

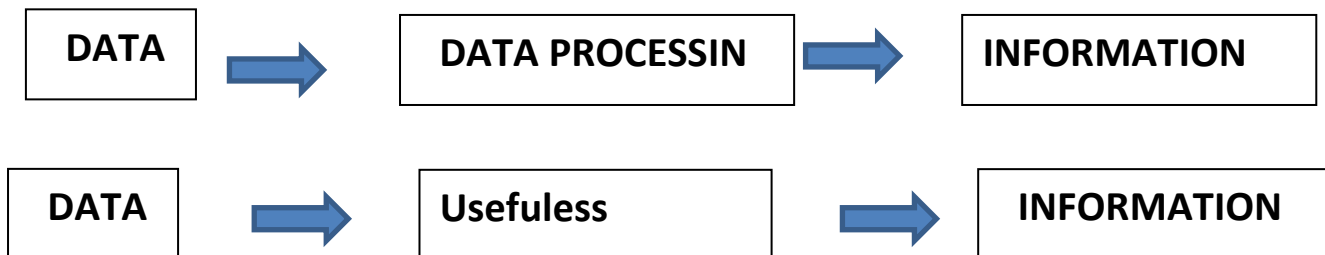
CHAPTER ONE

DEFINITIONS :

DATA : Is the raw material that includes any fact , figures , letters , words , or symbols to represent an object , idea , condition or situation such as medical statistics or production .

INFORMATION : Consist of selected data to be organized with respect to user , problem , time , place and function .

DATA PROCESSING : The main function of data processing is the converting of data information



REASONS FOR THE NEED OF DATA PROCEEING

Data processing methods are developed as a result of :

- 1- Growth of the volume of data .
- 2- The changing nature of data .
- 3- The technological progress .
- 4- The urgent need for better ways of handling data .

REASONS FOR THE NEED OF IMPROVED DATA PROCESSING

Improved data processing techniques are developed as a result of :

- 1- Growing volume of data :

The growing size and complexity of business organization have led to the growth of a large amount of paperwork especially after the second world war . This volume of data requires processing that cannot be handled by clerical staff using traditional methods .

2- Expensive clerical cost .

As a result of the growing volume of paper work, the number of clerical workers has been increased too , accompanied by rising wages that increase the clerical cost rapidly .

3- Accuracy requirements .

The clerical work is likely to consist of the routine and repetitive handling of large quantities of data which could lead to carelessness , that increases the chance of error . This resulted in the search for improved data processing techniques.

4- The need for more timely information.

The modern business activities demand accurate and rapid responses to changing conditions . On the other side , to achieve effective control over large organizations require that executives make daily or hourly decisions basses on up – to – date information .

FUNCTIONS (STEPS) OF DATA PROCESSING

1. ORIGINATION OF DATA :

Data origination in various forms called source document which could be hand written or typewritten such as invoices or purchase orders .

Source documents : are important because :-

- 1- They are the basis for all actions
- 2- Provides verifications for all transactions

2. DATA RECORDING :

Data is recorded in source form to be handled easily . The recording process includes these functions :

1- Editing :

Selecting the significant data and eliminating data that does not need processing .

2- Coding :

Reducing the amount of data by converting it to symbols including :
 Numbers.
 Alphabets .
 Alphanumeric symbols.

3- Classifying :

Systematically grouping data into classes according to its type ,
 location , age , price etc .

4- Copying and duplication :

The use of facsimile of data for distribution to more than one users .

5- Verifying :

Ensuring the storage of data into the processing system without errors.

6- Conversion :

Transforming data from one form to another.

3. MANIPULATION OF DATA

Converting data into a more useful form through the following procedures :

1- Sorting :

Arranging and selecting data according to :

1- Order or rank

2- Sequential order alphabetically or numerically .

2- Comparing And Analysing

To determine factors such as :

Relationship Nature

Order

Relative value

} of data

3- Calculating

Using arithmetical process (+, -, >, ×, ÷,) to convert data into a significant form ex: calculating the employee hours for payment.

4. SUMMARING AND REPORT PREPARATION

1- Summarizing :

- Condensing data into lists
- Condensing lists into totals
- Arranging data into categories

2- Report preparation

The output of data processing cycle is processed information recorded into output media such as reports ,books, magnetic tapes etc.

5. DATA COMMUNICATION

Communication : is the process of transferring data from one point to another point either :

- A. Internally by hand on machine
- B. Externally by mail

Today : Communication means electrical transmission of data by telephone , telegraph, microwave, ... etc.

6. DATA STORAGE

At the end of data processing cycle , data must be stored so that it is readily retrievable.

Organizations usually store data to design a storage system that will facilitate the retrieval of needed data.

METHODS OF DATA PROCESSING

1. Manual Method

To perform operations in the data processing cycle by hand and with the use of tools such as pens , files , folders , work sheet ... etc.

2. Mechanical method

To perform operations with the use of mechanical devices that requires manual aid such as typewriter, document control register, calculator ...

Advantages of mechanical over manual method

- 1- Eliminates recopying restoring of data.
- 2- Improves speed and accuracy of manual processing .
- 3- Eliminates human efforts in data processing by processing six operations automatically :
 - 1- Recording
 - 2- Sorting
 - 3- comparing
 - 4- calculating
 - 5- Summarizing
 - 6- Reporting

3. Electronic method

The use of computer to perform operations automatically by the movement of electrical impulses through programs rather than the movement of mechanical parts by hand .

Computer operations include :

- 1- INPUT : source document – codes – recording .
- 2- PROCESSING : make logical decisions .
- 3- OUTPUT : the result of data processing to be recorded on paper, magnetic tape, etc.

Advantages of the electronic method :

1. Speed of processing .
2. The processing of data is continuous while each machine performs only one operation .
3. Saving space with the use of compact equipment and storage .
4. Greater accuracy than other systems.
5. The capacity of computers and their versatility to complete tasks that could never performed with other systems .

CHAPTER TWO

The functions operated by machines :

1. Receive inputs after being converted from human readable into machine readable from through :
 - 1- Key boarding or key punching
 - 2- Optical scanning of printed pages
 - 3- Handing machines readable information as by continuous products of operation

2. Machine : Remember information by having stored in :
 1. Perforated :
 1. Unit records (cards)
 2. continuous records (paper tape)
 2. Magnetic :
 1. Tapes
 2. Discs
 3. Core memory
 3. Film :
 1. Unit record (mini card)
 2. Continuous record (micro film)

3. Machines compare and process the stored information by
 1. Sorting
 2. collating
 3. Internal processing such as :
 1. Addition
 2. Subtraction

4. Machines output information by :
 - 1- Printing out on sheet to be red by people .
 - 2- Activating a display through terminal for people to read .
 - 3- punching out for other machines to read :
 - 1- Unit record
 - 2- Continuous record
 - 4- Magnetic tapes for other machines to read.

PHYSICAL TOOLS

1- TABULATING CARD

1. Or punched card which is a unit record 7 1/4 X 3 1/4 inch.
2. Storing data in the form of punched rectangular holes .
3. Consist of eighty columns and 12 rows .
4. Information are recorded by columns as follow :
 - 1- one punch for numbers.
 - 2- two punch for alphabet characters .
 - 3- Three punches for special characters .
5. founded in 1887 and used in 1940 .

2-PEEK-A-BOO CARD

1. A unit record that is specialized in storing data in a machine readable from that permits further processing .
2. Each card is organized in a way that receive space for document numbers.
3. Each card represents a given subject or index term (entry) .
4. Arranged in a matrix form (10 × 10) or (100 × 100) matrix that provides (100 × 100) dedicated spaces.

PEEK – A- BOO CARD MACHINES

1. INPUT DEVICE .

Specialized in recording information on a card by drilling a hole to indicate the document number .

2. THE CARD READER

Identifies the positions numbers holes in the card visually

3. THE MARGINAL-HOLE PUNCHED CARD

A unit record which is specialized in recoding data by cutting down one of more of the drilled holes .

THE MARGINAL – HOE PUNCHED CARD MACHNES :

1. THE GROOVER

It is a hard punch device specialized in grooving (cutting down) the marginal hole punch cards.

2. THE SORTING NEEDLE

Specialist in searching marginal hole punched cards

3- PAPER TAPE

Continuous records which is specialist in recording (storing) data I machine readable form , numbers , alphabets , and symbols, are recorded in combinations of :

5- channels (5 holes)

6- channels (6 holes)

7- channels (7 holes)

8- channels (8 holes)

4- FLEXP WRITER

a. Which is specialist if recording data in paper tape .

b. Reading the recording data and printing it in a human readable form.

c. The punched tape is then used by other machines for further processing .

5- MAGNETIC TAPE

Continuous records which is specialist in storing data in computer processible from through the : 1- key board or through 2- Conversion of paper tape to magnetic tape. The data are recorded in binary form by using (1) to indicate magnetization and (0) for non-magnetization.

6- FILM RECORDS

Specialist in recording data in :

A. Binary form : where the data are represented by transparent or opaque spot on the film

B. As image form : where the entire page of printed data be recorded on film in reduced size where there are kinds of this form :

1- Continuous film (micro film)

2- Unit records film :

a. Mini card .

b. Microfiche .

c. Filmorex .

7- MIXED MEDIA :

1- MAGNETIC DISC .

a. It is a random – access media that read and write data directly.

b. Data are recorded in computer processible form .

c. Each disc face is divided into an assigned number of sectors .

d. Each disc face has from 50 – 100 inch circular tracks .

2-SORTER

Which is specialized in :

a. Arranging punched cards in a sequential order (alphabetically or numerically) by sensing the columns .

- b. Counting the cards
- c. Editing the punched information .
- d. translate punched information into human readable form in a report.

3-COLLATOR

Which is specializes in :

- a. Comparing two sets of punched cards to arrange them in a desire sequence.
- b. The machine can merge four files:
Two of them consist of matched card and other two of the unmatched cards .
- c. To ensure the sorting operations .

CHAOTER THREE

I- MECHANIZATION OF INPUT UNIT OPERATIONS

1. INPUT UNIT OPERATIONS

1. ACQUSTION

1. definition
2. Tasks of acquisition
3. Type of materials

a- Books b- Serials

1. Information required with acquisition
2. Information automatically produced
3. List and reports generated by the acquisition system

II. ANALYSIS

III. VOCABULARY CONTROL

IV. RECORDING RESULTS OF ANLYSIS

V. PHYSICAL TOOLS USED TO PERFORM INPUT UNIT OPERATIONS :

1. Marginal hole punched card
2. Peek a Boo card
3. Tabulating card
4. Film

1. ACQUSITON :

1. Involves the :

- Location

- Selection
- Ordering
- Receipt of source material for a collection

2. Tasks of acquisition

1. To determine the current and future needs of the users of information retrieval system .
2. To formulate the acquisition policy of acceptance materials based on subject coverage or publication type of source.
3. To compare the available or incoming sources with the policy to determine which material should be included in the information retrieval system .
4. To establish the procedures that ensure that the required materials are ordered and received .

3. Types of materials :

A. Books:

1. information required :
 1. author
 2. Title
 3. Publisher (code) and place
 4. date
 5. volume
 6. edition
 7. price
 8. accession
 9. ISBN
 10. number of copies

11. name or code of Jobber

2. Information automatically produced :

1. Purchase order
2. Checks
3. in order list
4. new books
5. pocket lable
6. Book card
7. expenditure records

3. List and reports generated by the system

1. Shelf list
2. Specialized bibliographies
3. Overdue list
4. Book inventories
5. Weeding reports
6. User studies
7. Fine lists / notes

B. SERIAL CONTROL

Serial represents difficult problems because :

1. Volume of serials
2. Bibliographic instability
3. User requirements

INFORMATION REQUIRED :

1. Title
2. Destination
3. Volume and number expected issue
4. Call no.
5. Language
6. ISSN

II. ANALYSIS

Analysis includes :

1. Perusal of source materials
2. Selection and analysis of source material to produce indexes, abstracts, and classification for an information retrieval system .
3. Today , computers are used and programmed to interpret the source materials contents automatically specially in two approaches :
 - A. The permuted index (KWIC index) .
 - B. The auto – abstracts .

III. VOCABULARY CONTROL

Vocabulary control involves :

Establishment of relationships among analyzed source materials to facilitate the identification of all source material in system relating to a given subject.

IV. RECORDING RESULTS OF ANALYSIS

Two systems for recording results :

1. Document system (or sequential file or main file or entry system) where each record represents a single document and used to records all characteristics relating to document.

2- Aspect system or inverted system (file) . Where each record represents a subject (or aspect or index term or descriptor or key word)

II. MECHANIZATION OF OUTPUT OPERATIONS

1. QUESTION ANALYSIS OF SEARCH STRATEGY DEVELOPMENT

It involves five functions which are :

- 1- Statement of question so that it can be examined.
- 2- Question analysis to identify its essential features.
- 3- Terminology control using terms to express the elements of the

question which could be :

Subject headings

Index terms

Descriptors

4- Question negotiation to adjust the structure of question in order to increase the likelihood that needed material is located.

5- Strategy determination :

To express the question in logical statement to get significant response :

$$A \times B \times C$$

2.CONDUCTING THE SEARCH

It involves manipulation of operations of the search mechanization in order to identify records in the file or to select a small number of documents from a collection of a large number of documents, based on terms relevant to a question.

PHYSICAL TOOLS USED IN CONDUCTIONG THE SEARCH

1- Marginal hole punched card

A. Recording the results of the analysis by grooving the holes.

B. Using sorting needle to select cards by inserting it in the hole position of file of cards that relates to the related subject.

C. All cards that represent documents containing the desired subject not held by the needle will drop out of the file.

2- Tabulating punched card

Can be used in both:

A. Document system : where each card represents a single document and a sorter can be used to select the desired cards .

B. Aspect system : where each card represents a single aspect or subject with the serial numbers of documents related to that subject then a collator is used to identify the desired document .

3- DELIVERY OF RESUKTS :

Which includes the provision of one or more of the following :

- a. Serial number of documents
- b. Bibliographic information
- c. An abstract of the document
- d. A copy of the original document

CHAPTER FOUR

PRINCIPLE OF SEARCHING

1. Steps of searching :

A. Recording of question

It is necessary to record the subject of the search for communication to the operation of an information system. The searcher must be familiar with :

- 1- The subject content of the file to be searched
- 2- The policy of analyzing the content of the record
- 3- The vocabulary control tools used in question analysis

B. Selection of clues from a question

Once a question has been formulated appropriate clues must be selected to serve as a basis for searching activity .

C. Organization of searching clues

Question must be analyzed in order to select clues that will be useful for formulating a strategy of search . These clues usually expressed in natural language and could be words or phrases and relationship among them .

D. Transforming of clues into language of searching system and coding system which could be :

- 1- Word indexes when the clues are selected from record during the analysis of the records . Example of word indexes are :
 - a. Concordances : An alphabetical list of words selected from the content of a book
 - b. Keyword in context (KWIC) or permuted index , list of terms

c. Natural language based system

2- Controlled indexes and classification which includes :

a. UNCODED

Where the selection of words or terms is based on the use of a subject authority list or a classification system

b- CODED

when a code is used to express the accepted terms in the authority list or a classification system

E. Strategies of Searchrch :

Once the clues is transformed into the language or codes of the system , it becomes possible to select the best search strategy that exploit the contents of a file of records in response to a question .

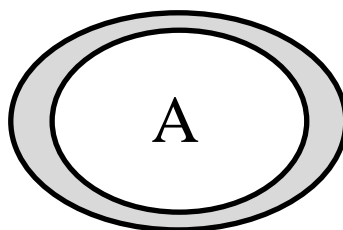
TYPES OF SEARCH STRATEGIES :

1- Single aspect or clue or term or keyword or subject search.

To identify all records in the file that have a single clue.

Example : all records in the file that have clue A

All records in the file that have the subject Orange .

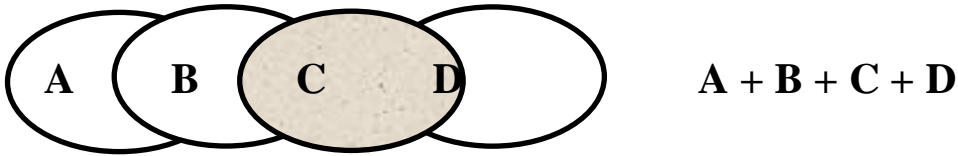


2- Strategy of the logical sum.

To specify two or more aspects or clues or subjects as points in a search and to identify all records in a file that are characterized by any of these clues .

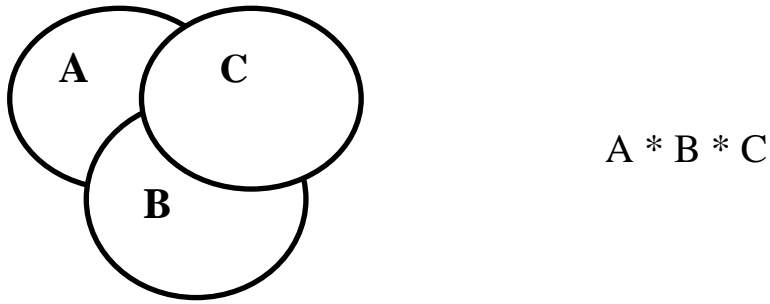
Example : all records that have clue A and / or B and / or C and / or D.

All records that have orange and / or Banana and / or Apple / or Apricot.
 All records authored by Allen.A.Kent or Allen S.Kent or Allen G.Kent
 or Allen C.Kent .



3- Strategy of the logical product

To identify all records in a file that have two or more clues.



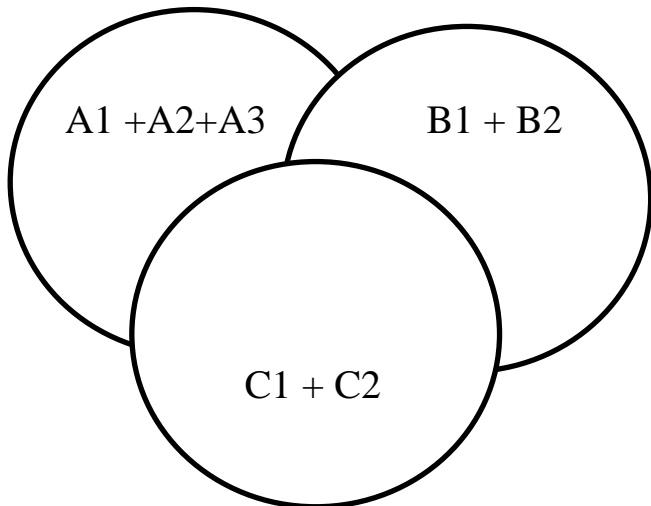
Example : all records that have A and B and C in Common.

All records authored by Allen , Kent that deal with information system in 1961.

All records about Orange and Banana and Apple.

4- Strategy of logical product of Logical Sum

To identify all records in a file that have one or more of several sets of clues in common.



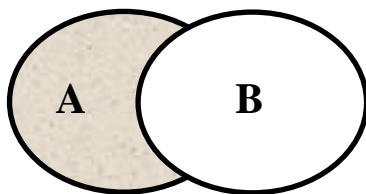
Example : $(A1 + A2 + A3) \times (B1 + B2) \times (C1 + C2)$

All records that have A1 or A2 or A3 AND B1 or B2 AND C1 or C2.

All records authored by Allen Kent or Allen S. Kent or Allen A. Kent deal with information retrieval or information system that are published in 1961 or 1962 .

5- Strategy of the logical Differences

To identify records in a file that are characterized by one or more clues and the absence of one or more clues



A – B

Example: All records that clue A which do not have clue B in Common

All records authored by Allen Kent (A) unless they were co-authored by John Daily (B)

6- Strategy of Sequence

To identify all records with one or more clues , but the clues must be found in record in an exact SEQUENCE

Example : All records that have clue (A) and (B) but clue (A) must precede clue (B)

$\langle A \times B \rangle$

7- Strategy of Greater than and less than

To identify records that contain numerical data .

Example : All records published between 1950 and 1960

> 1949

< 1961

Ch. Five

Codes and Notations

A code is any system of symbols in the communication process . Particularly a system that achieves some other desirable advantage over common language or numerical expression .

The sense of code involves the existence of two languages , a source language and a target language .

The Code is then the system of rules that enables messages in the source language to be transformed into a target language . (or code language) .

In the information retrieval field , two other have often been used interchangeably with the word code . These are notation and cipher .

The word notation will be used only to designate the particular set of symbols used to represent a code .

Why Establish a code or use a notation ?

Some of the obvious reasons for use of code and notation are as follows : -

- 1- To translate from a difficult – to – use – source language to a language that is easier to use for a particular purpose , or purposes .
- 2- To decrease the amount of space required to record information .
- 3- To supplement the information available in the source language .
- 4- To distinguish between alternative ideas or words that are not easily distinguished in the source language .
- 5-

Characteristics of codes and notation :

The development of systems for information retrieval has been approached from two points of view :-

- 1- The organization of a searching system for achieving effective and economical searches based on the information content of graphic records.

2- The organization of a searching system for achieving optimal use of a given tool . or set of tools .

With the first point of view , a system may be based on one or more of the principles discussed , the second point view implies that a predetermination has been made as to the toole , machine , or device that is to be used .

Natural language as a code

We mentioned ealier that from some points of view natural language is a code for ideas more accurately it should be considered to be one notation for such a code .

It has been found that the proportions of linguistic forms in a natural language remain quite constant for a given language , at a given time of its development , and for a sufficiently great number of observations . This characteristic of language makes it a relatively effective system , for the communication of thought by means of symbols .

Theoretically , the coding system that may be implied by language usually enables different messages to be distinguished easily and correctly . However , language is characterized by a change in the use of vocabulary that is unpremeditated , and follows the general tendency of the changes in nature , that is for the entropy of a system to approach a maximum .