked keywords. Quaries may use a like number of keywords, combining various and/or relationships. Inform-2 may be used individually or in conjunction with the IBM 1401 KWIC index system and/or the IBM 1401 SD1-3 system. The 4K 1401 must, in addition to two magnetic tape drives, have the HI-LO-Equal compare feature, a 1403 printer/132/, the advanced programming features and sense switches. The program is written in 1401 autocoder.

(1) Computing center, IBM corporation
1120 connecticut Avenue, N.W., Washington D.C. (1401 - 10.3.009)

(2) IBM corporation, washington system center/FSD/- 7220 wisconsin Avenue, Bethseda 14, Maryland (1401 - 10.3042)

Chapter V

Some Conclusions and Recommendations

It is an undisputable fact nowadays that the progress specially in the fields of science and technology necessitates the availability and accessability of information to research-workers promptly and comprehensively from the flood of literature which have reached an unprecedented degree in complexity & quantity.

Libraries and individual organizations have proved to be unable to cope with the present multi-dimentional & diversified needs & requests of scientists and research-workers.

Having this fact in mind the group tried in this first study to investigate some of the problems relating to both analytic and manipulative procedures in information retrieval. Some conclusions influenced by this study can be mentioned:

1. A comparison between traditional library techniques and nonconventional retrieval systems showed clearly that nonconventional systems whether manual or mechanized increase the capability of documentalists, scientists and Librarians to organize, analyse, assemble, translate, store, process, retrieve and dessiminate informations.

2. There are different techniques and procedures for analysis and manipulation of information. The optimum design and choice of a particular information retrieval system depends on certain criteria and unit operations pertinant to the needs of the users of the system, coverage of subjects, depth of analysis of these subjects, precision of the service offered as well as the costs that an institution is willing to pay in order to respond to its user's needs.

3. The study shows also that the costs of information retrieval systems are usually high in terms of library budgets. This is true when manual information retrieval systems (such as hand-sorted-punched cards) are used. Costs of mechanized systems, including those using computers(1), are apparently, much higher than the traditional library system although these costs diminish gradually with the increase of the users of the system. It should also be mentioned that availability of certain information at a particular time, with promptness, accuracy and comprehensiveness may justify these high costs.

4. From the study it can be concluded that there are two phases in information retrieval, namely analytic and manipulative. The analytic phase includes some unit operations performed primarily by subject specialists. The manipulative phase includes other unit operations which can be performed by scientists, engineers, documentalists and librarians, after being trained on information retrieval techniques. Since analysis of information, specially when performed in great depth, is time consuming on the part of subject specialists, thus documentation centers tend to buy (or exchange) magnetic tapes, disks, etc. from similar centers and use this prestored and preanalysed information with its own equipment.

---

(1) example of the average monthly rents of some digital computers are as follows:

IBM 7090	$64,000
General Electric 225	8,000
IBM 1401	7,500
IBM 1620	1,600

5. The experiment conducted at the Operations Research Center for analysis of the contents of books with a greater depth than ordinarily performed has showed, a considerable potential for the use of the available equipment. The experiment also showed the possibility of future utilization of these equipment for operational information retrieval systems in specialized fields.

6. It should be stated that the experiment conducted on the analysis of books has pointed out some problems and dimensions that should be tackled and overcome in the future experiments. First is the control of terminology and subject headings used and the application of different search strategies for answering multidimensional questions (through coordinate indexing and combination processes). Second is the design of IBM cards to serve as aspect cards and could contain a greater number of documents pertaining to that aspect. If these problems and dimensions would be tackled as a more developed manipulative technique and if this is combined with the use of indices of books (after uniform normalization). As a guide in the analytic procedures the result would be a more economic information retrieval system.

7. This report includes some theoretical discussion of the information and documentation techniques such as classification, abstracting, indexing as well as the unit operations and criteria involved in the choice and design of information retrieval systems. Other theoretical discussions of pertinant techniques in the machine language (from the syntactic, semantic and phonetic points of view) and search strategy to answer multi-dimensional questions as well as codes and notations would be included in the following reports.

Recommendations

1. Published and unpublished literature needed for persuit of research has reached an unprecedented degree in quantity and in complexity and thus it is no longer possible for individual organizations to acquire, process and dessiminate information to all persons who need it. Consequentely cooperative efforts are required on behalf of interested organizations in the country. The cooperation is assumed to cover both phases of information techniques, namely the analytic and manipulative ones.

2. Equipment used for information retrieval is usually expensive in terms of traditional budgets allocated for information & library services. Thus equipment available in the U.A.R., or would be available in the future, should be made accessible to information retrieval specialists, at least part-time, in order to perform the information and documentation work. Again, coordination of the use of these equipments may be achieved according to the kinds of unit operations involved and information systems applied.

3. The group who started this study should continue their studies in order to be able to follow up advanced techniques in the field of information retrieval and also to be able to develop operational systems utilizing actual and future equipment that may be available in the U.A.R.

4. The field of information retrieval is undergoing continuous progress, specially in United States, Europe and the Soviet Union, therefore missions-for short and long periods-should be sent to these countries. Members participating in these missions together with the group who conducted the present study are recommended to perform as consultants to interested organizations and also as specialists of information

retrieval entitled with promotion and dessimination of information techniques in the country..

5. Information science and information retrieval have become a formal study in many universities abroad. This study leading to Master and Doctor degrees is being offered for graduate students from different disciplines (science, engineering, economics, psychology, business administration, etc). The graduate study of librarianship and documentation being conducted at the Faculty of Arts, Cairo University, should be developed. Its curriculum should cope with the present progress in the information field and also to offer degrees in this field to graduate students from other disciplines.

6. The National Information and Documentation Center should have its own appropriate equipment and devices for information retrieval . It may begin with equipment used for manual information retrieval systems. Fully mechanized systems, including a computer, are also recommended to be acquired by the Center in the longrun, although part-time use of the 1620 and 1401 computers and other equipment available in the U.A.R. may be appropriate in the present stage.

7. The National Information and Documentation center is the national body entrusted with coordination and technical assistance to specialized documentation retrieval services in the U.A.R. Thus different institutions interested in the field should cooperate with the center to achieve maximum accessability of information among researchers in our country.

References

1. Garrido, "Superimposable punched cards as a Means of Reference to periodicals, Unesco Bull.Libr., vol. XII, No. 10, October 1958.

2. Hyslop, M.R. "Answering Metallurgical Questions by Machines, "Special Libraries 50, 435 - 42 November 1959.

3. IBM
   General Information Manual
   Storage Retrieval and Dessimination of Information
   B20 - 8095, 1962

4. IBM
   Literature on Information Retrieval & Machine Translation.
   320 - 1710, Nov., 1962.

5. IBM
   Reference Manual
   Index Organization for Information Retrieval
   C20 - 8062, 1961

6. Kent, Allen.
   Text book on mechanized information retrieval, N.Y., John Wiley, 1962.

7. Operating Instruction for Literature Searching Programs, a paper by Western Reserve University, Center for Documentation 1962.

8. Whaley, F.R. "The Use of Collator in an Inverted File Index" Special Libraries 1962, Vol 53 No. 2 P. 65.- 73.

٩ ــ الدكتور أحمد بدر ، التوثيق الآلى ، ... ثورة فى عالم المكتبات ، مجلة المكتبة العربية المجلد الاول ــ العدد الرابع سنة ١٩٦٤ .