	Master of Library and	**MLIS 202**
	Information Science	**Information Storage and Retrieval**
PUNJABI UNIVERSITY PATIALA	**UNIT No. 2**	

Department of Distance Education
Punjabi University, Patiala
(All Copyrights are Reserved)

Lesson No.:

LESSON NO. 2.1   **AUTHOR : R.S.R. VARALAKSHMI**

INDEXING : DEVELOPMENT AND TRENDS IN INDEXING

Structure :

2.1.1 Objective : This lesson introduces you to

the concept of indexing

the importance of indexing in information retrieval

the trends in the development of different indexing techniques.

2.1.2 Index and Indexing - Definition :

The word index is derived from the Latin word 'indicare' that means to indicate or point out the information where it is available. It is a location tools and helps to know the sources of information as well as its location. In other words it facilitates the identification or selection of desired documents containing the required information after they have been stored or shelved in a library.

Therefore index can be described as a list or systematic guide to the contents of a set of data, such as a periodical index or an index in a book. As you are aware, a back of the book index is an alphabetically arranged list of headings consisting of the personal names, places, and subjects treated in the work, with pages numbers to refer the reader to the point in the text at which information pertaining to the heading is found. Similar technique is being followed in libraries in preparing subject catalogue, a method of document representation that tells a user quickly the availability of an

item and where it is to be found. Hence index is prepared for the user to help him in finding information more quickly and easily. However, form librarian's point of view, it is not simply a list of the major terms in a publication. Rather, it is an organized map of the contents of a document, arranged to make the contents clearly visible and comprehensible to the user.

Index - Definition : According to Webster's Dictionary of Englihs Languages Index is "A usual alphabetical list that includes all or nearly all times. Considered of special pertineneces and fully of partially covered or merely mentioned in a printed or written work ... that gives with each item the place (as by page number) where it may be found in the work".

According to Encyclopedia of Library and Information Science an index is "a systematic guide to item contained in or concepts devied from a collection." An index gives the thought content of both documentary and non-documentary sources.

Harrod's Glossary of Library Terms defines index as "A systematic guide to items contained in or concepts derived from collection. There items or derived concepts are represented by entires arranged in know or stated, searchable order, such as alphabetical, chronological or numerical."

Index also refers to an open end finding guide to the literature of an academic field or discipline (example : Philosopher's Index) to works of a specific literary form (Biography Index) or Published in a specific format (Reader's Guide to Periodical Literature), or to the analyzed contents of a serial publication (New York Times Index).

Indexing - definition : Indexing is the art of compilling an index. It provides series of access points, which lead from information known to the user to additional, previously unknown information.

Thus indexing is the process of compilling one or more indexes for a single publication, such as a monograph or multivolume reference work, or adding entries for new documents to an open end index covering a particular publication format (example : newspapers), works of a specific literary form (biography, book reviews, etc.), or the literature of an academic field, discipline, or group of disciplines.

The professionally trained indexer reads or scans the text of each document to determine its content, then select appropriate headings (name, places, subjects) to facilitate retrieval. Cross-references are made from synonyms, and the entries are arranged in the desired sequence (alphabetical, numerical classified, etc.). In an open end index, content dexeriptors are usually selected from a list of preferred terms (controlled vocabulary), developed over by the indexing service.

Indexing Language : An artificial languages consisting of subject headings or content descriptors selected to facilitate information retrieval by serving as access points in

a datalog or index, including any lead-in vocabularly and rules governing from of entry, syntax etc. is called Indexing Language.

2.1.3 Purpose of Indexing :

The basic purpose of indexing is to math the ideas expressed in a document with the ideas embedded in a chosen heading/s from a prescribed list of subject headings. In other words its purpose is to help maximum of recall or retrieval of relevant information with minimum of noice (unrelated information). Index thus serves as an operating tool. No store can be efficiently operated without an index of some kind or the other. The indexing tool becomes necessary because any store of information of documents (library and information centre) containing information is required to be organized for repetitive use by the user.
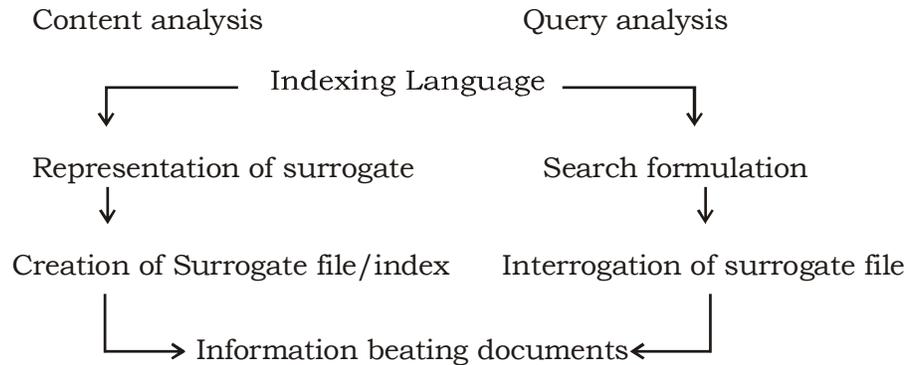
Uses : The indexed play a vital role in literature search and informing about availability and location of relevant documents in response to users request. The primary uses of indexes are :

(i) Overview of the subject : A general view of a subject can be obtained from a subject Index. It provides information on contemporary or current issues. It also reveals trends in the growth of a subject.

(ii) Access to information : It facilitates subject, author and title approach to seek information. It facilitates search through specific subject headings.

(iii) Resolve language problem : A major reason for indexes is to aid in solving the problem presented by the many languages in which material is now published.

(iv) Resolves problem of information explosion : The solution to the problem of information explosion (enormous number of documents published annually) can be provided by indexes, which facilitate rapid selection of relevant material, e.g, in the field fo science and engineering, it si now impracticable to search the primary literature owing to its size, the only practicable way through indexes.

2.1.4 Need for Indexing

To retrieve desired Information from any large collection of documents, we require two things of that collection. First the collection must be ordered so that when we know what we want, we will know where to find it. Second we must have a means of searching and matching; there should be some way of recognizing whether or not any given document contains the information we want. Reading the complete document is one approach to the later requirement, but it involves time factor and is still does not guarantee the searcher the required information. To resolve this, we use indexing as a tool for locating information. The need for indexing can be explained as follows :

1. Parallelism of indexing and retrieval work.

Content analysis   Query analysis

Indexing Language

Representation of surrogate   Search formulation

Creation of Surrogate file/index   Interrogation of surrogate file

Information beating documents

2.   To help maximum recall or retrieval of relevant documents and to achieve high precision ratio.

3.   Operating tool - to operate the documentary store.

4.   Necessary for the repretitive use of the store by user.

Mc Colvin (1958) stated the need for indexing as follows :

(i)   to facilitate reference to specific item

(ii)   to disclose relationships

(iii)   to disclose commissions

(iv)   to answer queries of discovery

(v)   to provide comprehensive overview of subject field

(vi)   to give nomenclature guidance

Self Check Exercise

Define index and explain its purpose

2.1.5 Indexing : development and trends

The indexing has a long history. It owes its origin to the traditional books index which has gradually seen considerable improvements and developed as bibliographic index/subject index. Specific aspects of a subjects are given in the form of subject headings following standard subject heading lists such as LC Subject Headings, Sears

List of Subject Headings, Chain indexing, PRECIS, POPSI, Citration indexing, KWIC, KWOC, indexing etc. These techniques have been introduced by emitent library and information scientists like Culture, Kaiser, Coates, Ranganathan, Luhn, Garfield, Austin, Bhattacharya and others.

Indexing is an old as writing. In 2nd Millennium B. C., Mesopotamian cueniform documents were enclosed in clay envelopes, which were as an Index of sorts; so that the documents couldn't be tampered with, they were sealed inside the envelope and then the contents were copied on envelopes.

During 3rd c B.C., abstracts of business records we maintained.

During 240 B.C., at Alexandrial library (Egypt), Callimachus of Cyrene, perhaps Alexandria's most famous librarian, created for the first time a subject catalog in 120,000 scrolls of the library's holding, called the *Pinakes* or Tables. It was more like a good subject index. Around 900 AD, There was at least one Encylopedia that was arranged in alphabetical order (but most such works had a classifed order).

In 12th C, ther was rise of universities and people started taking interest in various subjects as such. This, created interest in scholarship and subseruenct demand for alphabetical indexing.

First alphabetical listing of Vatican archives appeared in 13th C. True Alphabetical indexing, seems to have emerged in the 14th C. They consisted primary of the keywords in the theses or disputations, alphabetically arranged. However, this arrangement was not fully alphabetical in the modern sense. The ordering was extended mostly through the first syllable or first three letters of the first world. Mean while the first example for Catalogue with a index turned up in the 15th C when in 1483, a catalogue of the monastery at Melk, Austria had three Indexes-Author Subject categories and Anonymous works listed by the catch words in there titles (Vener).. Towards the end of 15th C the practice of supplementing the systematic listing with an alphabetical subject index was introduced in order to increases the mainipulative capacity of catalogue.

In 1545 Conrad Gesture's Bibliotheca Universalize listed the documents under the alphabetical order of the author's forename. Later in 1548, listed the same document in a Subject Classification order with an Alphabetic Subject Index to Classification codes. This can be considered as the genesis of all the present indexing systems.

18th C has witnessed increased presence of subject indexes to books and journals. In 1856, Andrea Crestandoro made an attempt to show the importance of titles of documents to derive indexes. Late 19th C witnessed more systematic and widespread use of subject indexing. Periodical indexes also appeared; For e.g., William Frederick Poole's 'Poole's index'.

In the field of subject indexing, the dictionary catalogue of C.A. Cutter brought easier way of organizing the documents resolving the conflict between subject and author

index through *Rule for dictionary cataloguing* in 1876.

In 1911, J. Kaiser in his *Systematic Indexing* laid down a logical solution to the problem of composite subjects that can be analysed in to combination of concepts indicating a concrete object and a process.

A need was felt for designing an appraratus, which will lead to some standard terms from several variant forms used by the authors for common concepts. Under the aegis of ALA Committee, Subject Headings were developed for use in dictionary catalogue. Lists of subject headings such as Library of Congress and Sears List of Subject headings, Medical Subject Headings (MeSH) were published that helped in compilations of good number of lists for special collections for special purposes. E. J. Coate's contribution towards subject indexing is very significant. His order of significance among various components of a compound subject is *Thing, Material and Action.*

However, S.R. Ranganathan's analytical mind contributed to approach the issue of subject headings more scientifically. He developed Colon Classification in 1930 based on fundamental categories (PMEST); translating the notation digit by digit back in to terms provides the names of the subject. This is popularized as Chain indexing. His methodology was developed at DRTC and his CC went into 7th Edition.

J.E.L.Ferradane has provided a new approach creating a new type of syntax in indexing language. His subject formulation is based on the relationship that exists between each pair of components. He has identified nine such relationships that are indicated by operators or symbols. Sharp is of the view that specific subject formulations, having all the components may not be of much use to the searchers of information and in his SLIC indexing system ensures optimum economy in indexing.

Timothy C Craven has developed the Nested Pharase Indexing System (NEPHIS) in which an input string of indexing terms is constructed as in PRECIS and POPSI.

The later part of 20th C was marked by information explosion and emergence of computers leading to automated indexing process. More over access to literature even by non-scholars necessitated the development of indexes in every field of universe of knowledge. As a result lot of innovative techniques in indexing procedure have emerged predominantly in U.S.A. From 1970s the rise of Selective Dissemination of Information (SDI) services, titles of scientific documents began to play a significant role in Science Communication. The title based Indexes, developed by H.P. Luhn, depend upon manipulation of all keywords in the title to give multiple entries, one entry for each significant word, known as key word indexing. e.g. KWIC, KWAC etc. At present the application of information technologies lead to automatic indexing and information retrieval processes.

Self Check Exercise

Trace the major trends of indexing in 20th century

2.1.6   Process of Indexing :

The essential operational functions of indexing include the following steps :

*   ***Scanning the collection :*** Indexer reviews the table of contents and introductory materials, skimming the entire publication to get sense of the topics and their interrelationships, noting chapter headings and subheadings. By this the indexer will know what the document is about.

*   ***Analysing the contents based on predetermined criteria :*** The Indexer tries to include in the index every significant terms of information of the document, choosing entries and subentries with an understanding of the way people will look up those terms in the indexes.

*   ***Choosing and tagging discrete items-descriptor–with appropriate identifies :*** Choosing the descriptors or index that represent the subject concept and tagging them with relevant identifiers like author etc.

*   ***Modeling the term to provide access points :*** Changing the term in natural language into controlled vocabulary using standard subject heading lists or thesauri.

*   ***Adding to each inetifier the precise location*** within the collection for the purpose of retrieval like the call number.

*   ***Providing information reliability and accessibility***

2.1.7   Functions of Indexing :

The primary functions of indexing include (i) synthetic and (ii) Syndetic i.e. connectivity of terms. These functions are being carried out at two stages subordinate and coordinate stages. The function of an index is to provide users with an efficient and systematic means for locating documents or parts of documents that may address information needs or requests. An index is therefore performing the following functions :

(i)   Indentify and locate potentially relevant information in the document or collection being indexed.

(ii)   Discriminate between information on a topic and passing mention of a topic.

(iii)   Exclude passing mention of topics that offer nothing significant to the potential user.

(iv)   Analyzes concepts treated in a document so as to produce suitable index headings based on its terminology.

(v)   Indicate relationships among topics.

(vi)   Group together information on topics scattered by the arrangement of the document of collection and organizes headings and subheadings into index entries.

(vii)   Directing users seeking information under terms not chosen as index headings to headings that have chosen, by means of see references.

(viii) Suggest users of a topic to lookup related topics also by means of see also reference. Ex: Cow see also Mammals.

(ix)   Arrange entries into a systematic and helpful order.

Indexes provide guide to material that the user may wish to recall or that he may not know exists; the is indexes are used for questions of recall or discovery. Without indexes the searcher would waste time by turning through documents page by page or by examining all items in a museum collection. Indexes save time and make practical searches that would other wise be given up.

2.1.8 Arrangement :

Usually an index is arranged alphabetically but may some times be arranged chronologically, geographically, numerically, or in any other suitable manner, depending upon the requirement. Indexing is a technique by which ideas expressed in a document are matched with the ideas inherent in the chosen heading/s from a dictionary of prescribed headings.

An **entry** usually consists of two elements : (i) descriptor and (ii) the location where the item or concept has been available. Depending on the nature of entry the indexes can be categorized as

Author Index : An alphabetically arranged index in which the headings are the names of the individuals and corporate bodies responsible for creating the works indexed. Author entries may be combined with the subject index or title index, rather than list separately.

Classified Index : An index in which entries are arranged under headings and subheadings indicating hierarchical divisions within classes based on the subject matter indexed, rather than in alphabetical or numerical sequence. To use such an index effectively, a subject index/classification scheme is required.

Subject Index : An alphabetically arranged list of headings selected by an indexer to represent the subject content of one or more works, with locators (usually page numbers) to direct the user to the corresponding text. Names are usually included in

the subject index, but some publications has a separate name index and even a separate geographic index of place names. In some publications, the subject index is combined with the author index in a single slphabetic sequence.

Title Index : An alphabetically arranged list of the titles of the works covered in a serial or nonserial publication, for example, *the Book and Film Title Index to America : History and Life.* Title and author indexes are sometimes combined.

Self Check Exercise :

        List three important functions of indexes

2.1.9 Summary :

Index is an alphabetized listing of names, places and subjects included in a work or a document store like library that give address to each item with the help of which the item can be identical and located. In library context, it is a systematic guide to times contained in or concepts derived from a collection and facilitate subject access to documents. It is the process by which information is made accessible. The basic purpose of indexing is to help in retrieval of relevant documents. The primary function of index is to act as an operational tool to organize the information and help the user in the retrieval of the same. With the growth in information and need for their organization the importance of indexing is growing steadily. They have improved over the years. Contemporary indexes are more accurate and there are more indexes of all kinds. Indexes are now stored in computer files and searched by correlation of subject headings. Computers also aid in the preparation of automatic indexing and retrieval.

2.1.10 Glossary

Controlled Vocabularly :

An established list of preferred terms from which a cataloger or indexer must select when assigning subject headings or descriptors in a bibliographic record.

Descriptor : A terms or a set of term used to surrogate the concept

Index : A systematic guide to items contained in or concepts derived from a collection. These items or derived concepts are represented by entries arranged in known or stated searchable order.

Indexing : The process of compilling one or more indexes.

Indexer : The person who compiles index.

Keyword searching is the searching of a database using a signification word from the title, abstract, or descriptor of a record as a point of reference to the article's overall content.

Subject headings : A word or phrase from a controlled vocabulary, which is used to describe the subject of a document. The most commonly used subject headings in libraries are the Library of Congress Subject Headings (LCSH), Sears List of Subject Headings.

Theasurus : Guide to use of terms, showing relationships between them, for the purpose of providing standardized, controlled vocabulary for information storage and retrieval.

2.1.11 References and further reading

1.   Prasher. R.G. (1989) Index and indexing systems. New Delhi, Medillion.
2.   Rajan, T.N. Ed. (1981) Indexing Systems. Calcutta, IASLIC.
3.   Riaz, Mohammad. (1991) Advanced indexing and Abstracting Practices New Delhi, D.K. Publishers.
4.   Khanna, J.K (1985) Advances in Librarianship. New Delhi, ESSESS Publications
5.   Borko, Harold and Bernier, Charles L. (1978) Indexing concepts and Methods. New York, Academic Press.

LESSON NO. 2.2   **AUTHOR : R.S.R. VARALAKSHMI**

TYPES OF INDEXING

Structure :

2.2.1 Objectives

After reading this lesson you will be able to understand

    (i)   basic principles of indexing;

    (ii)   different types of indexes;

      (iii)   importance of each type; and,

      (iv)   the modern methods of indexing.

2.2.2 Principles of subject indexing

The basic principle of indexing is encoding the subject concepts of a document in to retrievable index terms. The encoding to index terms includes two important functions - *synthetic* i.e. analysis of subject concepts and *syndetic* i.e. establishing relationship or connectivity. The relationship can be either subordinate as used in conventional classification schemes as method and representing each index term as an autonomous access point. The encoding can be done by assigning appropriate words that are devised from an indexing language like L.C. Subject Headings or deriving the subject terms from the title, preface, abstract as given by the author. The derived indexing is associated with inherent problems such as :

*   Dependence on author hence uniformity can't be achieved

*   It is difficult to connect the related terms, E.g. in computer search on teaching-teacher, teaching etc will appear but not the taught.

Hence assigned indexing has been given importance for indexing of documents of facilitate easy information retrieval using subject terms or index terms or key words. In any kind of indexing, etablishing relationship between two similar subject concepts depend on the judgment capability of indexing language. To achieve this, an indexing language must contain the equivalents of both a dictionary to show semantic relationships and a grammar to cater for syntactic relationships.

2.2.3 Syntax and Semantics

Syntax : That part of the grammar that deals with the way in which words are put together to form phrases, clauses or sentences is called as syntax. Also refers to the grammar and sentence structure of a language and to the branch of linguistics devoted to its study. It indicates the order in which the search terms and Boolean operators used in a keyword search statement are typed. Syntax determines the sequence in which a computer-based information retrieval system executes the search. In most bibliographic databases, commands are executed from left to right unless parentheses are used to indicate otherwise (a technique called nesting). Syntax is necessary to be able to search for the intersection of two or more classes defined by terms denoting distinct concepts. In a composite subject there could be many ways to express subject concepts.

Example :

      Children and television and (violence or aggression)

In this search statement, the Boolean "or" will be performed before the "and".

Thus the syntax provides the rules that determine how headings are constructed and how semantic relations among terms are indicated in an indexing language, for example, the conditions under which a parenthetical qualifier is added to a heading,

or an inverted heading is used instead of conventional word order.

Semantics : The branch of linguistics concerned with the meaning of the words, signs, and symbols that constitute the elements of change and evolution in a spoken or written language. It also deals with relationships of meaning between signs and their referents, within a system of communication.

Semantic relation : The connection in meaning between two or more concepts and between the terms (subject headings or descriptors) used to represent them in an indexing language. Semantic relations can be classified as follows :

Relation	Description	Example
Active	Action, process, or operation directly performed by one on the other	Scanner/Barcode
Associative	Linked conceptually but not hierarchically	Library statistics/ Bibliometrics
Causal	One responsible for occurrence of the other	Acquisitions/Collection growth Library/Academic library
Hierarchic	One a logical subclass of the other	Bookbinding/Binding
Locative	One located at, in, or on a place specified by the other	Mainz Psalter
Partitive	Part to whole	Chapter/Book
Passive	One influenced by or subjected to the action of the other with no reciprocal influence	Library collection/ Selection criteria
Antonymous	Opposite in meaning	Selection/Deselection
Synonymous	Having the same or nearly the same meaning	Booklet/Pamphlet

Source : URL : http://lu.com/odlis/odlis_s.cfm#semanticrelation

In indexing the major semantic relationships are considered in three groups - equivalent, hierarchical and associative or affinitive.

Equivalence relationships implies that there is more than one term denoting the same concept. E.g. synonyms, Acronyms, Abbreviations, Preferred spelling, current and established terms etc. In indexing language it is used to introduce a measure of control over this situation by choosing one of the terms as preferred and using only that in index by using a series of linkages called cross references e.g. 'see' 'use' 'x' etc.

Hierarchical relation represents class inclusion from generic to specific; whole part. It shows the kind of relationship.

E.g. Cattle

     NT Beef cattle

Cattle

     RT Hides

Beef Cattle

     BT Cattle

Hides

     RT Cattle

Affinitive relationships are not obvious and not clearly defined. Coordination may be a by-product of the generic relationship; it can also arise from whole-part relation. It merely shows relationship but not the kind.

E.g. Solar system

     See also   Sun.

Therefore semantic relations show alternative and substitute terms to searcher and indexer as well.

2.2.4 TYPES OF INDEXING : Over the period of time a number of indexing techniques have been developed to help the information retrieval achieve more recall and precision that is capability to retrieve more relevant and pertinent documents from the document store with the help of indexing. Some important types are briefly described below each of which have its own distinct characteristics, merits and demerits.

2.2.4.1.   Assigned indexing and derived indexing : Any indexing method has to indentify the subject bearing concepts to represent the document that will later facilitate recall or retrieval. The index terms can be either assigned or derived as explained under 2. Principles of indexing.

Assigned indexing : A method of indexing in which a human indexer selects one or more subject headings or descriptors from a list of controlled vocabulary to represent the subject(s) of a work. The indexing terms selected to represent the content need not appear in the title or text of the document indexed. Assigned terms are terms not occurring in the text. For example the manual library systems, in which books were classified according to an existing classification system, for example, Dewey or UDC, are assigned indexing systems; Assigned indexing otherwise known as concept indexing deals with five categories of concepts viz. entitles; activities; abstracts; properties; heterogeneous. The characteristics of assigned indexing are properties; heterogeneous. The characteristics of assigned indexing are

    (i)   Choice of terms i.e. five categories mentioned above

    (ii)   Choice of form of words i.e. noun form etc.

    (iii)   Establishing relationship either permanent or adhoc.

Permanent shows the relationship of related terms. E.g. Sea, River, Water. Adhoc or

associated shows the order of presenting the terms. E.g. poly ethylene bags, welding plastics i.e. unrelated terms form a composite subject. The permanent typeof relation is known as semantic while the adhoc is called as syntax.

Derived Indexing : A method of indexing in which a human indexer or computer extracts from the title and/or text of a document one or more words or phrases to represent subject(s) of the work, for use as headings under which entries are made. It is synonymous with *extractive indexing*. Derived indexing terms are terms occurring in the text to be indexed. Computerized IR-systems which extract keywords from the documents according to a weighing scheme are typically derived indexing systems. The derived indexing systems became very popular when the computer made it easy to create an inverted list of all the words occurring in a document base. E.g. Keyword indexing.

2.2.4.2.   Alphabetical indexing : An index may also draw its name from the arrangement of its contents. If the references in it are arranged alphabetically, it is known as an alphabetical index. It can be separate author, title and subject indexes in alphabetical order or can be a single alphabetical list of all the three.

Author Index : An alphabetically arranged index in which the headings are the names of the individuals and corporate bodies responsi ble for creating the works indexed. Author entries may be combined with the subject index or title index, rather than listed separately.

Title Index : An alphabetically arrnaged list of the titles of the works covered in a serial or nonserial publication, for example, the Book and *Film Title Index to America: History and Life.* Title and author indexes are sometimes combined.

Subject Index : An alphabetically arrnaged list of headings selected by an indexer to represent the subject content of one or more works, with locators (usually page numbers) to direct the user to the corresponding text. Names are usually included in the subject index, but some publications have a separate name index and even a separate geographic index of place names. In some publications, the subject index is combined with the author index in a single alphabetic sequence. In contrast to an alphabetical index, the index entries may be arranged systematically by the subject represented by class number called as classified index.

Classified index : An index in which entries are arranged under headings and subheadings indicating hierarchical divisions and subdivisions within classes based on the subject matter indexed, rather than in alphabetical or numerical sequence. To use such an index effectively, a parallel alphabetical subject index is required.

2.2.4.3.   Pre-and Post coordinate Index : Indexing involved three steps :

(i)   Conceptual analysis i.e. a composite subject is analyzed into its constituent concepts

(ii)   Translation involving conversion of the conceptual analysis in to a set of index

terms

(iii)　To put or synthesize the concepts in an order recommended by the rules of the language i.e. coordination of index terms.

Coordination

Pre coordination
Bringing in coordination of terms in anticipation of readers approach

Post coordination
The component conceptions are un-coordinated by the indexer. Coordination could be done at the time of search
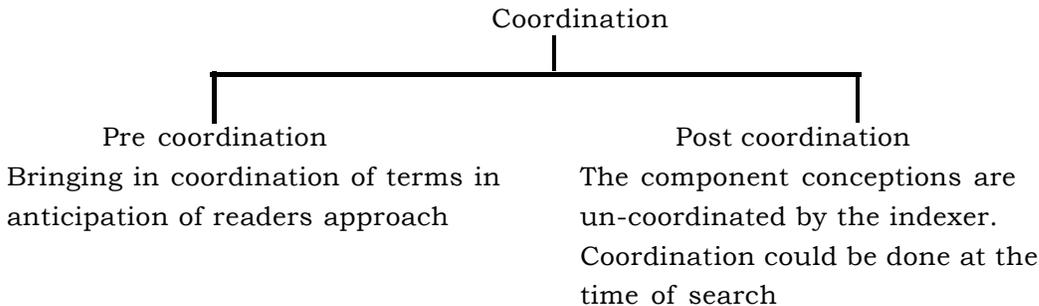
Pre coordinate indexing : A method of indexing in which multiple concepts are combined by the indexer to form subject headings or descriptors assigned to documents to facilitate the retrieval of information on complex subjects

Example : "Libraries and the blind–United States–Directories" instead of Libraries + Blind + United States + Directories.

The characteristics of pre coordination are :

*　　　Need for determination of order of significance. A great part of syntactical rules devoted to it.

*　　　It offers multiple entries and network of references

*　　　Based on System of rotation (POPSI); Context Preserved (PRECIS); Class number based (Chain indexing) etc.

PRECIS : PRECIS was developed by Derek Austin. It is an indexing system in which the initial string of terms, organized according to a schemeof role operators, are computer manipulated so that selected words function in turn as the approach term. The primary objective is to be independent of a classification system and the subject headings generated are coextensive with the contents of the document at all access points. It has an ordering procedure to organize termsof composite subjects.

PRECIS is an acronym for PREserved Context Indexing System. It was developed as an alternative procedure for deriving subject headings and generating index entries for British National Bibliography (BNB). The BNB has been following Chain Indexing since its inception for preparing its alphabetical index. There were two important factors that contributed to find an alternative method. These are :

(i)　　　Idea of replacing chain Procedure of indexing of BNB; and

(ii)　　The decision of the British Library to generate computer produced BNB with all the indexes.

Definition

Derek Austin very succintly defined PRECIS as a system in which "the initial string

of terms, organized according to a scheme of role operators, is computer-manipulated so that selected words function in turn as the approach term. Entries are restructured at every step in such a way that the user can determine from the layout of the entry which terms set the appropriate into its context and which terms are context dependent on the approach term."

Thus PRECIS allow a user to enter an alphabetical subject index at any one of the significant terms which together make up a compound statement and establish at that point the full context in which his chosen term has been considered by the author.

Objectives :

The main objectives of formulating PRECIS are :

1. Entries produced by this method should be co-extensive with the subject of the document.
2. Each of the entries should be meaningful and allow the users to interpret the entry correctly.
3. The use of computer should generate all index entries. The indexer's role would be only to prepare the input strings and give instructions to computer.
4. There should be adequate system of references for semantically related terms.
5. The system should be able to admit freely new terms into the system as soon as they are found in literature.
6. There should be common set of indexing rules in order to have consistency of work.

Levels of Operations

PRECIS has two levels of operations :

(i)   **Human Level :** Here a subject statement is analysed into a set of roles.

(ii)   **Computer Level :** Here the analysed subject statement is programmed to be manipulated into producing a variety of PRECIS subject index entries by computer-processing.

Syntax and Semantics of PRECIS

There ar two inter-related sets of working procedure : Syntactical and Semantic.

(1)   Syntactical relationships are governed by means of a set of logical rules and schema of role operators and codes.

(2)   Semantic relationships between indexing terms and their synonyms are controlled by machine held theasurus. A thesaurus is generated simultaneously with the preparation of input string.

Principles of PRECIS

It is based mainly on two principles :

(i)   **Principles of Context Dependency :** When component terms are arranged in a string, they are context dependent on each other. The meaning of each

string term depends upon the meaning of its preceding term, and express the single context together.

(ii) **Principle of One-to-One Relationship :** When terms are organized in context-dependent order, they form a one-to-one relation. It means that each term in string is directly related to its next term.

Entry Structure

The entry structure for the PRECIS entry is two-line-three part setup. A string of five tems A-B-C-D-E is represented in one-to-one related sequence. Term A or Term E can act as entry point. But, when C is considered as entry point, the different structure is necessary. This is fully expressed in the figure.

Lead → | C | | B   A | ← Qualifier

| D   E | ← Display

Figure : Entry Structure

Here, the first line consists of two units - Lead and Qualifier. This is called Heading. The second line is called Display. Lead consists of approach term or user's access point. Qualifier is that term(s) that set the 'Lead' to bring it into its proper context. Display includes those terms that depend upon the Heading for their context.

Input String, Role Operators and Codes

Input String is a set of terms arranged according to the role operators which act as instructions to the computers.

Role Operators refer to a set of notations. It specifies the grammatical role or function of the term which follows the operators and which regulates the order of the terms in an input string. The rules associated with role operators serve as computer instructions for generating index entries determines : format, typography, and punctuation.

Kinds of Role Operators

1. **Primary Operators :** These were earlier called Mainline Operators. These control the sequence of terms in the input string. They also determine the format of entries in the printed index. There are seven Primary Operators, 0 to 6, having built-in filing value for three sets of concepts.

2. **Secondary Operators :** These were earlier called Interposed Operators. These can be introduced into a string at any point to raise its level of exhaustivity. But these Operators cannot be used to start a string.

Schema of Role Operators

Primary Operators

Environment of Core Concepts	* Location
Core Concepts	1. Key System
	Thing when action not present
	Thing towards which an action is directed, e.g., object of transitive action, performer of intransitive action.
	2. Action; Effect of action
	3. Performer of transitive action (agent, instrument); Intake; Factor
Extra Core Concepts	4. View point-as-form
	5. Selected Instance : Study region, sample population
	6. Form of document : Target user

Secondary Operators

Co-ordinate concepts	f	'Bound' coordinate concept
	g	Standard coordinate concept
Dependent elements	p	Part; Property
	q	Member of quasi-generic group
	r	Assembly
Special Class of Action	s	Role definer; Directional property
	t	Author-attributed action
	u	Two-way attraction

Codes in PRECIS Strings

Primary Codes

Theme Interlinks	$x Ist concept in coordinate theme
	$y 2nd/subsequent concept in theme
	$z Common concept
Term Codes	$a Common noun
	$c Proper name (class of-one)
	$d Place name

Secondary Codes

Differences

Preceding differences	Ist and 2nd character
(3 characters)	$0 Non-lead, space generating
	$1 Non-lead, close-up
	$2 Lead, space generating

$3 Lead, close-up

3rd character = number in the range, 1 to 9 indicating level of difference

Date as a difference	$d
Parenthetical difference	$n Non-lead parenthetical difference
	$o Lead parenthetical difference
Connectives	$v Downward reading connective
	$w Upward reading connective

Topographic Codes

$e Non-filing part in italic preceded by comma

$f Filing part in italic preceded by comma

$g Filing part in roman, no preceding punctuation

$h Filing part in italic preceded by full point

$i Filing part in italic, no preceding punctuation

Example

Recruitment of Personnel in Textile Industries of India

As discussed under 'Entry Structure' the 'Lead' is printed in bold face to emphasize significance. 'Qualifier' consists of terms of wider context, and 'Display' that of narrower context. The 'Qualifier' and 'Display' positions are not necessarily occupied in all index entries. So the standard entries for Input String, therefore, will be as follows:

INDIA

Textile industries. Personnel. Recruitment

TEXTILE INDUSTRIES, INDIA

Personnel, Recruitment

PERSONNEL, TEXTILE INDUSTRIES, INDIA

Recruitment

RECRUITMENT Personnel. Textile Industries. India.

Steps in PRECIS

1. Analysing the thought content of the document, and indentifying the component terms denoting key concepts.
2. Organizing the component terms into a subject statement based on the principle of context dependency.
3. Determining the role of each term in terms of role operators.
4. Assigning the role operators which signify the syntactical role of each term.
5. Deciding which terms should be access points and which would be in other positions, and assigning further codes.
6. Adding further prepositions, auxiliaries or phrases for clarity and expressiveness in the index entries.
7. Generating index entries by the computer.
8. Generating supporting reference entries (i.e., sec and see also references)

from semantically related terms from a machine-held thesaurus.

Formats of PRECIS Index

There are three kinds of formats : Standard Format, Inverted Format, and Predicate Transformation.

1. **Standard Format :** Index entries are generated with primary operators (0), (1), and (2) through 'shunting' process.

   Example :   Training of Labours in India

           (0)   India

           (1)   Labour

           (2)   Training

           INDIA

           Labours, Training

           LABOURS, India

           Training

           TRAINING, Labours, India

2. **Inverted Format :** Entries here are generated when a term coded by an operator in the range of (4) to (6) or its dependent elements appear in the lead. The whole string is read from top to bottom and is written in the 'Display'.

   Example :   A Bibliography of Statistics for Librarians

         (1)   Statistics

         (5)   Librarians $01 for

         (6)   Bibliographics

         STATISTICS

             For librarians – Bibliographics

         LIBRARIANS

             Statistics – For Librarians – Bibliographics

         BIBLIOGRAPHIES

             Statistics – For Librarians

3. **Predicate Transformation :** When an entry is generated under a term coded (3) that immediately follows a term coded by (2) or (6) or (t) indicating action of one kind or another.

   Example :   In-service Training of Librarians by Teachers

         (1)   Librarians

         (2)   Training $21 In-service $v by $w of

         (3)   Teachers

         LIBRARIANS

             In-service Training by Teachers

         TRAINING LIBRARIANS

             In-service Training by Teachers

TEACHERS

In-service Training by Teachers

Use of PRECIS

In view of the successful use of PRECIS in UK, other countries such as USA, Canada, Australia, Germany, Denmark, Italy, Pland, China, etc. also started using it. One of the reasons for its outstanding success is the support given by the BNB where a large number of documents are indexed. It is being used in English as well as non-English indexing systems. It is here that PRECIS Manual contributed considerably to its application in many situations.

Limitations :

1. It is considered as complex system to learn and use.
2. It is regarded inefficient if used manually, as the generation of the appropriate terms is very time consuming.
3. The application of 'Role Indicators' for the formation of string may not produce the same results when different indexes interpret the operators with reference to the different concepts.

But despite all these limitations, PRECIS claims to be one of the best systems of indexing currently available.

POPSI : The system was developed by G. Bhattacharya of DRTC, India. It is based on Ranganathan's principle and postulates of theory of classification unlike class number in chain index. The POPSI string is formulated on the basis of verbal classification without using the notation part of it. The derived subject headings are useful for indexing and for feature headings also.

New research in verbal indexing had been going on at the Documentation Research and Training Centre (DRTC), Bangalore. These efforts resulted in the formation of the Postulate-Based Permuted Subject Indexing (POPSI). It is based on the general theory of classification involving postulates and principles. Further, it is now believed that POPSI is an improved version of Chain Procedure of indexing which was based on Class Number.

Purposes

It can be used for various purposes, according to Neelameghan and Gopinath, such as :

1. To formulate subject headings ;
2. To derive subject index entries;
3. To determine the subject of reader's query;
4. To formulate a strategy for searching information; and
5. To derive a base for the presentation of ideas in the text of a document.

Steps

The main steps, according to Neelameghan and Gopinath, involved in this indexing

system are as given below :

Step 1- Verbal Representation : The first step is the derivation of the postulate-based verbal subject representation. It is based on postulates and principles of the General Theory of Library Classification formulated by Ranganathan. It may be used to :

(i)   secure consistency in the analysis of subjects;

(ii)   determine the upper links for each component idea of the subject;

(iii)   indentify inter-relationship between component ideas of the subject;

(iv)   link the component ideas in a helpful sequence; and

(v)   ensure the co-extensive expression of the subject.

Step-2- Display of Components : It involves the determination of the pattern of display of the components of the verbal representation.

Step -3- Short Display : This step is to determine those component terms which are necessary to be retained in the verbal representation to be used for specific purposes.

Step-4-Approach Terms : This step is to find out all those approach terms which are likely to be used as "Approach Terms" by the readers in their search for information on the subject represented.

Step-5-Deriving Subject Index Entries : It helps to derive subject index entries in such a way that each of the approach terms occurs as a lead term of the entries.

Step-6-Display of Subject Index Entry : Choice of method/pattern/format for display of each subject index entry derived by the procedure selected at Step-5.

Step-7-Cross Reference : Determination of the pattern for subject Cross Referencing in the alphabetical subject index.

Step-8- Alphabetical Arrangement of Entries : This step is concerned with the arrangement of the Subject Index Entries and Subject Cross Reference Index Entries in alphabetical sequence in a single file.

Thus, the process in POPSI, is completed in several steps. These have been recognized by G. Bhattacharyya, as :

1. Analysis   5. Preparation of Entries for organizing Classification

2. Formalization   6. Decision about Terms of Approach

3. Standardization   7. Preparation of Entries for Associative Classification

4. Modulation   8. Alphabetization

Actual Work in POPSI

Guha stated that the actual work involved in each of the above steps can be demonstrated with the help of example of a document entitled "Oral Streptomycin Treatment of Human Lung Treatment : An Evaluation of Undesirable Effects".

Step-1- Verbal Representation (Derivation of Postulate-based verbal subject) : Once the component ideas in the subject are determined by,

(a)   linking up these ideas into sub-structures,

(b)   putting the components in a helpful sequence within each sub-structure,

(c)   arranging the sub-structure in a helpful sequence among themselves,

(d)   indicating the sequence and relationships, such as the indicator digits of colon classification.

The subject now can be represented as :

Medicine, Human body > Respiratory System > Lung : [Disease-Bacteria>Mycobacterium > Mycobacterium tuberculosis] = Tuberculosis : Treatment - Drug > Antibiotics : Streptomycin - Route of Administration : Oral; Effects > Side effects > Undersirable effects : Evaluation

In this example

(i) symbol > denotes inclusion relation

(ii) superordinate (upper) links of each specific idea are shown

(iii) all components are arranged in a helpful sequence

(iv) words and symbols are used to show the sequence and relationship among the components in the subject representation.

Step-2-Display of Components : Here, the display in a bibiligraphical tool, such as, a ctalogue or a documentation list is considered. Hence the display has to be linear.

Step-3-Short Display : The full representation prepared in Step-1 can be reduced to the following name-of-subject :

Medicine, Lung : Tuberculosis : Treatment -

Streptomycin-Oral administration; Undesirable effects :

Evaluation

Step-4-Approach Terms : The selection of approach terms is made from the representation of the subject as mentioned in Step-1. Synonyms and the alternative terms are also taken care of. In the above given exmaple, "Therapeutics' and 'Therapy' may be considered alternative terms for "Treatment'. Similarly, 'side reaction' is a synonymously used term for 'side effect'.

Step-5- Subject Index Entries : The subject index entries are to be derived from the full subject representation in such a way that each of the approach terms selected occurs as Lead Term of the entries.

Step-6-Display of Subject Index Entries : Each subject index preferably consists of three sections :

- Lead Term Section
- Subject Representation Section
- Reference Section

The Lead Term Section gives Lead Term and its immediate context in which the

idea denoted by the Lead Term occurs in the subjec concerned.

In the Subject Representation Section, the short display of the subject as stated in Step-3 is given.

The Reference Section gives the Class Number or any other identification number of the document.

Some of the subject index entries will be rendered as follows :

EVALUATION, UNDERSIRABLE EFFECTS

> Medicine, Lung : Tuberculosis : Treatment -
> Streptomycin-Oral; Undersiable effects : Evaluation

LUNG, MEDICINE

> Medicine, Lung; Tuberculosis : Treatment -
> Streptomycin-Oral; Undesirable effects : Evaluation

DRUG, TREATMENT, TUBERCULOSIS

> Medicine, Lung; Tuberculosis : Treatment -
> Streptomycin-Oral; Undesirable effects : Evaluation

Step-7-Cross Reference : Cross Reference Index Entries (see references) are to be prepared for each of the alternative, near-synonymous and synonymous terms for the component terms in the subject representation. Again, Cross Reference Entries (see also references) are to be prepared for each of the closely related or collateral idea or subject. For exmaple :

> Therapeutics see Treatment
>
> Therapy see Treatment
>
> Streptopenicillin see also Streptomycin
>
> Streptomycin see also Streptopencillin

Step-8-Alphabetical Arrangement : The Subject Index Entries, Subject Cross Reference Entries, and Subject Cross Index Entries, prepared according to the above procedure, are to be merged and arranged in a single alphabetical sequence.

Recent Meodification of POPSI

Bhattacharyya has carried forward POPSI, and made some modifications which enabled this new version to be referred to as Bhattacharyya's POPSI. He postulated a generalised subject indexing language with a set of elementary categories : Discipline (D), Entity (E), Action (A), Property (P), and a set of modifiers (M). The details are as given below :

1.   **D = Discipline :** It includes conventional field of study, or any of their aggregates, such as, Social Science, Physical Science, etc.

2.   **E = Entity :** This category includes manifestation having perceptual correlates, or only conceptual-existence, as contrasted with their properties, and actions performed by them or on them; e.g., Energy, Light, Plants, Animals, Place,

Time, Environment, etc.

3. **A = Action :** An elementary category that includes manifestations denoting th concept of 'doing action'. Function, Migration, Selection, Organization, Evaluation, etc. are its examples.

4. **P = Property :** It includes manifestations denoting the concept of 'attribute' - qualitative or quantitiative, e.g., Property, Effect, Power, Capability, Efficiency, Utility, etc.

5. **M =Modifier :** In relation to a manifestation of any one of the elementary categories, D, E, A, and P, the term Modifier refers to an idea used or intended to be used to qualify the manifestation without disturbing a conceptual wholeness of the latter, e.g., 'Infections' in 'Infectious disease'.

Working of New Version of POPSI

Example : Implications of Technological Changes in Rice Cultivation for India

1. **Analysis**

    D = Agriculture

    E = Rice

    A to E = Cultivation

    P of A to E = Technology

    P of P of A to E = Changes

    P of P of P of A to E = Implications

    M of the whole subject preposition = India

2. **Formulation**

    Agriculture (D), Rice (E), Cultivation (A to E), Technology (P of A to E), Changes (P of P of A to E), Implications (P of P of P of A to E), India (M of the whole subject proposition).

3. **Standardization**

    This step is concerned with semantics, and helps in standardizing the terms. Agriculture (D), Field Crop, Rice (E), Cultivation (A to E), Technology (P of A to E). Change (P of P of A to E), Implication (P of P of P of A to E), India (M of the whole subject proposition).

4. **Modulation**

    It augments the standardized subject formulation by interpolating and extrapolating the superordinates.

    Agriculture (D), Field Crop, Rice (E), Cultivation (A to E), Technology (P of A to E), Change (P of P of A to E), Implication (P of P of P of A to E), India (M of the whole subject proposition).

5. **Preparation of Entries for Organizing Classification**

    Here, the numbers of the POPSI Table are used to show the categories and the

positions of the components in the modulated subject formulation.

Agriculture 6 Field Crop, Rice 6.1 Cultivation 6.1.2 Technology 6.1.2.2 Change 6.1.2.2.2 Implication 6 (for) Asia, India

6. **Decision about Terms-of-Approach**

It is to be done carefully so as to achieve economy. Synonyms are controlled and references are generated from synonyms to standard terms. So, use each term other than 'Agriculture', 'Change', and 'Implication' as Term-of-Approach.

7. **Preparation of Entries for Associative Classification**

Here the entries are prepared under each term-of-approach, e.g. : Field Crop Agriculture 6 Field Crop, Rice 6.1 Cultivation 6.1.2 Technology 6.1.2.2 Change 6.1.2.2.2 Implication 5 (for) Asia, India.

Similarly, entries are to be prepared under each of the other Terms-of-Approach, or upper link terms may be used as heading to make it multifaceted. For example, the heading with 'Technology' as the Term-of-multifacted. For example, the heading with "Technology' as the Term of Approach may, in this case, consist of : "Technology, Cultivation, Rice".

8. **Alphabetization**

In the last step arrange the entries according to the alphabetical order by Terms-of-Approach.

Utility of POPSI

Although it is considered an extension of Chain Procedure, of indexing, yet POPSI is not based upon any scheme of classification. However, it uses the ideas and theory of classification, both in the analysis of subjects and also in the structuring of the names of subjects. In fact, POPSI can be used for various purposes.

i.   To construct subject headings for indexing purposes and for use as feature headings.

ii.   To generate subject index entries for classified catalogue, and back-of-the-book index.

iii. To determine the subject of the users' queries in a consistent and helpful manner.

iv.   To formulate a strategy for information search.

v.   To derive a base for the presentation of ideas in the text of a document.

Thus, it has imense utility, and that is why it is applied in the preparation of bibliographies and indexing lists. This system is used in a publication of ICSSR : Mohandas Karamchand Gandhi : A Bibliography, New Delhi : Orient Longman, 1974. This system is also used in the Machine Tool Abstracts of the Central Machine Tool Institute of India.

Chain Indexing : Developed by S.R. Ranganathan, it is a mechanical procedure to

derive subject index entries from class number of a document. It explicitly uses the classification number to assign index terms. It was conceived in 1930 and put to use in 1936. The components of a compound subject in chain indexing get automatically arranged in the order of Time, Space, Energy, matter and Personality. British National Bibliography followed this system from 1950-1970. LISA, Current Technology Index, British Catalogue of Music also uses chain indexing.

Chain procedure of indexing is developed by S.R. Ranganathan. It is a mechanical method which is used to derive subject headings from the class number of documents. The concept of 'chain' is the foundation of chain indexing. A chain is deemed to be a structural manifestation of a subject. Thus, a chain is said to be a modulated sequence of sub-class or isolate ideas. Therefore, the chain is derived from a classification scheme by converting the class number into natural language. These subject headings used for indexing are selected or sought subject headings having specific characteristics suitable to users.

Ranganathan defined chain procedure of indexing in his 'Classified Catalogue Code' (Ed. 5, 1964) as follows :

> "Procedure for deriving Class Index Entry (i.e. Subject Index Entry) which refers from the name of a class to its class numbers in a more or less mechanical way."

Therefore, this method may be used to derive Class Index Entry in a classified catalogue, and Specific Subject Entries, Subject Analyticals, and See also Subject Entries in a dictionary catalogue.

Characteristics of Chain Indexing :

Guha has identified the following characteristics of Chain indexing :

1. The main characteristics is the explicit use of a classificatory base.
2. It makes use of a classificatory language.
3. There is hardly an idea of 'marking-and-parking' function of classification.
4. Indexing language to represent any specific subject co-extensively.
5. The central ideas are the concepts of Facets and Fundamental Categories.

Steps in Chain Indexing :

The class number of a compound subject constructed according to a scheme for classification, may form the basis for applying chain procedure. In such a case, the following steps are involved in chain procedure :

1. Determination of the specific subject by analysing the subject content of the document, based upon the title.
2. Representation of the name-of-the specific subject in terms of its fundamental components. All the auxiliary words from the title are removed. In other words,

specific subject is represented in Kernal terms.

3. Determination of the category or status or role of each fundamental component according to a set of principles and postulates. This is an 'Analysis of the name of Subject'.

4. Transformation of the analysed name-of-subject by rearranging, if necessary, fundamental concepts, according to after additional postualtes and principles.

5. Standardization of each term in the transformed name-of-subject. If the standard term is not adequate in the scheme of classification, it should be replaced with an equivalent standard term as given in the schedule, or tesaurus, or a glossary.

6. Translation of the name-of-the-subject of the document in standard terms into the class number in the preferred classificatory language.

7. Determination and representation of the chain of which the name-of-the-specific-subject is the 'Last Sought Link'. The name-of-the subject is to be allotted standard term, which may be done as follows :

    (a)   Make the first link from the first digit.
    (b)   Make the second link out of the two digits and so on, upto the last link.
    (c)   Write the links one below the other in succession.
    (d)   Write against each link its translation in natural language.
    (e)   Connect each link with its translation by an "=" sign, and
    (f)   Join the "=" sign of each link with that of the next succeeding link by a downward arrow, if necessary.

8. Determination of different kinds of links, e.g., False Link, Unsought Link, Sought Link, and Missing Link.

(a)   **False Link (= FL) :** A link which does not represent any valid concept, and mostly it is a connecting symbol, or digit representing Phase Relation.

(b)   **Unsought Link (= UL) :** A link which is not likely to be used as access point to seek material.

(c)   **Sought Link (= SL) :** A link which is neither false nor unsought. It denotes the concepts that the user is likely to access.

(d)   **Missing Link (= ML) :** A link in a chain-with-gap. It represents that concept which is not available in the preferred classification scheme.
       It is inserted by the indexer at the proper place by means of verbal extension at the chain-with-gap whenever there is such a need.

9. Derivation of a subject heading from each of the Sought Links in the chain. Derive specific subject heading for the specific subject entry from the last sought link. Then move towards the term(s) of upper links, and if necessary, in a reverse/backward rendering process. If the subject includes a space or a time isolate or a form, then break the chain into parts of the point(s). In that

case specific subject heading is to be derived from lst sought link of each part in reverse rendering process. During this process, the chain just before the part denoting time, space or form must be inverted, and the rest of the chain appended to it.

10. Derive subject reference heading for the subject reference from each of the upper sought links. This process will continue until all the terms of upper sought links are exhausted and indexed.

11. Prepare subject references or 'See also' references from each subject reference heading to its specific subject heading. When a subject heading begins from last sought link denoting space, time or form, prepare 'See' references instead of 'See also' references from subject headings to specific subject heading.

12. Prepare 'See' references for each alternative and synonymous term used in the specific as well as specific reference headings.

13. Merge specific subject entries, subject references (i.e., 'See also' references) and 'See' references. These should be arranged in a single alphabetical sequence.

Examples :

1. To illustrate the point a simple example of a book "Retinal diseases of the eye in Children" is taken. The classifier fully analyses the content and translate this into the following formulation in standard terms :

    Medicine (BF).   Child (Special).   Retina (IPI)
    Disease (E).   Treatment (2E).

This is translated into the classificatory language as follows :

    L9C, 1857: 4:6

The indexer/subject cataloguer starts from this point. The subject headings prepared from class number are transcribed in an unidimensional form. If these are represented without the facet indications, they look like a single chain.

L	=	Medicine
	↓	
L9C	=	Child
	↓	
L9C, 185	=	Eye
	↓	
L9C, 1857	=	Retina
	↓	
L9C, 1857:4	=	Disease
	↓	
L9C, 1857:4:6	=	Treatment

(Here, False Links and Unsought Links have been climinated).

Subject Heading

1. Treatment, Disease, Retina, Child, Medicine

To provide access to various subjects, apart from the last link, entries will be prepared from other links :

Subject Reference Headings

1. Disease, Retina, Eye, Child, Medicine
2. Retina, Eye, Child, Medicine
3. Eye, Child, Medicine
4. Child, Medicine
5. Medicine

Cross Reference Entry

Retina, Eye, Child, Medicine

See also

Treatment, Disease, Retina, Child, Medicine

There will be similarly references from all the other headings also.

2. Geetanjali of Rabindranath Tagore

Class Number   0157, IM61, 2

0	=	Literature (Sought Link)
	↓	
01	=	Tentonic Literature (Unsought Link)
	↓	
015	=	Sanskrit Literature (Sought Link)
	↓	
0157	=	Bengali Literature (Sought Link)
	↓	
0157,	=	(False Link)
	↓	
0157, 1	=	Poetry, Bengali Literature (Sought Link)
	↓	
0157, IM61	=	Tagore (Sought Link)
	↓	
0157,IM61,	=	(False Link)
	↓	
0157,IM61,2	=	Geetanjali (Sought Link)

Subject Headings

Geetanjali	=	0157, IM61,2
Tagore	=	0157, IM61
Poetry, Bengali	=	0157, 1

Bengali Literature	=	0157
Sanskrit Literature	=	015
Literature	=	0

Application :

1. The British National Bibiligraphy (BNB) adopted chain procedure of indexing in 1950, based on class number constructed from DDC, and continued upto 1970. From 1971 onward, a new indexing system PRECIS replaced chain indexing.

2. The Indian National Bibliography (INB) has been following chain indexing with some modifications since its inception.

3. Chain procedure was also used by E.J. Coates for alphabetical subject index and for derivation of subject headings for British Technology Index. A modified version was later used in Current Technology Index which replaced British Technology Index.

Advantages :

The chain indexing has some of the advantages as given below :

1. The classifier analyses the subject of a document for formulation of the subject based on class number.

2. Classes are retranslated to provide alphabetical approach through Class Index entries.

3. It is based on classification number and terminology given in the schedule of classification system adopted.

4. Chain indexing is more suitable for analytico-synthetic scheme of classification.

5. It can be applied to any classification scheme in which notation symbols indicate the subordination of each step of division.

6. It provides alternative approach to the classified file through reverse rendering, thereby helping in retrieving the needed information.

7. Chain indexing is found to be fully amenable to computerization.

Limitations :

1. It has been seen in the examples given above that a subject heading derived through chain consists mainly of concept terms which constitute a subject. The correct meaning is brought out by a pre-determined sequence. The terms are in the reverse order of the digits of the class number. Hence, they are in the sequence of abstract to concrete, and at the same time, are in the order of specific to general. The problem, therefore, is, that of the citation order.

2. Another limtation is the problem of disappearing chain. In the example, it is clear that only the index entry derived from the last link is co-extensive with the specific subject of the document. All the other headings are actually broader

subject headings. Hence, strictly speaking, there will be only one specific subject entry in the whole index.

3. Chain indexing is closely linked with the scheme of classification. Therefore, the index would reflect all the defects and shortcomings of the scheme. The indexer could hardly add anything from his own understanding of the subject.

Improvement in Chain Indexing :

1. In view of the above mentioned limitations of cittaion order, disappearing chain and difficulty of rigidity by the classification scheme, some efforts had been attempted to make improvement in chain indexing. Ranganathan himself suggested different citation order by following Forward rendering method. He stated that instead of deriving the specific subject from the Isst Sought Link, it could be from the first Sought Link to the last. Accordingly, in the Example 1 above, it would be :

    Medicine, Child, Retina, Disease, Treatment.

2. Regarding the problem of disappearing, it has been suggested to follow the rotation of constituent terms of the heading. Instead of dropping the terms one by one, rotation of the terms would keep all the headings complete. It would be :

    1. Child, Eye, Retina, Disease, Treatment/MEDICINE
    2. Disease, Treatment/MEDICINE, Child, Eye, Retina
    3. Eye, Retina, Disease, Treatment./MEDICINE, Child
    4. MEDICINE, Child, Eye, Retina, Disease, Treatment
    5. Retina, Disease, Treatment./MEDICINE, Child, Eye
    6. Treatment/MEDICINE, Child, Eye, Retina, Disease

3. The problem of economy is not so easy to be solved. It could be achieved either by eliminating permutation of component terms, or another set of rules. But permutation may complicate the rules of syntax and economy may be sacrificed.

The major limitation of pre coordinate indexing is that the file may fail to provide a particular combination of access points as per user's point of view. So the dissatisfaction with pre-coordinated indexes lead to devise schemes which bypass the deciding on an order of terms that controls the approaches. The result was emergence of post coordinate indexing systems.

Post-Coordinate Indexing : With the advent of computers post coordination gained momentum, though manually also it is being practiced. The major contributions are by W.E. Batten, G. Cordonnier, Calvin Mooer, Mortimer Taube et al. In Post-Coordinate indexing, usually called 'Coordinate Index', the words and phrases selected at the time of indexing can either be searched individually or be combined at the time of searching. The post coordinate indexing systems are Key word indexing - KWIC, KWOC; Uniterm indexing; automatic indexing. In this method of indexing the subject

headings or descriptors assigned to documents represent simple concepts that the user must combine at the time of searching to retrieve information on a complex subject.

Example : Annotation + Bibliography for "Annotated bibilography".

2.2.4.4   Keyword Indexing : KWIC index, keyword in context index, a king of indexing developed in 1958 at IBM by H.P. Luhn. Significant words from the title and/ or text of the work or works are used as headings. The KWIC index system accepts an ordered set of lines, each line is an ordered set of words, and each word is an ordered set of characters. Any line may be "circularly shifted" by repeatedly removing the first word and appending it at the end of the line. The KWIC index system outputs a listing of all circular shifts of all lines in alphabetical order. Practical instances of it are widely used by computer scientists. For example, the "permuterm subject index" of Citation indexes of Institute of Scientific Information, Philadelphia is essentially such a system. KWAC (Keyword alongside context) and KWOC (Keyword out of context) are modifications of KWIC. KWIC, KWAC, and KWOC are simple, mechanical term extraction indexes for text (usually titles) which retain some of the context (i.e. adjacent words).

Examples :

KWIC:

for Croatians.	**Cataloguing** and classification
Cataloguing and	**Classification** for Croatians.
for	**Croatians.** Cataloguing and classification.

KWAC

Cataloguing and classification for Croatians.

Classificator for Croatians. Cataloguing and

Croatians. Cataloguing and classification for

KWOC

Cataloging Cataloguing and classification for Croatians.

Classification Cataloguing and classification for Croatians.

Croatians. Cataloguing and classification for Croatians

Uniterm indexing : The system was introduced by Mortmer Taube in 1953. It uses single terms as opposed to composite headings. The subject content of a document is anlyzed and individual concepts separated. For example if the subject of a document is represented by five terms, then five separate entries or term cards are prepared. Searching for relevant documents is simple because the user can select a document(s) that matches all relevant concepts.

Example : Treatment of retinal diseases. Select terms : **treatment, diseases, retina.** After this simple formulation, the term cards for each one of the three terms will be picked up one by one from their alphaebtical positions; the posted book numbers are

compared to find common numbers between all the cards. The common numbers that are found in all cards are the most sought documents. Further details can be had from the accession file.

Automatic indexing : it is creation of catalog or indexing records using computer programs, not human cataloguers. H.P. Luhn of IBM was pioneer in this field. Each significant word in the title becomes an entry point, appears in the middle, with the rest of the title on either side.

Example :

digital computation/**Field** plotting by Fourier synthesis and

computation/Field **plotting** by Fourier synthesis and digital

Field plotting by **Fourier** synthesis and digital computation

plotting by Fourier **synthesis** and digital computation/Field

Fourier synthesis and **digital** computation/Field plotting by

synthesis and digital **computation**/ Field plotting by Fourier

"Automated indexing software" is, according to the common definition, software that analyzes text and produces an index without human involvement. A good automatic index will exist once there's good artificial intelligence, something that presently doesn't exist.

2.2.4.5.   Citation indexing : A citation in the literary sense means any written or spoken reference to an authority or precedent or to the verbatim words of another speaker or writer. In library usage, a written reference to a specific work or portion of a work (book, article, dissertation, report, musical composition, etc.) produced by a particular author, editor, composer, etc., clearly identifying the document in which the work is to be found. The frequency with which a work is being cited is considered as a measure of its importance/value in the literature of the field. Citation format varies from one field of study to another but includes at a minimum author, title, and publication date. An incomplete citation can make a source difficult, if not impossible, to locate. Citation style manuals and online style guides are available to make a standard pattern of citations of other works.

CITATION INDEXING

The limitations of the traditional indexing systems were mainly responsible for the development of a non-conventional indexing system. Citation indexing is one such indexing system. It represents an entirely new approach to the problem of file organization. It does not depend for its indexing technique upon (a) either word derivations from text, or (b) word assignment.

Association of Ideas

Citation indexing is essentially a method of detecting the relationship of documents through citations. A citation index, according to Eugene Garfield is "an ordered list of cited articles each of which is accompanied by a list of citing articles". The citing

article is identified as a source, and the cited article as a reference. When a document refers to some other documents it can be assumed that there is some similarity or association of ideas between the citing documents, and the cited documents. An author usually mentions along with his work all those documents which he has consulted or referred to. This is called citing or referring documents. The list of references is a key part of any scientific paper, since it helps to put the research into its proper context in the development of that subject. Thus they try to show the relevance of their work to what has already been done in that field.

Citations are the references made to other documents. A document to which a reference is made is called a cited document. A document which makes reference to the cited document is called a citing document. Thus, a citation implies a relationship between a part or the whole of a cited paper and a part or the whole of a citing paper. The basis of citation indexing is the relationship between the cited documents and the citing documents. By following the subsequent citations, the history of an idea can be traced where and how it has been applied and whether it is sustained, rejected or absorbed into later work.

Development :

Considered as a new approach to the bibliographic file organization, the basic idea of citation indexing is not new. The first application of this idea was made in 'Shepard's Citations', developed by Frank Shepared in 1873 as an index to American legal cases. It provides a listing of individual American Court cases, each case being followed by acomplete history written in simple code. The listing under each case being followed by a complete history writtenin simple code. The listing under each case displays the publications that have been referred to the case, other court decisions that have affected the case, and any other references that may be useful. This type of listing is particularly important to the lawyer to base his argument because, in law, precedent is considered relevant for future judgements.

Having seen the usefulness of Citation Indexing, Eugene Garfield stressed the need for compiling such as index in the field of science and technology in 1950s. After some preliminary studies he brought out an experimental Science Citation Index (SCI) in 1961. Institute for Scientific Information, Philadelphia published the first SCI in 1963 covering the literature of 1961. It has been a regular publication since then. Since 1972, Social Science Citation Index and since 1978 Arts and Humanities Citation Index are also being brought out by the Institute.

SCISEARCH, the online version of SCI, is being published from 1974. It is available through some major online host systems, such as DIALOG, etc. The CD version of SCI is also available. A new feature **related records** is also incorporated. It leads the searcher to other records having references in common with the ones already retrieved.

Science Citation Index (SCI)

Institute for Scientific Information, Philadelphia, under the dynamic leadership of Dr. Eugene Garfield developed **Science Citation Index**. He nurtured SCI and brought it to the present position of the popularity. It is perhaps the best example of the integrated search system for a comprehensive coverage of the periodical literature in science and technology. It covers about 3000 journals, and about 1200 books representing more than 100 subjects. It has three indexes : Citation Index, Source Index, and Permuterm Subject Index, though separate but related to each other.

Citation Index (CI)

A citation index inSCI is arranged alphabetically by authors. It contains two types of information : (a) information about a cited item (reference), and (b) information about citing items (source). An entry for a cited item consists of :

- Name of first author and initials,
- Year of publication of the cited item,
- Title of the document in abbreviated form,
- Journal in which published,
- Volume number.
- Starting page number.

In different papers of an author are cited, the papers are arranged in chronological sequence. If different authors in different places cite a particular document, all the citing items are displayed immediately under the cited item by alphabetical order of the source authors. A citing item includes the following information :

- The citing author's name and initials,
- Title of the source document/periodical,
- Year of publication,
- Volume number,
- Starting page of the citing item.

There is also coded symbol indicating the nature of citing item. For instance, A (for abstract), E (for editorial), N (for technical notes), etc. Further, Citation Index part includes anonymous cited documents. These are arranged separately alphabetically by titles. Another separate section within CI contains a Patent Citation Index. It is a list of all cited patents (foreign and domestic) arranged in numerical sequence by patent number. It provides the year of issuance, the inventor's name, and country.

Example (Cited reference followed by Citing reference)

Cited Author	Citing Author	Cited Year	Publication	Citing Year	Volume	Page
SWEAT JA		61	SCI		134	163
	BROOKS BC		J DOC	65	24	30
	BROWN HL		AM DOC	66	19	30
....		62	SP LIB		31	56
	COOPER WS		J DOC	68	L29	22

In this example, a paper by JA Sweat published in the year 1961 in the periodical <u>Science,</u> Volume 134, page 163 has been cited in two papers. BC Brooks cited it in his paper published in 1965 in Journal of Documentation, Volume 24, page 30, and HL Brown cited in his paper published in <u>American Documentation</u> in 1966, Volume 19, page 30. The dotted line indicates that another paper by the same author, i.e., JA Sweat, has also been cited by W.S. Cooper and its bibliographical details are given. The document code L indicates that the citing item is a letter.

Source Index (SI)

It is a complete author index of citing items, i.e., all items in source periodicals included in the SCI. It is arranged alphabetically by the last name of the first author of the source item. An entry for a citing item contains :

-       Name of the first author,
-       Names of the co-authors (upto ten), if any,
-       Full title of the citing (source) item, i.e., article,
-       Periodical title, i.e, source document,
-       Volume number
-       Issue/part/supplement number,
-       Starting page number,
-       Year,
-       Coded symbol for type of document (review, letter, etc.)
-       Number of refeences in the bibliography of the citing item,
-       Accession Number of the source periodical (as filed at ISI),
-       Followed by the full title of the article.

The SI has two more sections on anonymous items, and a <u>Corporate Index.</u> Anonymous items are arranged alphabetically by the titles of periodicals. These are given at the beginning of the SI.

In the <u>Corporate Index,</u> all the source items are listed alphabetically by author under the name of the organization where the work was done. If more than one organizationis involved in a given project, an entry is created for each organization. The index will

indicate under each organization the names of staff who atuhored articles indicating periodical, volume, and page.

The SI can be used independently, because each entry is complete in itself. For example, one can find out if a particular author has published any new contribution.

Example

Citing Author	Co-authors	Journal	Vol	Page	Year	Code	No. of Ref.	Acc. No.
BROOKS BC		J DOC	24	41	65		12R	48284
THE MEASUREMENT OF INFORMATION RETRIEVAL EFECTIVENESS PROPOSED BY SWEAT (Full Title)								
BROWN HL	JAMES T	AM DOC	19	30	66		10 R	
EVALUATION OF DOCUMENT SEARCHING SYSTEMS (Full Title)								
COOPER WS	LIB ASSOC (Gr Britain)	J DOC	29	22	65	L	6 R	
EXPECTED SEARCH LENGTH (Full Title)								
JAMES T	SEE BROWN HL							

This example shows that HL Brown in collaboration with T James published a paper in the American Documentation in 1966 which appears on page 30 of volume 19, having ten references. The full title of this paper is also given in the next line. All co-authors in the SI are cross referenced (as shown in the last line).

Corporate Index Part

Organization attributed	Citing Author	Journal	Year	Vol	Page
LIBRARY ASSOCIATION (Gr. Britain)	COOPER WS	J DOC	68	29	22

In this index section under each organization is found the citation for all items attributed to that organization. Here, Cooper's article is listed under LIBRARY Association (Great Britain) as this article is attributed to that organization.

Permutation Subject Index (PSI)

In the PSI, significant terms are selected from the titles of papers appearing in the source periodicals. These terms are then permuted so that each of these terms serves as a primary term, and every possible two-word combination with every other the title as a secondary term. It also reveals the author of the paper in which the primary

term and co-term appeared in the title. The format of PSI as given in SCI and two related SI entries are displayed in the example below :

Example :

INTERSTELLAR

ATOMS	GREENEBER	JM	These authors used the word
BORON	MORTON	DC	"Interstellar" in the title of
....			their articles in addition to
....			the word shown opposite their
			names.
CLOUDS	CHAFFEE	FH	
	COHEN	JG	
	GOLDSTEI	JS	
COLD	CRUTCHER	RM	
COLLIDING	GOLDSTEI	JS	

Source Index Listing for two articles, from the above example, containing "Interstaller" and "Clouds" in their titles. This is displayed below :

```
CHAFFEE   FH
      LINE SPECTRA IN INTERSTELLAR CLOUDS
            ASTROPHYS J 189 (3) : 427-440       74       4ZR
GOLDSTEI JS
      MAZZELLA AJ - PARTIAL COALESCENE     MODEL FOR
      COLLIDING INTERSTELLAR CLOUDS
      NUOV CIMB 21 (1) : 142-150           74       9R
```

Search Strategy

The use of SCI is relatively simple. A searcher does not start with a subject; rather he begins with a reference or an author he has identified through a footnote, book, encyclopedia article, conventional word, or subject index. In CI part of the SCI, the search is carried out to identify that particular author's name. Finding author's name, searcher sheets to see which of several possible references fits the one he is interested in. Another search through the Source Index will provide complete bibiligraphical description of the source items.

This helps the searcher to judge the relevance of these documents to his query.

Though CI is highly specific, yet a search may be expanded in order to build a more extensive bibliography for a particular query. Once the searcher locates a number of source articles, he can use the bibliographics of one or several of these as other

entries into the Citation Index. This process is called "Cycling".

In case the searcher is not aware of any reference, he may start search through Permutation Subject Index (PSI). He is required to compile a list of terms relevant to his topic, and look into PSI to identify one of these terms used as primary term. He can simultaneously locate the author using the term in his article title. Thereafter, he can look into Source Index to identify complete bibilographical detail of the document. Using primary term and co-terms he can get more defined subject, and can locate appropriate authors to be used in SI. This information can be used in Citation Index to locate subsequent citing sources.

Advantages

1.   It enables to link up all recent papers on a subject with papers published earlier having association of ideas.

2.   It is unaffected by terminology problems, whereas the conventional subject indexes have problems of changing terminology.

3.   It does not involve any intellectual activist in its compilation, as it is fully computerized depending upon citing and cited documents.

4.   This index is faster and more efficient for searching than conventional indexes.

5.   It provides the bst access to interdisciplinary literature.

6.   It helps in tracing the genesis of interidisciplinary or newly emerging subjects.

7.   It can help an author of a cited document to know how far his ideas or research results have been appreciated, applied, or citicized by others.

8.   It can be used as a tool to identify the interconnected papers on a subject to trace its history and development.

9.   It can be used as a measuring tool for the progress of science, as a means of automatic content analysis, and as a way of evaluating journals (to find out the core ones), papers, auhtors, etc. for awarding prizes, promotions, appointments.

10.   It reveals the obsoloescence of the published material.

11.   It is a self-updating index as each new citation of an earlier paper is automatically listed making the list of citing documents upto date.

12.   It avoids "noise" or irrelevancy by using careful selection of references as access points.

Disadvantages :

1.   It does not provide any logical or conventional subject arrangement to which the users are accustormed.

2.   It is necessary that to begin with the search, a user should know at least one reference of a document on the subject.

3.   A citation index retrives only the related documents, and not the contents of the documents.

4.   Another disadvantages of CI is that it does not satisfy the exhaustive approach of a user.

5.   It is said that a search in the SCI may retrieve high proportion of irrelevant items. In other words, it suffers from a low precision ratio.

6.   Citation index depends on the correct citation practice of authors. Any discrepancy at this level will generate "noise" at the output stage.

2.2.4.6.   Indexing and the Internet : The World Wide Web facility of the Internet has made information resources available to a wide audience of potential users. The size ofthe internet also implies great variety in the resources available like web pages, FTP files, websites etc in which thousand of pages of text. Images andsound are stored in HTML or XML format. To provide access to such vast amounts of information, currently different approaches are in practice. Broad subject guides or information gateways or directories like SOSIG. BUBL etc have been prepared to access Internet resources. Besides, software tools like search engines and meta search engines are designed to allow users to perform relatively specific subject searches of these sources.

2.2.5. Summary : Indexing as discussed is a tool to retrieve documents from the document store, whether it is conventional library or web resources. To perform this job indexing systems with different characteristics have been developed. Broadly they can be categorized as assigned i.e. standard controlled vocabulary used to assign index terms and derived indexing i.e. vocabulary selected from natural language as used by author from title abstract or keywords in the text. The alphabetical indexing is the traditional and easy method of organzing the indexes that may be either author title, or subject separately or may be a single index combining all. Another important categorization is pre and post coordinate indexing systems. In the first type coordination is being done in anticipation of search by user. i.e. predetermined while in later coordination will be done at the time of search by user. A good num ber of systems like PRECIS, POPSI, and Chain indexing etc. developed under pre-coordination while Keyword, Uniterm, automated etc as post coordinated systems. The application of computers paved way for automated index generation through keyword and citation index. Thus different types of indexing systems are aiding information retrieval to print as well as web information sources.

2.2.6. Glossary

Citation : a written reference to a specific work or portion of a work (book, article, dissertation, report, musical composition, etc.) produced by a particular author, editor, composer, etc., clearly identifying the document in which the work is to be found.

Internet : A network of computer networks which operates world-wide using a common set of communication protocols.

Search engine : the software used to retrieve information from a database or from

the Internet (eg, the WebLUIS catalog or Yahoo!). A search engine generally includes features such as Boolean operators, search fields, display format, etc. E.g. Google, Yahoo, Lycos

2.2.7. References and further reading

1.   Buckland, M. (2003) http://www.sims.berkeley.edu/courses/is245/s03/verbal.html

2.   Foskett, A.C. (1982) The subject approach to information, 4th Edition. London, Clive Bingley.

3.   Kumar, P.S.G. (2004) Information analysis, repackaging, consolidation and information retrieval. New Delhi, B.R. Publications.

4.   Lancaster. F.W. Abstracting and indexing in theory and practice.

5.   ODLIS. URL: http://lu.com/odlis/odlis_s.cfm#semanticrelation

6.   Prasher. R.G. index and indexing systems. New Delhi, Medillion, 1989

7.   Taylor, A. (2000). Wynar's Introduction to Cataloging and Classification. 9th edition. Littleton, Colo. : Libraries Unlimited. (p. 408-411).

LESSON NO. 2.3   **AUTHOR : R.S.R. VARALAKSHMI**

INFORMATION STORAGE AND RETRIEVAL SYSTEMS

Structure :

2.3.1. Objectives

The objective of this lesson are to make you

* to understand the dimension of information representation.
* to understand the components of information storage and retrieval system.
* to understand the various formats and standards for organization of information.
* to understand the process of information retrieval, manual and online.
* to know the importance of search and search stagies in the retrieval of information.

2.3.2. Introduction

The graphic record of humanity is represented by its accumulated data, information, knowledge and wisdom. This is being stored in different forms like text, graphic, image, sound etc and on different formats such as paper, audivisoual, microforms, and electronic media. There is a need to store, retrieve and preserve this documentary information for the use by present generation and for posterity. Therefore it is necessary to organize the store of information to facilitate retrieval of the same

44

according to user's requirements. In this context this lesson explains you the features of information storage and retrieval system (ISAR). We will discuss the role of index as a means of searching and matching that shows the way of recognizing whether or not any given document contains the information we want.

2.3.3. Information Storage and Retrieval (ISAR) -Definition

ISAR is the systematic process of collecting and cataloguing data so that they can be locted and displayed on request to selectively recall recorded information from a file of data. It can be defined as "the ystematic process of collecing and cataloguing data so that they can be located and displayed upon request." (The Concise Columbia Electronic, Encyclopedia, 1994)

In other words ISAR is the process of recovering information bearing symbols from a storage place in response to queries from intending users of the information. In libraries generally, searches are typically for a known item or for information on a specific subject, and the file is usually a human-readable catalog or index, or a computer-based information storage and retrieval system, such as an online catalog or bibliographic database. However, computers and data processing techniques have made possible the high-speed, selective retrieval of large amounts of information for government, commercial, and academic purposes. There are two basic types of information storage and retrieval systems.

I.   *Document-retrieval systems* store entire documents, which are usually retrieved by title or by key words associated with the document. This permits full text searching, enabling retrieval on the basis of any words in the document. In others, a digitized imageof the document is stored, usually on a write-once optical disc. These Database systems store the information as a series of discrete records that are, in turn, divided into discrete fields (e.g., name, address, and accession number); records can be searched and retrieved on the basis of the content of the fields. The data are stored within the computer, either inmain storage or auxiliary storage, for ready access.

II.   *Secondary/Reference-retrieval systems* store references to documents rather than the documents themselves. Such systems, in response to a search request, provide the titles of relevant documents and their physical locations. For example the catalogues, OPACS, and other such systems. They have proven extremely effective in libraries, where material is constantly changing.

2.3.4 Information representation

The basic functions of an information store are :

    (i)   Subject analysis of a document by an indexer,
    (ii)   The translation of the subject analyzed into the indexing language,
    (iii)   The organization of the files of which the database is comprised.

Generally information is represented in a store after subject analysis with a purpose

to recall. The analyzed concepts provide the access points that are represented in notational form e.g. class number or as subject headings or index terms in natural or controlled vocabulary. As already discussed in the previous lessons different indexing languages and techniques are in use to represent a document in the store and provide access points for its easy retrieval. Therefore the information can be described and represented as a bibliographic record in a store as follows :

(a)   Traditional cataloguing and classification;

(b)   Indexing languages - controlled and uncontrolled e.g. L.C Subject Headings, MeSH

(c)   Encoding standards e.g. MARC, Metdata, Dublin Core, XML

Bibliographic record is a record consisting of bibliographic etails of documents, their storage and maintenance. May be a library catalogue or an indexing or an abstracting periodical or a bibilography or a documentation list, either in printed or in machine-readable form. In a computer-based system however, the master bibliographic record for an item is held in the machine and one or more computer programs generate entries from it. The entries are displayed as OPAC (Online Public Access Catalogues). The degree of compatibility among library/information centres depends on the extent to which they follow a common set of standards for information handling in manual or computerized information store. Some important standards in application are ISBD, ISO 2709, andMARC 21 etc.

Organizing Internet Resources : As the Internet becomes an accepted source of electronic information; libraries and information specialists strive to improve methods for the description, organization and retrieval of remotely accessed digitized objects. Designing ways of navigating the wide range of resources is a major challenge. The use of various ross-searching and linking technologies is now high. New techniques of resource description and access have been developed viz., RDF (Resource Description Format), Metadata formats such as Dublin core, Text Encoded Systems (TEI) etc and Z39.50 is a major facility in this area.

Self Check Exercise

How does the information represented in information store or a database ?

2.3.5. Information Retrieval - the features

Information retrieval is the proces,s methods, and procedures used to selectively recall recorded information from a store or file of data. In libraries, searches are typically for a known item or for information on a specific subject, and the file is usually a human-readable catalog or index, or a computer-based information storage and retrieval system, such as an online catalog or bibliographic database.

Definition : Harrods glossary defines Information Retrieval as "Finding documents or the information contained in documents in a library or other collection, selectively vary from simple index or catalogue to the documents in a computer based system. Classification, inexing and matching, searching are all systems of information retrieval"

Information retrieval (IR) is the art and cience of searching for information in documents, searching for documents themselves, searching for metadata which describe documents, or searhing within databases, whether relational stand alone databases or hypertextnetworked databases such as the Internet or intranets, for text, sound, images or data. (Wikipedia).

Role of Indexing in Information Retrieval : Indexes have an important role to play in information retrieval.

(i)   Indexes acts as a link between a source of information and its user. It informs the existence of documents containing documents surrogates, such as author, title, imprint and call no. etc.

(ii)   An Index is a systematic guide to concepts derived from a collection of documents represented by entries arranged in aknown and searchable alphabetical, numerical classified order.

(iii)   If size of collection is quite large then index plays a major role to retrieve relevant Information. A good index minimizes the search efforts and ensures optimum results.

(iv)   In Information Retrieval System, index performs two simultaneous functions:- Retrieving Information from documents that are required and Holding back Information about required documents based on a particular subject.

(v)   In the context of information retrieval, the term index is primarily used as a tool capable of retrieving information about required documents.

(vi)   The two characteristics of indexing i.e., exhaustivity and specificity affect twoimportant measures of information retrieval viz. recall and precision, which operate the search stage or output stage of the system.

        The IR process is possible by manual or machine methods. In the manual process, to avoid the problem of scanning primary literature, a number of secondary sources like bibliographies, indexing and abstracting periodicals etc. are developed but the process of searching and locating informationis

tedious. Hence, the current trends is to develop the information store in amachine readable form and use a computer for searching the desired information. Whatever be the method of operation, the retrieval comprises three main functions :

(a)   The analysis of the search question ; Translation of concepts in the question into the indexing language of database.

(b)   The formaulation of search statement i.e. the order in which the terms chosen are to be sent to computer, the relationships between the terms etc.

(c)   Data out put that include display or distribute the data.

Thus search is the key component between the store and the user and the quality of search strategy defines the effectivenss of ISAR.

The structure of information retrieval system

| Selection of items | → | Organization of records using indes/ Controlled ocabulary | → | Development of Primary store/ database | → | Development of catalogue/reference database |

Matching request with reference database ↑

Search strategy ↑

Search the store ↑

| Need arises | → | Request made | → | Conceptual analysis of subject in request | → | Translation of request into indexing language | ↑
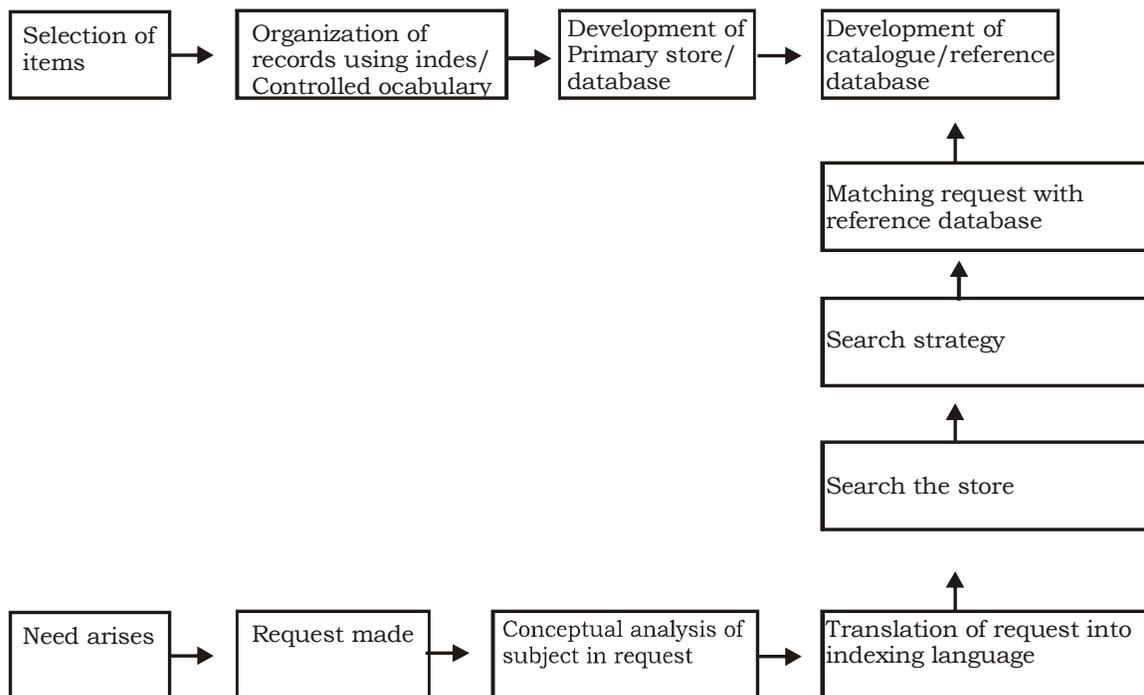
2.3.6. Evaluation of information retrieval

So far we have discussed the characteristics of information storeage and retrieval process. Now we have to know the methods of measuring their performance. Evaluation is essential to assess the performance of the retrieval system and the effectivenss of the processes adopted for retrieving desired documents. Two important measures used to ascertain the effectiveness of information retrieval are recall and precision, both related to the 'relevance' of documents/information to the needs of the requester.

Relevance: The retrieval system recalls some books from the store, using the indexing language, as per the user's request. The number of documents which are actually relevant to the user out of the retrieved books is known as relevance. It is the essence of a retrieval system as the system selects and uses an indexing language to retrieve as much relevant information as possible. The relenvance is expressed in recall and precision ratios.

When an index is consulted and asearch made in response to a qery, the collection can be visualized to be divided into two sets. The sets are documents retrieved and documents not retrieved. Further analysis of the sets may reveal that each set is having within it documents which are relevant and those which are not. Guha has explained them as follows :

	Retrieved	Not retieved	Total
Relevant	a (hit)	b (missed)	a+b
Not relevant	c (wasted)	d (dodged)	c+d
	------------------	--------------	-----------
	a+c	b+d	a+b+c+d=N

In the above matrix, a+c is what is retrieved and a+b is what is relevant. Out of a+c, only 'a' is relevant; so the size of 'a' determines the relevant out of total retrieved. Index efficiency can be obtained by determining the size of 'a' inralation to a+b. Since this measure indicates the amount of relevant documents in the store/file, it is called recall ratio.

Recall is the measure of the system's ability to let through wanted documents i.e., the degree to which it delivers all relevant documents. In other words, it is essentially a measure of how well the system performs at yielding up all the relevant items within it, and the extent to which relevant items are retrieved is referred to as the recal ratio. A simple formula to represent Recall ratio is :

$$\frac{a}{a+b} \times 100 \text{ i.e.} \quad \frac{\text{Relevant documents retrieved (in a search)}}{\text{Total No of relevant documents (in store)}} \times 100$$

Precision is a coplimentary measure that indicates the ability of the system in holding back unwanted documents. It describes how efficiently the system provides only the relevant items. Precision acts as filter in information retrieval system. It is measure of the system's ability to hold back unwanted item.

The user will judge the results of the search by their relevance to his requirements and also by the time taken to get the results. It is customary to express the results in terms of the proportion of items retrieved and delivered which the requester judges relevant. This porportion is called as precision ratio. In detrmining the recall ratio, only the number of the relevant documents retrieved is taken into consideration. The number of non-relevant documents rtrieved in response to a query (i.e. 'c' which

is a waste) if taken into consideration canindicate how precise is the indexing system in retrieving only relevant documents. A simple formula to represent Precision ratio is :

$$\frac{a}{a+c} \times 100 \text{ i.e. } \frac{\text{Total No of relevant documents retrieved (in a search)}}{\text{Total No of douments retrieved (in a search)}} \times 100$$

Infact recall and precision are inversely proportional to each other and hence cent percent recall and cent percent precision are not possible. If recall increases, precision direases and vice versa. Information retrieval system can also be judged on the basis of

(i)   Indepth analysis of documents
(ii)   Selection and development of suitable retrieval mechanisms
(iii)   Identification of user requirements
(iv)   Matching user requests wth stored/incoming information
(v)   Timeliness of suppy.

Self Check Exercise

What is meant by Recall ratio and Precision ratio ?

2.3.7. Characteristics of online information retrieval or online searching

Online is a general term for using devices and peripherals which are inteacting directly and simultaneously with a computer central processing unit (CPU). Online information retrieval is using a computer for searching and retrieving of selected information from the data held on a computer as database. When we are online, we are connected directly to the (CPU) of a computer using a terminal. This enables the user community to communicate directly in a conversational mode with a remote computer through network facilities.

Online information retrieval (or online searhing) is the acquisition of information

from a distant computer via is terminal or PC, invovling aninteractive dialogue between enquirer and computer. The computer handles a number of databases stored in electronic form, consisting of references to journal articles, conference papers, reports, books etc, which the Information Retrieval Service (IRS) or 'host' makes available to interested parties, such as university librries, on a commerical basis. The computer matches any input search terms against its files and displays any resulting matches which can then be printed out or downloaded by the searcher. For example the Internet is a storehouse of world's information. It is estimated that there are hundreds of millionsof pages on the web that is doubling almost for every fourmonths. Some important information sources on the Internet are e-books, e-journals, e-reference sources like encyclopaedias, dictionaries, directories, manuals etc., technical reports, theses, patents, educational material etc., in full text or bibliographic description. It also provides access to collections of a number of libraries through OPACs. In online, there are programs that help the searcher to browse or surf the net or websites known as 'search engines'.

2.3.8. Search and search strategies

Search: Words that are input to a catalogue/index/computer program to match user request against information in a store/database is called as search.

Search strategy : It is a systematic process used to find the most relevant information on a topic. During this process, one considers all potentially useful reference sources, selects works appropriate for the research need, locates information, and evaluates the information found. Series of search statements are organized to, etrieve records that will answer the intellectual search request. Harrods Glossary defines search strategy as "The plan adopted for answring a particular enquiry, or more specifically the search statements used to answer an enquiry."

Models of search strategies : Some common models are (a) Boolean logic model, (b) Extended Boolean model and (c) Fuzzy model.

Booleanlogic is more popular in manual and computer search methods. Search formulation involves several steps; formulation means to link the chosen terms in an order according to their relationship. The process involves

(i)   Selection of search/key words.

(ii)   Adoption of search logic either to expand or control the search to pertinent subject.

(iii)   Identifying appropriate data bases to start the search

(iv)   Conduct the search

Evaluate results of search out put

Fuzzy logic : Boolean Models are algebric models while fuzzy logic is derived from fuzzy set theory dealing with reasoning that is approximate rather than precisely deduced from classical predicate logic. It can be thought as the application side of

fuzzy set theory dealing with well thought out real world expert values for a complex problem.

Boolean search : Generally the relationships between terms are expressed in boolean logic operators AND, OR, NOT

E.g. Rivers and Salinity

OR placed between two key words generally for synoyms

E.g. Fruits OR Vegitables

NOT placed between two key words to avoid an unwanted term

E.g. Fruit NOT Apples
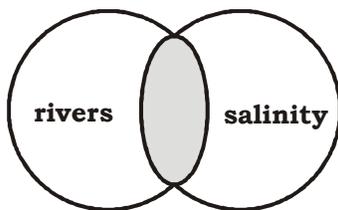
Venn diagram of
Boolean AND
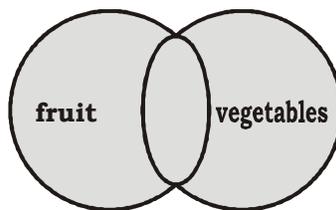
Venn diagram of
Boolean OR
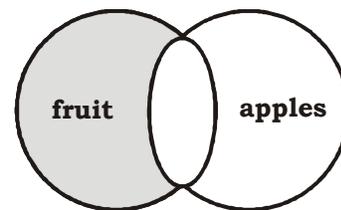
Venn diagram of
Boolean NOT

Truncation : Truncation is used to expand results by instructing the computer to look for the root word and all alternate word endings. Truncation symbols may differ depending on the database or engine in use. Some common truncation symbols are:

   * (asteric);

   ? (a question mark); and,

   # (the pound sign)

For e.g.

Computer

Computing

Computation

Compute

Computers

While doing the search the indexes or search engines automatically truncate the end words entered to look for 'plural' or ing or 'ed' and takes only the root term 'comput'

Search engines: A search engine is a searchable online databse of internet resources. It is a tool or program which allows keyword searching for relevant sites or information on the Internet. It has several components: seach engine software, spider software, an index (database), and a relevancy algorithm (rules for ranking). Internet searchengines (e.g. Google, Alta Vista) help users find web pages on a gien subject. The search engines maintain databases of web sites and use programs (often referred to as "spiders" or "robots") to collect information, which is then indexed by the search

engine. General and topic-specific search engines are prevalent today, for example, Education World, WebCrawler, Infoseek, Lycos, and Yahoo are examples of search engines. A search engine generally includes features such as Boolean operators, search fields, display format, etc.

2.3.9 Online search services

There are different kinds of online search services:

1. Search engines : These rely on "software spiers" to index Web sites. Examples of search engines are Google, Excite, and Lycos.

2. Directories : These rely on submissions from users and Web site owners to populate their indexes. Most directories add your site to their index, but generally they link only to your home page rather than indexing the full text of each page on your site. Examples, of directories are Yahoo! and Open Directory.

3. Meta search engines : A type of search tool that will allow a user to conduct a search across two or more search engines and directories at a time. These tools don't hold an index of their own, but are programmed to search across many of the top search tools simultaneously. Meta-search engines do not own a databse of Web pages; they send your search terms to the databases maintained by search engine companies.

4. Subject gateways : An Internet site set up to provide an easy link to other information servers giving useful subject based information in a particular discipline. Examples can be found in our Subject Pages. Organized lists of Web pages, divided into hierarchically grouped subject areas as a result of human effort. They are also known as subject directories. E.g. SOSIG, BUBL Information Services, Pinakes: a subject launch pad.

Operation of search engines: The search engines generally have the following strategies to retrieve information on the web.

Type of search: Keyword

Search options: Simple or Advanced search, search refining.

Domains searched: Web, Usent

Search refining: Boolean "AND," "OR" and "NOT"

Relevance ranking : Ranks according to how many of the search terms a page contains.

Results presented as: First several lines of document or "Detailed" summaries

User interface: Advanced query to further refining search at the end of each results page.

Self Check Exercise

Explain different search strategies

```



```

2.3.10 Summary

Information storage and retrieval systems have a long history and traditionally they are beign managed with the help of subject catalogues, indexing systems. The information explosion posed a problem for efficient organization of information storge and retrieval. The timely advent of computer and communication technologies paved new ways of information representation and retrieval. The most important comonent of retrieval process is search that matches the relevant document ofthe information store or database to the specific needs of the requester. There are several search strategies in practice like Boolean logic, truncation, Fuzzy sets etc. that are applied both in manual as well as online searching procedures. Online searching especially the Internet search is much more complex in view of the huge data it handles. Search engines, directories, subject gateways etc. have been developed and applied for effective searching of web resources.

2.3.11 Glossary

Information Retrieval (IR): The word was coined by Calvin Mooers in 1948-50. It is a broad interdisciplinary field thta draws on many other disciplines. IR is the art and science of searching for information in documents, searching for documents themselves, searhing for metadata which describe documents, or searching within databases, whether relational stand alone databases or hypertext networked databases such as the Internet or intranets, for text, sound, images or data.

Internet : The **Internet,** or simply the **Net,** is the publicly accesisble worldwide system of interconnected computer networks that tansmit data by packet switching using a standardized Internet Protocol (IP). It is made up of thousands of smaller commerical, academic, domestic, and government networks. It carries various information and

services, such as electronic mail, online chat, and the interlinked Web pages and other documents of the World Wide Web.

Online searching: Online information retrieval (or online searching) is the acquisition of information from a distant computer via a terminal or PC, involving an interative dialogue between enquirer and computer.

Precision : In information retrieval, a measure of search effectiveness, expressed as the ratio of relevant records or documents retrieved from a database to the total number retrieved in response to the query

Recall : In information retrieval, a measure of the effectiveness of a search, expressed as the ratio of the number of relevant records or documents retrieved in response to the query to the total number of relevant records or documents in the database Search strategy: In information retrieval, a systematic plan for conducting a search.

2.3.12 References and Further Reading

1.   Rowley, J.E. and Turner, CMD. (1977) The Dissemination of Information. London, Andre Deutsch.
2.   Guha, B. Documentation and Information. Calcutta, World Press, 1983.
3.   Foskett, A.C. (1982) The subject approach to information, 4th Edition. London, Clive Bingley.
4.   Answers. com URL: www.answers.com/topic/
5.   Wikipedia. URL: en.wikipedia.org/wiki/Wikipedia
6.   ODLIS: URL: http://lu.com/odlis/odlis_s.cfm#semanticrelation

---

LESSON NO. 2.4   **AUTHOR : DR. V. CHANDRA KUMAR**

VOCABULARY CONTROL

Structure :

2.4.0   LEARING OBJECTIVES

After studying this lesson, you will be able :

    (a)   To understand the concept of information retrieval system;

    (b)   To gain knowledge about the vocabulary control; methodology; trends and developments

2.4.1   OUTLINES

    1.   Introduction

    2.   Information retrieval system

    3.   Vocabulary control

    4.   Summary

2.4.2   INTRODUCTION

The term information retrieval was coined in 1952 by Colvin Moore and gained popularity in the information science community from 1961 onwards. The concept of information retrieval presupposes that there are some documents or records containing information that have been organised in an order suitable for easy retrieval. The documents or records concerned with contain bibliographic information which is quite

different from other kinds of information or data. If we take a database of information pertaining to an office, or a supermarket, all we have are the differentkinds of records and related facts, likes names of employees, their positions, salary, and so on, or in the case of a supermarket, names of employees, their positions, salary, and so on, or in the case of a supermarket, names of different items, prices, quantity, and so on. The retrieval system here is desgined to search for and retrieve specific facts or data, like the salaryof particular employee, or the price of a product, and os on. The major objective of bibliographic information retrieval system is to retrieve bibliographic details of those items containing the user's required information.

2.4.3 INFORMATION RETRIEVAL SYSTEM

The concept of information retrieval system is to some extent self-explanatory from the terminological points of view. One can simply denote such a system as one that stores and retrieves informaton. Three major areas that constitute an information retrieval system: items of information, users' queries, and matching of these queries with the document database. An information retrieval system, thus, has three major components - the document subsystem, the user subsystem, and the searching/ retrieval subsystem. These subsystems are quite broad and each one is designed to serve one or more functions, such as: (a) analysis of documents and organisation of information (creation of a document database); (b) analysis of users' queries, preparation of a strategy to search the database; (c) actual searching or matching of users' queries with the database, and finally (d) retrieval of items that fully or partially match the search statement.

An information retrieval system is designed to retrieve the documents or information required by the user community. An information retrieval system aims at collecting and organising information in one or more subject areas in order to provde it to user as soon as asked for. Thus, information retrieval system serves as a bridge between the information generator or creator and the users of that information.

The system input consists of documents. That is, the information centres based on the selection criteria and policies to acquire certain documents, which implies a detailed and accurate knowledge of the information needs of user community. Once the documents are acquired, they need to be 'organised and contrlled', so that they can be indentified and located in response to various types of user demand. Organisation and control activities include classification, cataloguing, subject indexing and abstracting.

Subject indexing process involves two distinct steps: 'conceptual analysis' or content analysis of a document and the'translation' of the conceptual analysis into a particular vocabulary or 'index language'. The index vocabulary, that is a limited set of terms that must be used to represent the subject matter of the documents. Such a vocabulary must be a list of subject headings, a classification scheme, a thesaurus, or simply a

list of 'approved' keywords or phrases. The uncontrolled vocabulary usually implies the use of words of phrases occurring in the documents being indexed.

The steps involved in the output side are similar to those involved in input. The user community submits various requests to the information centre, and the staff prepares search strategies for the requests. This process also involves two steps of conceptual analysis and translation. In the first step analaysis of request to determine what the user is really looking for, and the second involves the translation of the conceptual analysis into the vocabulary of the system. The conceptual analysis of the request, translated into the language of the system, is the 'search strategy'.

Once the search strategy has been prepared, it is 'matched' is against the database of the document representations. This could involve a search of card files, printed indexes, microfilm, or magnetic tape or disk. Document representations that match the search strategy, that is, satisfy the logical requirement of the search, are retrieved from the database and delivered to the user the enquier.

There are factors controlling the effectiveness of an information retrieval system. These may be separated into two groups: database factors; and factors associated with the exploitation of the database. The databse factors are 'input factors' which include: (a) what documents are included; (b) how completely and accurately the subject matter of these documents is recognised and represented in the indexing operation; and (c) how adequate is the vocabulary of the system to represent the subject matter of these documents. The exploitation factors are 'output factors' include: (a) how well the staff of the information centre is able to understand the information needs of the users (user-system interaction); (b) how well they can transform needs of the users (user - system interaction); (b) how well they can transform these needs into searching strategies; and (c) how adequate is the vocabulary of the system to represent the subject interests of system users.

Two broad categories of information retrieval system can be identified: in-house and online. In-house information retrieval systems are set up by a particular library or information centre to serve mainly the users within the organistaion. One particular type of in-house database is the library catalogue. Online public access catalogue (OPACs) provide facilities for library users to carry out searches, and then to check the availability of the item required. Online information retrieval system means as a system that have been designed to provide access to remote database(s) to a variety of users. Such services are available on commercial basis and available in CD-ROM technology. Recent developments in computer and communication technologies have widened the scope of online information retrieval systems. The Internet and World Wide Web have made information available for use by anyone virtually anywhere. This is leading us the concept of a digital library system where information can be

generated and made available in electronic form on the Web for use by any user from any corner of the world.

2.4.4 VOCABULARY CONTROL

Vocabulary control is one of the most important components of an information retrieval system. The term vocabulary control can be defined in a simple manner and reproduced as follows: Indexing may be thought of as a process of labelling items for future reference. Considerable order can be introduced into the process by standardizing the terms that are to be used as labels. This standardization is known as vocabulary control, the systematic selection of preferred terms.

As in the process of information retrieval, the process involves two quite distinct intellectual steps: the 'conceptual retrieval' and 'translation'. The second step in any information retrieval environment involves a 'controlled vocabulary'. The vocabulary (index language) of a retrieval system exerts a considerable influence on its performance. It greatly influences the construction of each search strategy, and the actual search (match with database). It is necessary to consider the major characteristics of a controlled vocabulary and the reasons why such vocabularies have been considered necessary in information retrieval system.

Particular topic of document representative may be represented in many different ways in different documents or by different indexers in which the indexers use natural-language terms to describe the subject matte of the document. In other words, lack of controlled vocabulary result inconsistency in the representation of identical subject matter.

Synonym control : A topic may be represented using a term or by any of the synonym to the term. This uncontrolled vocabulary situation the searcher must think of all synonymous words or expressions in order to find out all the relevant literature on a particular topic. One of the major functons ofthe controlled vocabulary is to control synonyms, that is, to specify as to which of th several synonymous expression is to be used by indexers and searchers and thus to avoi the separation of identical subject matter under different terms in the system. Such control can be achieved by choosing one of the possible alternatives among the terms and designated as preferred term and other synonyms designated as non-preferred terms under which certain user may likely to approach the system. The selection of preferred term among the synonyms under which documents are actually indexed and searched for, must be one under which the majority of system users are likely to look first.

2.4.4.1 Near synonym control : The way treated in synonym control is the way terms are likely to be near synonyms, apart from abbreviations, there are comparatively few words in English that are exactly synonymous.

2.4.4.2 Quasi-synonyms control : The term 'quasi-synonyms' is not very precisely expressive. Mandeoot et al illustrated it in terms of its implications. They

considered quasi-synonyms are antonyms that represent the opposite extremes of values. Let us consider the pair roughness and smootheness; versa. These quasi-synonyms are treated in the same way of synonyms, i.e. one is chosen as preferred term and a reference is made from the other.

2.4.4.3 Differentiation of Homographs : Homographs that is, words with identical spelling but different meaning. Let us consider the term MERCURY; this term is to be used for mythological character, a planet, a metal, a car or any other possible context. This problem can be overcome usually by means of using a parenthetical qualifier or scope note in controlled vocabulary tool.

Another major objective of vocabulary contrll is to link terms together that are semantically related in order to facilitate the conduct of comprehensive search. In an ideal controlled vocabulary tool, it groups related terms that are hierarchically (whole-part; genus-species) related and also across the hierarchies-non-hierarchically (whole-part; genus-species) related and also across the hierarchies-non-hierarchical associative related. In general two kinds of relationships - paradigmatic and syntagmatic - is recognised in contrlled vocabulary tools. A paradigmatic relationship always exists between the terms - Aluminium, Magnesium and Light metals - whereas a syntagmatic relationship is transient, one that is true in certain situations only - Aluminium may be related to Beer barrels but aluminium is not always related to beer barrels and beer barrels are not always related to aluminium.

Two major functions of vocabulary control in an information retrieval environment are: (a) to promote the consistent representation of subject matter by indexers and searchers, thereby avoiding the dispersion of related materials. This is achieved through the control (merging) of synonymous and nearly synonymous expressions and by distinguishing among homographs; and (b) to facilitate the conduct of a comprehensive search on some topic by linking together terms whose meanings are related paradigmatically or syntagmatically.

There are controlled vocabulary tools used to control the vocabulary in indexing and retrieval such as classification schemes, subject heading lists and thesauri. Classification schemes, being tools for organzing knolwedge, could be of great help for vocabulary control. The subject heading lists and theasuri contains alphabetically arranged terms with necessary cross references and notes that can be used for indexing or searching in an information retrieval environment. Subject heading lists were initially developed to prepare entries or heading in a subject catalogue. Therefore, they include rather broader subject terms or headings. On the other hand, theasuri have been developed on specific subject fields with various representations of terms such as synonyms, spelling variants, homonyms, etc, along with indicating the broader (superordinate), narrower (subordinate) and related (coordinate and collateral) terms.

However the distinction between a subject heading lists and theasuri has gradually faded and the latest editions of Library of Congress Subject Headings and Sear's List of Subject Heading indicate the terms' features as shownin normal thesauri.

2.4.5 SUMMARY

Information retrieval system is designed for storage and retrieval of information. The major objective of bibliographic information retrieval system is to retrieve bibliographic details of those items containing users' needed information. Information retrieval system constitutes three major areas such as documents, users' queries and matching with document database. Six components/subsystems are in the information retrieval system: (a) The document selection subsystem; (b) The indexing subsystem; (c) The vocabulary subsystem; (d) The searching subsystem; (e) The user-system inrerface; and (f) The matching subsystem. The process involves in information retrieval system are conceptual analysis and translation in input as well as output levels. There are factors influencing in the performance of information retrieval system in the input and out put levels.

Vocabulary control is one of the important components in the information retrieval system. Lack of controlled voabulary results inconsistency in subject representation. By controlling synonyms, near synonyms, and quasi-synonmys and distinguishing among homographs, the controlled vocabulary avoids dispersion of like subject matter and the collocation of unlike subject matter.

A number of vocabulary control tools have been designed over the years: they differ in their structure and design features, but they all have the same purpose in an information retrieval environment. Availability of vocabulary control helps both the indexer for indexing the documents as well as the end-users in the formulation of their search expressions. A large number of software packages are now available that allow the record creator to automatically switch to one or more chosen online vocabulary control tools in order to select appropriate terms for represening the document in hand. This helps to represent the terms thata are representative of the subject contents of the documents, but these are also standardized (in terms of their usage, spelling, form and so on) and are likely to be chosen by user for searching purposes. Similarly, there are programs available by which the end-users can choose appropriate terms from online vocabulary control tool. Vocabulary control tools also help end-users modify their previosuly formulated search expressions by either widening or narrowing the search expressions.

2.4.6 KEY WORDS

      Antonym : Words opposite in meaning, e.g., Cold Hot

      Homograph : Words with identical spelling but different in meaning.

      Information Retrieval System : Any system that is designed to facilitate the process of searching some collection of documents, using the term document, in order to identify those documents which deal with a particular subject may

be called an information retrieval system.

Synonym : Words with more or less same meaning

2.4.7 REFERENCES AND FURTHER READINGS

Chowdhury, G.G. Introduction to modern information retrieval. London: Library Association Publishing, 1999

Foskett, A.C. The subject approach to information. Ed 5. London: Library Association Publishing, 1996

Frants, V.I., Shapiro, J. and Voiskunskill, V.G. Automated information retrieval: theory and methods. San Diego: Academic Press, 1997

Korfhage, R.R. Information storage and retrieval. New York: John Wiley, 1997

Lancaster, F.W. Information retrieval systems: characteristics, testing and evaluation. Ed 2. New York : John Wiley, 1978

Lancaster, F.W. Vocabulary control for information retrieval. Ed 2. Arlingon, V.A. Information Resources, 1986

Vickey, B. and Vickery A. Information science : theory and practice London: Bowker-Saur, 1987

Vickery, B.C. Techniques of information retrieval. London: Butterworth, 1970

2.4.8 SELF-CHECK EXERCISE

Note : (i) Use the space provided below for your answer.

1. Define information retrieval system.

    ----

2. Enumerate the components/subsystem of aninformation retrieval system.

    ----

3. What do you meant by vocabulary control ?

    ----

4.   Describe the methodology is used contrl over the vocabulary in information system.

LESSON NO. 2.5 **AUTHOR : DR. V. CHANDRA KUMAR**

MEDICAL SUBJECT HEADINGS

LEARNING OBJECTIVES

After studying this lesson, you will be able to

(a) Understand the concept and trend of subject heading lists

(b) Know the structure, scheme and features of the Medical Subject Headings (MeSH)

OUTLINES

2.5.1 Introduction

2.5.2 Medical Subject Headings (Mesh)

2.5.2.1 History of MeSH

2.5.2.2 Features

2.5.2.3 Entry Structure

2.5.2.4 Cross References

2.5.2.5 Entry Display

2.5.2.6 Summary

2.5.3 Summary

2.5.4 Key Words

2.5.5 References and further readings.

2.5.6 Self check exercises.

2.5.1 INTRODUCTION

Controlled vocabulary tools are used to control the vocabulary of indexing and retrieval. These natural language tools contain natural language terms that canbe used for indexing and retrieval purposes. They fall into two major categories: (i) classification schemes; and (ii) subject headings lists and thesauri. Classification schemes, being a tools for organising knowledge in an artificial language, but both subject heading lists and theasuri containalphabetically arranged terms with necessary cross references and notes that can be used for indexing or searching in an information retrieval environment.

Subject heading lists were initially developed to prepare entries/headings in a subject catalogue that could be replicate the classified arrangement of document records.

Subject headings are provided for documents in catalogues with two objectives: (a) to identify a relevant or pertinent document of a given subject; and (b) to enable to find materials on all related subjects with references to any given specific subject. These objectives are sought to be achieved by subject inexing through standard rules and procedure for deriving subject headings for documents.

Subject headings lits having a traditional way of entry display forms such as see, see also, scope note. There is a trend in display semantic relationships in the subject hedings list to conform to the particular standards for theasuri. Subject headings list uses the labels BT, NT, RT, SA and UF for broader terms, narrower terms, related terms, See Also and Used for.

2.5.2 MEDICAL SUBJECT HEADINGS (MeSH)

MeSH is the National Library of Medicine's (NLM) controlled vocabulary thesasurus. The MeSh thesaurus is used by NLM for indexing articles from 4,800 leading biomedical journals for the MEDLINE/PubMED database. It is also used for the NLM-produced database that includes cataloging of books, documents, and audiovisuals acquired by the Library.

2.5.2.1   HISTORY OF MeSH

The first official list of subject headings published by the National Library of Medicine appeared in 1954 under the title Subject Heading Authority List. It was based on the internal authority list that had been used for publication of Current List of Medical Literature which in turn had incorporated headings from the Library's Index-Catalogue and from the 1940 Quarterly Comulative Index Medicus Subject Headings. With the inclusion of Index Medicus in 1960, a new and throughly revised Medical Subject Headings appeard.

With the 1954 Subject Heading Authority List, there appeared a "Categorical Listing" of standard subheadings. In the 1960 Medical Subject Headings, the number of subheadings was reduced to sixty-seven. They could be used under any kind of main heading if the combination was not patently foolish or impossible. These sixty-seven subheadings were applied with more generalized meanings.For instance, the subheading "therapy" was used to mean "therapy of," "therapeutic use of" or just "therapeutic apects." Though this solution was simpler, many problems still remained. Categorized lists of terms were printed for th efirst time in the 1963 Medical Subject Headings and contained thirteen main categories and a total of fifty-eight separate groups in subcategories and main ategories. These categorized lists made it possible for the user to find many more related terms than were in the former cross-reference structure. In 1963, the second edition of Medical Subject Headings contained 5,700 descriptors, compared with 4,400 in the 1960 edition. Of the headings used in the 1960 list, 113 were withdrawn in favor of newer terms. In contrast, the 2006 edition of MeSH contains 23,885 descriptors.

The Medical Literature Analysis and Retrieval System (MEDLARS) would speed the publication process for bibliographies such as Index Medicus, facilitate the expansion of coverage of the literature,a nd permit searches for individuals upon demand. The new list of subject headings introduced in 1960 was the underpinning of the anlaysis and retrieval operation. MeSH was a new and thoroughly revised version of lists of subject headings complied by NLM for its bibliographyies and cataloging. Frank B. Rogers, then NLM director, announced several innovations as he introduced MeSH in 1960.

There is another departure from traditional practice represented in this list. This is the adoption of standard topical subheadings for cataloging books, as well as for indexing periodical articles. The topical subheading is in effect a substitute for a phrase heading, and on the whole it is a preferable substitute. The main heading-topical subheading combination is a pre-coordination of terms, reducing the problem of term permutation, which looms large in most manual retrieval systems in book form.

There are 22,997 descriptors in MeSH. In addition to these headings, there are more than 151,000 headings called Supplementary Concept Records (formerly Supplementary Chemical Records) within a separate theasurus. There are also thousands of cross-references that assist in finding the most appropriate MeSH Heading, for example, Vitamin C see Ascorbic Acid. These additional entries include 24,050 printed see references and 112,012 other entry points.

The Medical Subject Headings Section continually revises and updates the MeSH vocabulary. In additon to receiving suggestions from indexers and others, the staffs collect new terms as they appear in the scientific literature or in emerging areas of research; define thee trms wihtin the context of existing vocabulary; and recommend their addition to MeSH. Professionals in various disciplines are also consulted regarding broad organizational changes and close coordination is maintained with various specialized vocabularies.

2.5.2.2   FEATURES

(i)   MeSH contains several different types of terms.

(ii)   Descriptors (main headings): characterize the subject matter or content.

(iii)   Qualifiers: are used with descriptors and afford a means of grouping together those documents concerned with a particular aspect of a subject. A list of qualifiers appears following the list of new descriptors.

(iv)   MeSH also has special types of headings available for indexing, cataloging, and online searching:

(v)   Publication Types: characterize what the item is, i.e., its genre, rather than what it is about.

(vi)   Geographics: include continents, regions, countries, states, and other

geographic subdivisions. They are listed in category Z of the tree strutctures.

(vii)   Entry terms or see references: synonyms or closely related terms that are cross-references to descriptors.

2.5.2.3   ENTRY STRUCTURE

MeSH descriptors are organized in 16 categories: category A for anatomic terms, category B for organisms, C for diseases, D for drugs and chemicals, etc. Each category is further divided into subcategories. Within each subcategory, descriptors are arrayed hierarchically from most general to most specific in up to eleven hierarchical levels. Entries are at the most general level of the hierarchical structure in very broad headings such as "Anatomy" or "Mental Disorders." More specific headings are found at more narrow levels of the eleven-level hierarchy, such as "Ankle" and "Conduct Disorder." The user may consult the trees to find additional subject headings which are more specific than a given heading, and broader headings as well. For example, under Abnormalities, there are specific abnormalities :

> Abnormalities C16.131
> Abnormalities, Drug Induced C16.131.42
> Alagille Syndrome C16.131.77.65
> Angelman Syndrome C16.131.77.95

In the trees, each drscriptor is followed by the number that indicates its tree location. The trees also contain geographic descriptors (Category Z), which also do not appear in the Alphabetic List. These descriptors may be used in searching MEDLINE.

Complex concepts can be described in three different ways:

(i)   The first is coordination, the combined use of two or more separate descriptors. For exmaple, jejual enteritis may be expressed by the use of Jejunum and Enteritis.

(ii)   Secondly, qualifiers can be used in conjunction with appropriate descriptors. A deficiency of monoamine oxidase may be indexed as Monoamine Oxidase/deficiency. The direct linkage of the qualifier to the dscriptor to which it relates avoids the possibility of false coordination that may occur if two descriptors are used to represent a single concept.

(iii)   Thirdly, many pre-cordinated descriptors are contained in MeSH for frequently encountered subjects. If MeSH has a pre-coordinated descriptor such as Heart Surgery, the indexer or cataloger uses it rather than a descriptor-qualifier combination. If a descriptor-qualifier combination is availabe, it will be used in preference to coordinating two descriptors.

Form of heading : There is a preference for headings using the direct rather than the inverted form; thus Mitral Valve, rather than Valve, Mitral. In a number of

instances, however, inverted forms are used.

Spelling : American spelling is preferred rather than British spelling, thus Anesthesia rather than Anesthesia.

Englishing: There are some terms in the foreign literature which defy translation into English; they designate concepts which are unknown in the English literature. In some cases we have found no solution except to anglicize the foreign trm. Thus radiesthesia, and reflexotherapy, describe concepts appearing in the French and German literature, and in terms similar to the original French and German.

Choice of terms: Entry terms are used in the form and with the definitions set forth in standard reference works. The names of bacterial genera are from the latest edition of Bergey's Manual, the names of neoplasms are the terms preferred in the American Cancer Society's Manual of Tumor Nomenclature; the names of enzyme groups are those promulgated by a Committee of the National Research Council; the names of drugs are the generic terms proposed by the American Medical Association's Council on Drugs: and so forth. Exceptions do occur; excepting oversights, or those cases in which it has been necessary to accept a term prior to its normalization in a standard reference work, exceptions have been made only after due deliberation.

2.5.2.4   CROSS REFERENCES

Many synonyms, near-synonyms, and closely related concepts are printed in MeSH as see references to help users find the most relevant MeSH descriptor for the concept they are seeking. Three kinds of informative refeences suggest other descriptors in MeSH that relate to the subject and that may be useful in indexing, cataloging, or searching a particular topic.

The entry terms, sometimes called "See cross-references", indicate that information related to one term will be found under a different term. They are not always synonyms of the descriptors to which they refer. Some terms, because of their narrow focus, are not useful as subject headings; they are instead created as cross-references.

Abdominal Abscess
  X Abscess, Abdominal
  X Abscess, Intra-Abdominal
Abscess, Abdominal see Abdominal Abscess

See related references are used primarily to indicate related headings not occurring in the same subcategory of the MeSH Tree Structures. See related referrences indicate the presence of other descriptors related to the topic conceptually, e.g., NAVAL MEDICINE see related DIVING.

  Indicators of Entry and Related Terminology

Referring to another term	Corresponding "referred from" designation

| see | X |
| see related | XR |

Consider also is used primarily with anatomical descriptors. It indicates the presence of other descriptors that relate to the topic linguistically, e.g., BRAIn consider also terms at CERBER- and ENCEPHAL-. Note consider also refers to groups of descriptors beginning with a common stem rather than to a single descriptor.

The descriptor/qualifier combination notations refer an invalid (and prevented) descriptor/qualifier combination to the preferred pre-coordinated descriptor expressing the equivalent concept. For example, at the entries for the descriptors ACCIDENTS, AORTA, and ARM, the references read:

ACCIDENTS/prevention & contrll <u>see</u> ACCIDENT PREVENTION

AORTA/radiography <u>see</u> AORTOGRAPHY

ARM/injuries <u>see</u> ARM INJURIES

2.5.2.5   ENTRY DISPLAY

MeSH descriptors are arranged in both an alphabetic and a hierarchical structure. In the alphabetic listing, each descriptor is followed by one or more alphanumeric expresions (tree numbers) that indicate where in the Tree Structures the term is to be found. A plus sign following a tree number in the Alphabetic List indicates there are more specific descriptors in the Tree Structures at that number under that descriptor. A sample MeSH descriptor data and tree of an entry is shown below:

MeSH Descriptor Data

MeSH Heading	Acquired Immunodeficiency Syndrome
Tree Number	C02.782.815.616.400.040
Tree Number	C02.800.801.400.040
Tree Number	C02.839.040
Tree Number	C020.480.040
Annotation	caused by HIV; coord IM with HIV-1 or HIV-2 (IM) if pertinent; / epidemiol: consider also HIV SEROPREVALENCE; AIDS-RELATED OPPORTUNISTIC INFECTIONS is available if particularly discussed: see note there; for lymphoma with AIDS, use LYMPHOMA, AIDS-RELATED
Scopr Note	An acquired defect of cellular immunity associated with infection by the human immunodeficiency virus (HIV), a CD4-positive T-lymphocyte count under 200 cells/microliter or less than 14% of total lymphocytes, and increased susceptibility to opportunistic infections and malignant neoplasms. Clinical manifestations also include emaciation (wasting) and dementia. These

elements reflect criteria for AIDS as defined by the CDC in 1993.

Entry Term	AIDS
Entry Term	Imunodeficiency Syndrome, Acquired
Entry Term	Imunolgic Deficiency Syndrome, Acquired
Entry Term	Acquired Immune Deficiency Syndrome
Entry Term	Acquired Immuno-Deficiency Syndrome
See Also	AIDS Arteritis, Central Nervous System
See Also	AIDS Dementia Complex
See Also	AIDS Serodiagnosis
See Also	HIV Seropositivity
See Also	HIV Seroprevalence
See Also	Lymphorma, AIDS-Related.
Allowable	BL CF CI CL CN CO DH DI DT EC EH EM EN EP ET GE HI
Qualifiers	IM ME MI MO NU PA PC PP PS PX RA RH RI RT SU TH TM UR VS VE VI
Entry Version	ACQUIRED IMMUNODEFIC SYNDROME
Previous Indexing	Immunologic Deficiency Syndromes (1979-1982)
History Note	83
Unique ID	D000163

MeSH Tree Structures

Virus Diseases [C02]

   RNA Virus Infections [C02.782]

      Retroviridae Infections [C02.782.815]

         Lentivirus Infections [C02.782.815.616]

            HIV Infections [C02.782.815.616.400]

                        Acquired Immunodeficiency Syndrome [C02.782.815.616.400.040]

                        AIDS Artritis, Central Nervous System [C02.782.815.616.400.048]

                        AIDS-Associated Nephropathy [C02.782.815.616.400.050]

                        AIDS Dementia Complex [C02.782.815.616.400.700]

                        AIDS-Related Complex [C02.782.815.616.400.080]

                        AIDS-Related Opportunistic Infections [C02.782.815.616.400.100]

           HIV-Associated Lipodystrophy Syndrome [C02.782.815.616.400.400]

HIV Enteropathy [C02.782.815.616.400.480]

HIV Seropositivity [C02.782.815.616.400.500]

HIV Wasting Syndrome [C02.782.815.616.400.520]

2.5.3 SUMMARY

Classification schemes, subject headings list and thesauri are the controlled vocabulary tools used for indexing and retrieval purposes. Subject headings are used with two objectives: to identify the relevant documents on given subjects and to find out the related documents. Though the structure and purpose of subject headings lists differed with theasuri, the structure of the subject heading lists is similar with thesauri.

MeSH was develped for indexing the medical literature in MEDLINE/PubMED database. MeSH has its own intrinsic features that being used for comprehensive literature in medicine. It has principles to inclusion of entry term in the subject headings list. Direct as well as indirect entries can be found in the list. Complex subject have some kind of pre-coordination in the entry heading. For synonyms and equivalent concepts are linked with cross-references. Three kinds of references are recognized by MeSH. The entries are displayed in alphabetical and tree structures.

2.5.4 KEYWORDS

Complex subject: A subject with a basic class and one or more facet ideas.

Entry term: A term selected to be a subject heading.

Indexing language : It is a set of terms and devices for handling the relationships between them in a system.

2.5.5 REFERENCES AND FURTHER READINGS

Chain, Loiis Mai. Cataloguing and Classifiction: an introduction. 2nd ed. New York, McGraw-Hill, 1994

Chowdhury, G.G. Introduction to modern information retrieval. London: Library Association Publishing, 1999

Foskett, A.C. The subject approach to information. Ed 5. London : Library Association Publishing, 1996 http://www.nlm.nih.gov/mesh

Lancaster, F.W. Information retrieval systems: characteristics, testing and evaluation. Ed 2. New York: John Wiley, 1978

Lancaster, F.W. Vocabulary control for information retrieval. Ed 2. Arlington, V.A. Information Resources, 1987.

Taylor, Ariene G. Wynar's Introduction to Cataloguing and Classification. 9th ed. Englewood, Colo, Libraries Unlimited, 2000.

2.5.6 SELF-CHECK EXERCISE

Note: (i) Use the space provided below for your answer.

1.   What do you meant by subject headings list ?

----

2.   Describe the rules underlining in the choice of terms, form of heading, spelling
      in the MeSH.

----

3.   List out the types and functions of 'cross references' in the MeSH.

LESSON NO. 2.6   **AUTHOR : DR. V. CHANDRA KUMAR**

THESAURUS OF ERIC DESCRIPTORS AND THESAUROFACET

LEARNING OBJECTIVES

           After studying this lesson, you will be able to

(a)   Understand the concept of thesaurus in general and Thesaurus of ERIC Descriptors in particular.

(b)   Understand the concept of Theasurofacet.

OUTLINES

2.6.1 Introduction

2.6.2 Thesaurs of ERIC Descriptors in Online form.

       2.6.2.1 Structure

       2.6.2.1 Thesaurus of ERIC Descriptors in Online form.

2.6.3. Thesaurofacet

2.6.4. Summary

2.6.5. Key Words

2.6.6. References and further readings

2.6.7. Self-check

2.6.1 INTRODUCTION

UTISIST defines thesaurus in terms of its function, a terminological control device used in translations from the natural language of documents, indexers or users into a more constrained 'system language' (documentation language, information language). In terms of structure, a thesaurus is a controlled and dynamic vocabulary of semantically related terms which covers a specific domain of knowledge.

A thesaurus which acts efficiently as a controlled language for indexing and retrieval therefore has to provide specific terms, general terms to which the specific terms are related in the subject literature. Thesaurus is, in fact, an improved version of the 'list of subject headings.'

Recognition of the thesaurus as a widely used form of indexing tool came with the first international standard for the construction of monolingual thesauri in 1974.

Since then the processes of developing and maintaining thesauri have been standardized. There are international standards like (ISO 2788 and ISO 5964), British (BS 5723 and BS 6723) and UNISIST (called UNISIST Guidelines).

2.6.2 THESAURUS OF ERIC DESCRIPTORS

The Education Resources Information Center (ERIC) sponsored by the Institute of Education Sciences (IES) of the U.S. Department of Education produces the world's premier database of journal and non-journal education literature. The ERIC database is the world's largest single source of education information containing more than 1.1 million bibliographic records of education-related journal articles and non-journal materials. ERIC provides ready access to education literature to augment American education by increasing and facilitating the use of educational research and information to improve practice in learning, teaching, educational decision-maing, and research.

The ERIC Thesaurus, or Thesaurus of ERIC Descriptors, is a list-controlled vocabulary of education-related terms used by ERIC to organize database materials by subject. A popular and widely used reference tool for sets of education-related terms established and updated by ERIC lexicographers to assist searchers in defining, narrowing, and broadening their search strategies.

The theasurus content is the result of 30 years of work and refinement by the **ERIC system**. The Theasurus of ERIC Descriptors, cited by the American National Standards Institute as an exemplary model in its 1993 edition of the z39.19 Standard (Guidelines for the Construction, Format, and Management ofMonoloingual Thesauri), offers scope notes, or definitions, articulated by the U.S. Department of Education's Educational Resources Information Center (ERIC). In addition to offering definitions, arrays of related terms are offered to help the searcher to pinpoint the desired concept as presented in ERIC Theasurus.

The 14th printed edition contains a total of 10,773 vocabulary terms with 206 descriptors and 210 "use" references that are new to this edition

2.6.2.1   STRUCTURE

Two kinds of descriptors are recognized in ERIC: major and minor. "Major descriptors" have an asterisk ("*") in front of them to indicate they are the primary topics of the document. "Minor descriptors" are listed without an asterisk to indicate that they are secondary topics. The structure of theasurus is similar with traditional structure. This includes a Scope Note, Use For (UF) & Use (USE) references, Narrower Terms (NT), Broader Terms (BT), & Related Terms (RT).

Scope Note: In the Scope Note Special indexing notes are often included in form of a brief statement. It is used to clarify an ambiguous term or to restrict the usage of a term.

"UF" reference: The "UF" reference is employed generally to solve problems of

synonymy occuring in natural language. Terms following the UF notation are not used in indexing. They most often represent either (1) synonymous or variant forms of the main term, or (2) specific terms that, for purposes of storage and retrieval, are indexed under a more general term. The examples below illustrate both types of UFs:

LIFELONG LEARNING

Use For

Continuous Learning (1967, 1980)

Education Permanent

Lifelong Education

Life Span Education

Permanent Education

Recurrent Education

A former Descriptor that has been downgraded to the status of a UF term is accompanied by a "life span" notation in parentheses: e.g., (1966 1980). This indicates the time period during which the term was used in indexing. It provides useful information for searching older printed indexes, or computer files that have not been updated.

"USE" reference: The USE referenc, the mandatory reciprocal of the UF, refers an indexer or searcher from varient terms (non-preferred terms) to the preferred terms. A coordinate or multiple USE reference is also used in this thesaurus. The following example illustrates the use of two main terms together to represent a single concept, both for indexing and searching:

FOLK DRAMA (1969 1980)

USE DRAMA AND FOLK CULTURE

BT (Broader Term) and NT (Narrower Term) : These indicate the existence of a hierarchical relationship between a class and its subclasses. Narrower terms are included in the broader class represented by the main entry. The [+] symbol besides a term indicates that there are further narrower terms.

LIBRARIES

Narrower Terms

Academic Libraries [+]

Branch Libraries

Children Libraries

Depository Libraries

Public Libraries [+]

Research Libraries

School Libraries

Special Libraries [+]

The Broader Term (BT) is the mandatory reciprocal of the NT. Sometimes a term may

have more than one Broader Term :

REMEDIAL READING

Broader Terms

Reading [+]

Reading Instruction [+]

Instruction [+]

RT (Related Term) : Part-whole relationships, near-synonyms, and other conceptually related terms appear as RTs.

HIGH SCHOOL SENIORS

Related Terms

College Bound Students

Grade 12

High School Freshmen

High School Graduates

Noncollege Bound Students

To resolve the problem of homograhs, ERIC also used a Parenthetical Qualifier. Examples Includes LETTERS (ALPHABET) and LETTERS (CORRESPONDENCE); SELF EVALUATION (INDIVIDUALS) and SELF EVALUATION (GROUPS).

Theasurus of ERIC Descriptors display in the following ways of presentation:

(i)   Alphabetical Descritor Display

(ii)   Rotated Descriptor Display

(iii)   Two-Way Hierarchical Term Display

(iv)   Descriptor Groups and Descriptor Groups Display.

2.6.2.2   Thesaurus of Eric Descriptor in Online form

The Eric database can access the database on this Web site or through on this Web site or through several commercial vendors and public networks Researchers have come to rely on this thesaurus to locate precise terms from the controlled vocabulary used to index the ERIC database.

ERIC database can be searched using thesaurus either alphabetically or by category wise browsing the descriptor list.

If you have more than one descriptor, one can use Boolean operators to expand or narrow down the search. All searching is based on Boolena logic : the search engine retrieves document records based on the way you tell it to combine subject terms with AND, OR and NOT operators.

2.6.3 THESAUROFACET

This is a faceted classification and a thesaurus designed for engineering and related subjects, which was developed for use in the English Electric Company. It is a specialized kind of retrieval language with both a thesaurus, type and a classification type display, each containing some term unique to itself. These two parts are

complement to one another. The faceted component takes care of hierarchical relationships, all other relationship appear in the thesaurus component. One of the advantages of the thesaurofacet is that, it can be used for arranging books on the shelves of a special library as well as for indexing the items in a database.

In this scheme, related terms indicated by the classification hierarchy are not repeated in the index to the scheme. The index includes additional related terms only. A brief extract from the schedules is given below :

SV   Sewage engineering

SV2   Sewage

SV4   Domestic sewage

SV6   Industrial sewage

SV7   Soil sewage

SV9   Surface water sewage

SVB   Sewers

SVD   Outfall sewers

SVD   Strom sewers

SVE   Manholes

SVG   Sewage pumping

SVH   Syphone (Sewage)

SVS   Sewage disposal)

The entry for 'Sewage' from the index/thesaurus is :

Sewage   SV2

UF   Sewerage

RT   Corrosion atmospheres

BT (A) Wastes

It is noted that where term 'Sewage' is concerned, the broader term 'Sewage engineering', the narrower terms 'Domestic sewage', 'Industrial sewage', and so on, and the related terms 'Sewer', 'Sewage pumping', and so on are indicated in the classification schedules and are therefore not repeated in the index. thesaurus. The thesaurus entry does however contain related terms which are not apparent from the schedules and there is the broader term 'Wastes' additional to the broader term 'Sewage engineering' revealed by the classification hierarchy. (The 'A' in BT(A) stands for 'Additional').

2.6.4 SUMMARY

The saurus is one of the controlled vocabulary devices like classification schemes, subject headings list. There are many standards and guidelines evolved for establishment and maintenance of thesaurus. The standards are at international and national levels. Many discipline-oriented thesauri are available for use. Thesaurus of ERIC descriptors is from ERIC system used for indexing and retrieving education related subjects.

The Thesaurofacet is a type of controlled vocabularly that combines the alphabetical thesaurus with a faceted hierarchical classification, including notation. The Nesco's thesaurus can be considered a type of thesaurofacet. In this thesaurus, however, the alphabetical and faceted components are not completely complementary to each other.

2.6.5 KEYWORDS

Preferred term : A term used consistently when indexing to represent a given concept; some times known as descriptor.

Non-preferred term : The synonym or quasi-synonym of a preferred term.

2.6.6 REFERENCES AND FURTHER READINGS

Chowdhury, G. G. Introduction to modern information retrieval. London : Library Association Publishing, 1999.

Houston, James E. (ed.). Thesaurus of ERIC Descriptors. 14th Edition. New York, Greenwood Publishing Corp;, 2001

http://eric/ed.gov/

Hunter, Eric, J. Classification made simple, Alderslot, Gower, 1988.

Lancaster, F.W. Information retrieval systems : characteristics, testing and evaluation. Ed 2. New York : John Wiley, 1978.

Lancaster, F. W. Thesaurus construction and use : a condensed course. (PGI-85/WS/11). Paris, Unesco, 1985.

2.6.7 SELF-CHECK EXERCISE

Note : (i) Use the space provided below for your answer.

1.   What are the guiding principles followed in the Thesaurus of ERIC Descriptors ?

----

2.   Describe the structure of thesaurus entry of ERIC descriptors ?

----

3.   What do you meant by Thesaurofacet ?

MASTER OF LIBRARY AND INFORMATION SCIENCE   MLIS 202
INFORMATION STORAGE AND RETRIEVAL

LESSON NO. 2.7   AUTHOR : MOHINDER KAUR

SEAR'S LIST OF SUBJECT HEADINGS

Structure

2.7.0   Objectives

2.7.1   Introduction

2.7.2   What is Subject Heading ?

2.7.3   Sears List of Subject Heading

      2.7.3.1   Purpose

      2.7.3.2   Origin

      2.7.3.3   History of Different Editions

      2.7.3.4   How to determine the Subject work ?

      2.7.3.5   Specific and Direct Entry

      2.7.3.6   Common Usage

      2.7.3.7   Form Headings

      2.7.3.8   Grammar of Subject-Headings

      2.7.3.9   Compound Subject-Heading

      2.7.3.10   Methods of Sub-Divisions

      2.7.3.11   Subdivision by Physical Form

      2.7.3.12   Subdivisions by showing Special Aspects

      2.7.3.13   Sub-Divisions through Chronology

      2.7.3.14   Sub-Divisions with the Name of Place

      2.7.3.15   Subject Heading of Non-book Materials

      2.7.3.16   Subject Headings of New Subjects

2.7.4   Self-Check Exercises

2.7.5   Reference Books

2.7.0   Objectives :

After studying this lesson you would be able to understand :

      (a)   Concept and need of Subject Heading; and

      (b)   Theoretical base of Sears List of Subject-Headings.

2.7.1   Introduction :

Library acquires documents in any form to serve users. Library Catalogue is supposed to be up-to-date to provide key to the holding of the library. The library catalogue

provides access to the collection by providing a multiple-sequence of entries. The library catalogue should be an instrument equipped to deal with the several channels of enquiry of its users applied for successful identification and retrieval of books and documents, no matter what pattern of arrangement is applied to the material on the shelves.

The catalogue is a finding list which helps to identify, to trace and to locate a document of other reading material in a library. According to C. A. Cutter the purpose of a catalogue is to enable a person to find a book of which either (i) "the author is known, or (ii) the title is known and (iii) the subject is know." Here in this lesson or topic, key importance is going to be given to the 'Subject-approach'. AACR-II does not provide any rules or guidelines for the preparation of subject entries which is the top most demand of users of a special library. To meet the subject approach of information, every library has to provide some sort of alphabetical subject-catalogue. Since the subjects-approach to books is mostly through the words of the natural language representing the subject content of reading material. A number of devices have so far been used by the library professionals to provide for subject demand to the reading materials. For this purpose there was demand for a list of subject headings so that a particular word be made in use by different cataloguers to minimise confusion. At present, there are two most popular lists of subject-heading which are :

      (1)   The Library of Congress List of Subject-Heading (LCLSH).

      (2)   The Sear's List of Subject-Headings.

Here is this lesson, we will study about one list of subject-heading i.e. 'Sears List of Subject-Headings'.

2.7.2 What is a Subject Heading ? :

Subject-heading is a specific and commonly or universally accepted term for the identification of a particular subject or topic. It is very difficult for the users to find material on a particular topic with author or title approach. In one tries to locate material on a particular topic by knowing only the name of the author, one has to trace all the books written by that author. It would take a lot of time and energy and sometimes may find nothing. If one knows the subject-heading or subject of the topic, can find material in a quick manner. Here user demands specific material and time is very important for him or her. Thus subject-headings same the time of readers, and it satisfy the very important Law of Science i.e. save the time of a reader.

2.7.3 Sears List of Subject Heading :

Now we will study the purpose, origin, history, salient features, pattern and how to use this list of subject-headings.

2.7.3.1   **Purpose :** The main purpose of subject cataloguing is to prepare a list under one uniform word or phrase all of the material on a given topic that a library has in its collection. A subject may be a topic treated in a book, film strip or other

kind of document. A subject-heading is the word or phrase used in the library catalogue to show its topic. A subject is arranged above the main or author entry as an access point regardless of the format of the catalogue of library record.

Subject entries are made of the library materials in the catalogue in order to show, what information or material the library has on a subject which a reader demands. These entries are just like author entries to show the works that the library has by a given author. Intelligently made, subject entry is a very important work to supplement to reference tools in a library because it may help the reader as well as the reference librarian to find quickly and rightly the material needed to answer a particular questions on a subject. Subject entries are a fastest way to find material on a particular topic or subject. Ordinarily one consults the author entry for a specific work, but here if the reader is not sure about the author's name, one may find the individual piece of work more rapidly under the subject-entry. For example, if a reader tries to find a book 'Documentation' by S. R. Ranganathan, would be difficult to consult all the cards in the catalogue made under Ranganathan but there can be to consult all the cares in the catalogue made under Ranganathan but there can be a quick finding if one consults under the subject-heading 'Documentation'. Other benefit here would be that reader can find many other books on this topic written by different authors along with the particular books which one demands.

2.7.3.2   Origin of Sears List of Subject-Heading : Minnie Earl Sears prepared the first edition f the list of subject-headings in response to demands suitable to the needs of the small library. Before her publication lists of subject-headings were already available - one list of subject-headings authored by A.L.A. and the other by Library of Congress. Sears list of subject headings was published in 1923 under the name 'List of Subject-Heading for Small Libraries'. These headings were used by nine small libraries that were known to be well catalogued. Minnie Earl Sears felt the need for uniformity in the chosen subject-headings and thus the followed the pattern of the Library of Congress subject-headings with some exceptions. She followed this (LCLSH) keeping in mind to extend the list of subject-headings on the same pattern as and when the collection of a library increases. The Library Congress List of Subject-Headings is meant for large and specialized collection and is one of the most comprehensive lists every prepared. On the other hand, Sears List of Subject-Headings is meant for small and medium sized libraries. LCLSH was compiled in 1909-1914 and Sears List on this list come in publication in 1923. For the purpose of this work i.e. subject-entries, the Sears List has been selected because of its worldwide popularity and suitability for small and medium sized libraries.

2.7.3.3   History of its different editions : The first edition of the Sears List was published in 1923 under the title "List of Subject-Headings for Small Libraries". It was prepared by Minnie Earl Sears in demands and requests for a list of Subject-

headings which could be less comprehensive in comparison to those of issued by the American Library Association and the Library of Congress. This list was based on the "Library of Congress List" so that there would be no confusion for librarians those were using Library of Congress headings or that might wish to add terms from that list. A number of editions were brought out with the 18th edition as the latest but for teaching purposes its 12th edition is in use.

Minnie Earl Sears used only 'See' and 'refer from' references in the first edition. It 2nd edition was published in 1926. Here she added 'See also' reference at the request of teachers of cataloguing who were using the List as a text book. To make the List more useful as a textbook, she added a chapter on "Practical Suggestions for the Beginner in Subject-Heading Work" in here third edition of the list.

Fourth and Fifth editions of this book were edited by Isabel Stevenson Montro respectively in 1939 and 1944. New thing added in this edition was the inclusion of Dewey Decimal Classification numbers. These numbers were given after the proper subject-headings. Some new subject-headings wer also included in this edition which were based on Standard Catalogue Series and from the catalogue cards issued by the W.H. Wilson Company. Another new feature was the printing in *Italics* of those subdivisions that had a more general application.

The sixth edition in 1950, 7th edition in 1954 and 8th edition of Sears List published under the editorship of 'Bertha M. Frick'. Now the title of the book came under the name of new title. i.e. Sears List of Subject-Headings. This title was changed with the publication of its Sixth edition. Now this list was not in use. Only in small libraries rather medium-sized libraries also started to have subject-headings from this book, thus the last 'phrase for small libraries' was deleted. With the publication of its 9th edition, the Dewey Decimal Classification (DDC) numbers disappeared. libraries which were usingh the list of subject-headings complained that they had to verify the class numbers from DDC, as there was instruction to consult DDC also for the given classification numbers. Keeping in mind, the given suggestions, DDC numbers were omitted by the publisher in the 9th and 10th editions. This practice was followed for at least twelve years. Now the larger libraries started following network and computerized cataloguing which was too costly and beyond the reach of small libraries. This was also not the needs of small and medium sized libraries. With limited budget and less staff members, these libraries and librarian of these libraries were left with no assistance in the classification of their collections. Earlier editions with classification numbers alongwith subject-headings were helpful to them in the work of classification. With the help of these class numbers, they could verify the class numbers which they allotted to their documents. Largely in demand and needs of these librarians, Dewey Decimal Classification numbers again appeared in the eleventh edition of Sears List of Subject-Headings. From that time continuously these

class numbers are assigned by the editor of the Standard Catalogue, Department of the H. W. Wilson Company. These classification number are taken <u>from the eleventh edition of the Abridged Dewey Decimal Classification and Relative Index published in 1979 by Forest Press.</u> Here only broad classification numbers are given. Sometimes, same class numbers are given to more than one point of view and the result is that one number is consequently for the subject-headings. For example, of t he subject-heading "Food" there are given three class number i.e. 641; 641.3 and 664. Now which is more appropriate depends on the selection of a cataloguer. Few of the numbers selected from 12th edition of "Abridged Dewey" carried out more than four places beyond the decimal point. Detailed numbers are only the requirement of large libraries, thus these numbers are not satisfying the needs of small libraries.

The headings added in the 12th edition were suggested by librarians working in various sizes and types of libraries and by the cataloguers at the H. W. Wilson Company. Some headings were taken from the latest edition of Library of Congress Subject-Headings.

The <u>thirteenth</u> edition appeared in 1986 which was prepared by Carmen Rovira and Caroline Reyes and first time took the advantage of computer. It also responded to the changing theory in subject analysis which was in demand due to development of online public access catalogue. This effort was taken further in the <u>14th edition published in 1991.</u> This was published under the editorship of Martha T. Mooney. He reduced the number of compound terms and simplified many subdivisions. In this edition, 'Subject-Headings for Children's Literature' included.

In the <u>15th edition published in 1994</u>, the first edited by Joseph Miller, the interval between publication of editions was shortened to provide a more timely updating of subject-headings. In this edition a List of Concealed and Replacement-Headings was added to facilitate the updating of catalogues, and the legend "Former heading" was introduced within the List to identity earlier forms of headings. 15th edition also included many headings relating to individual works of fiction, poetry, drama and other imaginative works such as films and radio and television programs, based on the 'Guidelines on Subject' Access to Individual Works of Fiction, Drama, etc'.

Its <u>16th edition published in 1997</u> included classification numbers from the 13th edition of Abridged Dewey Decimal Classification. Further it included instructions for the applications of subdivisions, and headings in the field of religion so that undue detail given to Christain be reduce and made it so to be useful for other religions.

Important features of <u>seventeenth edition published in 2000</u> was the revision of the headings for the native people of the Western Hemisphere. The headings Indians, Indians, of North America, Indians of Mexico, etc. were cancelled in favor of Native Americans, which may be subdivided geographically by continent, region, country, state of city. The heading Native Americans is now the pattern heading for all ethnic

groups.

In the year 2000, Sears List of Subject-Headings has completed its 80 years. During these years, the new and revised heading in each edition have reflected developments in the literature and in the use of the English language. The changes in the form of the headings and in the structure have reflected the prevailing philosophy of subject cataloguing. Now we have its <u>18th edition</u> in the market.

(1)   There are three major changes in the new edition of Sears List. The first is the inclusion of five hundred new subject-headings.

(2)   The second is the revision of the classification numbers. In this edition numbers are taken from the 14th edition of the 'Abridged Dewey Decimal Classification' published in 2004.

(3)   The third major improvement is in the Principles of the Sears List. The Principles of the Sears List, is intended both as a statement of the theoretical foundations of the Sears List and also as a concise introduction to subject-cataloguing in general. It has been expanded in this edition to provide guidance to libraries that select to allocate topical and geographic headings to individual work of fiction, drams and poetry. This difficult area of cataloguing has been much discussed in recent years in the library community and in the committees and subcommittees of the American Library Association. The List of Commonly Used Sub-Divisions, which was selected in the previous editions of the Sears List in favour of a more exhaustive treatment of subdivisions within the body of the List, has been restored in this editions and renamed List of Sub-Divisions provided for in the Sears List. It now list, for the purpose of easy reference, every subdivision for which there is a provision in Sears, no matter have specialized. At the same time, for every subdivision there is an entry in the alphabetical list full instructions for the use of that partial subdivision.

2.7.3.4   How to determine of the Subject work ? : For the selection of Subject Heading, the fist step is to ascertain the real subject of the document and the purpose for which it is written. Sometimes, there is clearly determined, e.g. 'Flowers' is obviously the subject of the book titled 'About Flowers'. In other cases, the subject is not so easy to ascertain because it may be a complex one or the author may not express it in a manner clear to someone who is unfamiliar with the subject. The subject of a work cannot always be determined from the title alone. The title information is often misleading and undue dependence on it can result in error. A book entitled 'Rainbow' immediately suggests the subject part of 'Physics' but closer examination reveal the book to be a novel written by D. H. Lawrence. There is no subject-heading for Rainbow but under 'Spectrum analysis' ther is class reference entry under the work 'Spectrum' (x). The actual subject-heading of this book will be "English Fiction". Sometimes one has to read the contents, introduction, other time to scan the whole document. Other reference tools may also help in the selection of

subject-heading of a document. For non-book material one has to examine the container, the label, any accompanying guides, etc. and view or listen to the contents. The cataloguer will be in a position to determine the subject of the times in hand only after the preliminary examination has been made. If the meaning of a subject is not clearly understood, one should consult reference source, not only dictionaries and encyclopedias, but also specialized reference books, as well. Only when the cataloguer has decided on the subject-content of the work and identified it with explicit words, can the Sears List be used to advantage. The Library catalogue will be more useful if the cataloguer consider material from the reader's point of view. Finding of subject-heading is just like classifying a book. The classification numbers give opposite to subject-headings can also help in the right finding of subject-headings.

2.7.3.5   Specific and Direct Entry : Principle of specific entry is fundamental for both in using and making a modern subject catalogue. The rule of specific and direct entry is to enter a work directly under the most specific term, i.e., subject-heading, which accurately and precisely represents its content. This word serves as a key abstract of the work. It a reader wants information about 'Resticides', The direct approach, is to consult the catalogue under the heading 'Pesticides', not under the large topic Agriculture or Farming. However, in practice the list very often provides broader headings, so as to meet the demands of small libraries, thus sacrificing the principle of specific entry.

2.7.3.6   Common Usage : Sears List of subject-heading has used the words which are more common in America. For example, color, cataloguing, labor, elevator in place of lift. But for the clarification and help to readers cross reference entries are given in British Spellings i.e. color, cataloguing, labour, lift etc. Sometimes some subject-headings are give in plural form. This all information in given in 'Principles of the Sears List'.

2.7.3.7   Form Headings : In Sears List of Subject-Headings, there are headings of another type, generally known as form headings or form subject-heading those have the same appearance as regular subject headings but refer only to the Literature and forms of Art, e.g. Essays, Poetry, Drama, Fiction, Hymns, Songs etc. Literary form headings are usually used for collected works rather than works of individual writers. For example, the form heading Essays in used not for works of an individual author but for the collection of essays by authors of different countries. Works of individual authors are entered the name of authors and a reader consults the author or title entry in the catalogue for such works.

For a work about essay as a literary form, e.g. the appreciation of the essay or how to write it, the heading Essay represents a true subject and not a form heading. The distinction between form headings and subject-heading can sometimes be made by using the singular form for the true subject-heading and the plural form for the form

headings, e.g. Essay; Essays. But the appearance or presence of language do not always permit this, rather one has to add further something i.e. collected works. For example, the heading English Drama is sued for a book about English Drama, but in order to show that the book is a collection of dramas by several English authors the subdivision. Collected works must be added, e.g., English drama Collected Works.

Entry is also provided for general format of material and purpose of the work, e.g. Almanaes; Dictionaries; Encyclopaedias; Filmstrips; Yearbooks etc. These further can be subdivided by Language, Subjects and Nationalities.

2.7.3.8   **Grammar of Subject-Headings :** Simple subject-heading consists of a single noun and it is the ideal type which the language supplies. Such terms are not only the simple in form but often the easiest to comprehend. Some subjects can be expressed with direct headings, e.g. Art; Agriculture; Educational Religion; etc. There can also be direct objects, e.g., Apple; Poetry; Trees; Chairs; etc. But many words have synonyms from which a choice has to be made, and conversedly a word may have two or more quite different meanings; for others there may be a choice in spellings; another consideration can be user of singular or plural form. For example in the case of a 'Book on Pottery', other worlds that might be used are : Crockery, Earthenware; Stoneware; etc. In the Sears List, the terms chosen is "Pottery" but references are made from the given terms. As far as possible, the headings, terminology and subdivisions in Sears List follow the practice of Library of Congress subject Headings. Therefore, it becomes easy for a library to change over to library of Congress List, when a library becomes large in its collection.

2.7.3.9   **Compound Subject-Heading :** Sears List has joined two nouns or headings by the word "and" under one heading. These are closely related material which cannot be separated easily in idea and which is usually treated together, e.g., Black in literature and art; Blind Book and reading; Boat and boating, etc. There can be two different subjects that are treated in their relation to each other, e.g. Children-Diseases; Children-Food; Children-Language. Here in these example '–' is used to connect these two subjects. Sometimes there are two areas of knowledge which are expressed in Sears List by more or less complex phrases. These are the least-satisfactory headings as they offer the greater variation in wording and are often the longest words which may not be through of readily provided by the maker or thought by the user of catalogue. Even English language seems to offer no more compact terminology for this kind of subjects e.g. Information storage and retrieval system; Freedom of information; and so on.

2.7.3.10   **Methods of Sub-Divisions :** Sears List of Subject-Headings has used the methods of subdivision which has enlarged its scope beyond the actual heading printed in the list. The principle of specific entry can be achieved in some cases only by subdividing a general subject by worlds or phrases which indicate special aspects e.g.;

Birds

Birds –   Eggs and nests

Birds

Birds –   Eggs and nests

Birds –   Migration

Birds _   Protection

      -

Classical   –   Antiquities

Classical   –   dictionaries

Classical   –   education

Classical   –   literature

      –

Cold

Cold   –   Physiological effect

Cold   –   Therapeutic use

Cold   –   (Disease)

Cold   –   Storage

In each of the specific fields above, the subdivisions are appropriate and characteristic of it and those used under one are not applicable to the other two listed here.

Subdivision by Physical Form :   Some materials present a subject not in expository or narrative form but as lists, outlines, or tables; or graphically as maps, picture, or filmstrips. The work may be a directory of book publishers, a bibliography of children's literature, a dictionary of Library Science, a collection of geographical maps etc. In such cases, it is important to show to the users of the catalogue that they are not works about publishers, or children's literature, or library science, respectively. This will enable the reader to locate material directly without having to read through all the cards under the main heading. e.g.

Engineering   –   Periodicals

Publishers   –   Directories

Children Literature   –   Bibliography

Library Science   –   Dictionaries

Geology   –   Maps

Literature   –   Yearbooks

Motion picture   –   Catalogues; etc.

So always keep in mind while making subject-heading that there are given Standard terms known as "form division" for making common subdivisions and may be used whenever appropriate. Since they show what the material 'is', rather than what it is 'about', they are as necessary for a small library as for a large one. None of these headings appears in this form in the list, unless in has been cited as an example.

Therefore, the 'See' or 'See also' under the name of the form is to be interpreted as directions for use. Only when a heading has been located and added can be words 'See' or 'See also' be interpreted literally.

Subdivisions by showing Special Aspects : Any general subject may be presented from a particular point of view. A work may be a history of the subject, the most common of the special aspects; or it may deal with philosophy of the subject, or research in the field, or the laws about the subject, or how to study and teach a subject. These concepts appear to general subjects are expressed by such headings 'a's :-

Art	–	History
Art	–	Exhibitions
Art	–	Study and teaching
Art	–	Technique
Aret	–	Galleries and museums
Mathematics	–	Study and teaching
Aeronautics	–	Research
Religion	–	Philosophy
Radio	–	Law and legislation

Sub-Divisions through Chronology : In an Indian library there will be so many books on Indian History. If these all are entered under general headings, its users must look through many entries to find a book on specific era. However, with chronological subdivisions corresponding to generally accepted period of Indian History or to the spans of time most frequently treated in materials, a search can be narrowed, such as :-

India – History	–	Early time to – 1400
India – History	–	1401 AD – to 1600 AD
India – History	–	1601 AD – 1947, and so on

Sub-Divisions with the names of Place : In Sears List of Subject-Headings there is a provision under many headings to divide geographically in order to keep in mind the size of the list. For example, "Agriculature (May subdiv. geog.)". There is given one example and on that base one can divide any subject-heading in hand. Such as :

Agriculture	–	France
Agriculture	–	India
Agriculture	–	Punjab
Agriculture	–	Patiala

Under the heading "United States" so many subject-headings are given, thus one can adjust the topics from these subject-heading related to any part of the world, e.g. :

United States.Army	– India.Army
United States.Army	– Examination–India.Army–Examination
United States.Army–Officers	– India.Army–Officers

United States.Climate	– India.Climate
United States–History	– Indian.History
United States–Population	– Indian–Population
United States–Population	– China–Population
United States–Politics and Govt.	– India–Politics and Govt.
United States–Statistics	– China–Statistics

Subject Headings for Non-book Materials : Non-book materials are chart, games globe, kit, model, filmstrip, pictures, sounds recordings, etc. The assignment of subject-headings for audiovisual and special instructional material are guided to follow the same principles that are applied to books. The heading must specifically describing the contents of the material should be used, and the same headings are applied to book and non-book material alike. This is especially to keep in mind if the catalogue integrates all the media. One integrated catalogue is seems to be preferable because it would bring all materials on one subject together regardless of format. Non-book materials often concentrate on very small aspects of larger subjects, the cataloguer may not find in the List the specific heading that should be used. If the catalogue is kept separately then there is no problem to given about the format. It is necessary to give information about the format according to AACR-2 rules, if the catalogue is one for book and non-book material. For this purpose, on page 40 there is give list of non-book materials.

Subject Headings for New Subjects : No printed list in the world for subject-heading can be entirely upto-date. There is fast flow of new ideas, new inventions, or new countries being created. Headings for these new topics, ideas of current interest will have to be added by the cataloguer as and when needed. These should be constructed in the same way as headings for related topics and as the catalogue has been guided in the instructions given in the list.

It is not always possible to decide at once not he permanent form of a new subject-heading, but the catalogue cannot always wait for the subject to develop before giving headings to new thought. Approximate headings can be given and used until the terminology becomes standardized. A list of these tentative subject-headings should be kept in use for sometimes as it will never be long in use. But after sometime they can be reconsidered later and either adopted permanently or changed with some modifications and added in the permanent List. We must keep in our minds that new term is not merely a new name or a colloquial lists for a subject already in the catalogue.

In the end, we can say that subject-headings well always be a of greatest Usefulness to the notice in a subject area. The expert tends to know the names of authors and the sources of continuing information in a subject area. The novice must have some place to begin, which is provided by the subject-access to the material. It is only

when the novice begins to study the means of subject access closely that one becomes aware of its deficiencies. Usually one well condemn one's own lack of knowledge long before one realizes that his ignorance was shared and compounded by the makers of the subject-headings.

2.7.3 Self-Check Exercises

1. Define Subject Headings.
2. Who is the author of Sears List of Subject-Headings ?
3. When was first edition of Sears List of Subject-Heading published ?

2.7.4 Reference Books

1. Sears, Minnie Earl : Sears List of Subject-Headings. Edited by Barbara M. Westby. 12th ed. The H.W. Wilson Company, New York, 1982.

2. Sears, Minnie Earl : Sears List of Subject-Heading. 18th ed.

3. Girja Kumar Krishan Kumar : Theory of Cataloguing, 5th rev. ed. Vikas Publishing House Pvt. Ltd., New Delhi, 1986.

4. Singh, S.N., Prasad, H.N. : Cataloguing Manual AACR-II. B.R. Publishing Corporation, Delhi, 1985.

5. Sengupta, Benoyendra, Kesavan, B.S. : Cataloguing : Its Theory and Practice. 3rd ed. rev. The World Press Private Ltd. Calcutta, 1974.

6. Viswanathan, C. G. : Cataloguing : Theory and Practice. 4th ed. Rev. Today and Tomorrow's Printers and Publishers, New Delhi, 1970

7. Krishan Kumar : An Introduction of AACR-2, 3rd ed. rev. Vikash Publishing House Pvt. Ltd., 1990.

8. Viswanathan, C. G. : Cataloguing : Theory and Practice. 5th ed. rev. Print House (India), Lucknow, 1983.

LESSON NO. 2.8   **AUTHOR : DR. AMRITPAL KAUR**

LIBRARY OF CONGRESS LIST OF SUBJECT HEADINGS (LCSH)

2.8.1 Background

Library of Congress List of Subject Headings which is now running in its 20th ed. (1997) was first published in 1914 by Library of Congress (LC) under the title: Subject Headings used in Dictionary Catalogue of the Library of Congress. Since then, it has become the standard list used by most large general libraries, special libraries and some smaller libraries in the United States and abroad. It is now available in three formats : print, machine readable and microfiche. The machine readable version is also available in CD-Rom, called CDMARC subjects. The list is revised weekly and the CD-Rom version quarterly.

2.8.2 Format of Headings

In the List, authorized headings appear in boldface a.g. Books, Book Binding, Reference books, etc. Each valid heading is followed by scope notes, cross reference, and the subdivisions, if any.

Entries printed in light face roman type are not to be used as subject headings. The general format of an entry in the list is as under :

Reference books

    [Z 7 1 1]
    UF   Bibliography – Reference books

> Books, Reference
> Libraries – Reference books
> Reference books – English

BT   Bibliography
> Books and reading

NT   Bibliography – Best books
> Children's reference books
> Encyclopedia's and dictionaries

2.8.3 Formation of Subject Headings

In LCSH, subject headings are constructed in a variety of ways ranging from single noun headings to complex and compound headings. Various form of heading which have been used in the list are described below :

Single noun heading : the simplest form of main heading consists of a single noun e.g., Agriculture, Botany, Cataloguing.

Abstract ideas are usually stated in singular e.g. Library, Love where as concrete objects and things are rendered in plural e.g. Short Stories, Doctors.

Phrase Heading : When a subject or concept cannot be properly expressed by a single noun, a phrase is used. There are several patters of phrase headings which are discussed below :

Adjectival Phrase Headings : The most common type of adjectival heading consists of a noun or noun phrase with one of more adjectival modifiers e.g. Russian Art, English Drama, Higher Education etc.

Conjuctive Phrase Headings : The headings consist of two or more nouns, noun phrases or both, with or without modifiers, connected by the word 'and'. They are mainly used for the following purposes :

> I.   To connect topics or concepts which are usually treated together in books e.g. Emigration and immigration.
> II.   To connect opposite subjects that are usually treated together in books e.g. Open and closed shelves.
> III.   To express a relationship between two concepts e.g. Literature and science, Libraries and schools.

Prepositional Phrase Headings : These heading consist of nouns, noun phrases or both, with or without modifiers, connected by a preposition e.g. Photography of children, Divine rights of kings.

Inverted Phrase Headings : In many cases, a phrase heading is inverted in order to bring a significant word in a prominent position as the entry element e.g. Art, Medieval; Chemistry, Organic.

Subdivisions :

In LCSH, four kinds of subdivisions i.e. form, topical period and geographic are

recognised. Each form is discussed below :

Form Subdivisions

From subdivisions are used to indicate the form in which the material on a subject is organised and presented e.g. Economics–Periodicals; Chemistry0–Dictionaries.

Topical Subdivisions

Topical subdivisions are used under main headings or other subdivisions to limit the concept expressed by a main heading to a special subtopic e.g. the heading Agriculture–Accounting means accounting as applied to the field of agriculature and does not mean accounting as a kind or a division of the subject agriculture.

Geographic Subdivisions

A geographic subdivision indicates the origin or the locality of the main subject. Heading that may be sudivided by place carry the designation (May Subd. Geog.) immediately after their listed in the list. Geographic Subdivisions are added to the subject heading either directly or indirectly, depending on the place in question.

       Direct geographic subdivisions

       Horticulture – Zeneva

       Art – India

       Indirect Geographic Subdivisions

       Forticulture – Switzerland – Zeneva

       Charities – Italy – Florence

Period Subdivisions :

In LCSH, we can add the time division by five different ways :

1.   Noun Adjective Heading : Ancient, Medieval, Renaissance.
2.   Phrase of Historical Events without any date e.g. Colonical Period Revolution.
3.   History period followed by a date.
        Middle English (1100-1500)
        Early Modern (To 1700)
4.   In the form of century : English Fiction – 19th century.
5.   Subdivisions constructed ith the preposition 'to'
        H.S. History – to 1455.

Free–floating Subdivisions

Free–floating subdivisions are listed either separately or under representative main headings called 'pattern heading', with the intention that they may be combined with appropriate main headings at the time of application. There are four categories of free–floating subdivisions. These are :

Free-floating subdivisions of general application

These subdivision are applicable to a large number of headings. A list of these subdivisions appears in Subject Cataloguing Mannual : Subject Headings. Following

are some examples of free-floating subdivisions.

* Abstracts
* Library resources
* Lighting
* Software
* Study and teaching

Under each subdivision in this list, instruction is given as to the types of headings to which the particular subdivision is applicable e.g. the subdivision-lighting is applicable only under main headings representing vehicle, building, rooms, installations etc.

Free-floating subdivisions under specific types of headings

These subdivisions are used with the following categories of main headings :

* Classes of persons
* Ethnic groups
* Names of corporate bodies.
* Names of persons
* Name of places
* Name of bodies of water, streams, etc.

Examples :

Asian - Americans - Race identity

American Library Association - Employees

Milton, John 1608-1674 - Political social views

Free-floating subdivisions controlled by patter headings

Some form or topical subdivisions are common in a particular field. Instead of authorising them heading by heading and repeating them under each heading within the category, they are listed under a chosen heading in the category. This chosen heading then serves as a pattern heading of subdivisions for headings in that category. In LCSH the applicator subdivisions are displayed under the pattern heading. The subdivisions listed under a patterns heading may be transferred and used with another heading in the same category even though the combination does not appear in the list e.g. under English language, the pattern heading for languages, the subdivision-Phonology is listed. Therefore the combination Russian language-Phonology may be used, even though the combination does not appear in the list.

Free-floating subdivision indicated by multiples

Certain subject headings carry multiple subdivisions e.g.

Vietnamese Conflict, 1961-1975-Foreign public opinion-British [German, Russian, etc.]

In this example the terms given in square brackets serve as examples of similar subdivisions that may be used without prior authorisation e.g. Vietnamese Conflict, 1961-1917-Foreign public opinion-French.

2.7.4 Cross References

From the 11th edition of LCSH in 1988 'sec' and 'see also' references and the complementary 'x' and 'xx' wer replaced by the thesaurus symbols UF, USE, BT, NT, RT and SA. In the LCSH, main types of relationship expressed by cross references are equivalence, hierarchy and association.

Equivalence Relationships : USE/UF (used of) references are made from the unauthorized or non-preferred terms to the authorised of preferred headings for the subject in question. USE/UF (used for) references are made for :

Synonyms

Bookseller's Catalogs	Books - Catalogs
UF Books - Catalogs	**USE Bookseller Catalogs**

Virinat spellings

Archeology	Archaelogy
UF Archaelogy	

Earlier form of headings

Online catalogs	Catalogs, Online
UF Catalogs, Online	**USE Online catalogs**

Abbreviations and acronyms

CAD	**Computer-aided Design**
USE Computer-aided Design	UF CAD

Opposite tems

School attendance	Absence from school
UF Absence from school	**USE School attendance**

Alternative endings

Theory of knowledge	Knowledge, Theory of
UF knowledge, Theory of	USE Theory of knowledge

Hierarchical Relationship

Previously included under see also reference hierarchical references indicate topics that are either broader or narrower in scope than the one in question. Two symbols BT (broader topics) and NT (narrower topics) are used for these purposes e.g.

Reference Books

BT Bibliography

Books and reading

NT Bibliography-Best books

Children's reference books

Encyclopedias and dictionaries

Associate Relationship :

The symbols RT (related topic) is used to link headings that are related in concept but not in a hierarchical sense. RT references are provided under both terms involved.

Such references are usually made for the following types of relationships.

(i)   Heading with meaning that overlap to some extent :

Aliens

RT Citizenship

          Immigrants

          Nationalization

(ii)   Heading representing a discipline and the object studied :

Entomology   **Insects**

RT Insects   RT Entomology

(iii)   Headings representing persons and their field of endeavour

Medicine   **Physicians**

RT Physicians   RT Medicine

General References :

General References are represented by the symbol SA (see also), which refer from one heading to a group of headings or to subdivisions used under other headings e.g.

          Atlases

          SA subdivisions Maps under names of countries, cities etc., and under topics

13.5   Subject Headings for Special Types of Materials :

Detailed rules have been provided for literary works, biography and children's literature.

Examples :

          American drama - 20th century

          Novelists, Russian - 19th century - Biography

          United States - Biography

          Physicists - Biography

          Art - Biography

          Twain, Mark, 1835-1910 - Biography

          Children's Poetry, American

          English language - Dictionaries, Juvenile

          Mars (Planet) - Juvenile literature.

2.7.6   Comments on LC List of Subject Headings :

LC List of Subject Headings is no double and authoritative list based on the experience and practice of a large national library, but it has certain drawbacks which are an under :

          1.   The list is useful for preparing the subject catalogue of libraries and not for documentation lists.

          2.   The list is not based on scientific principles but on artitrary decisions. No attempts has been made to apply the latest theory of subject headings. Hence contains a large number of headings.

3. It does not solve all the difficulties of catalogue in the selection and use of subject headings. Headings are not specific enough for subjects covered by many documents especially when subjects are becoming increasingly complex.
4. Reference are not always constructed systematically.
5. Related subjects are scattered from Alphabets A to Z.

2.7.2 Summary

Subject cataloguing is an important area of cataloguing. It attempts to bring out the subject content of a document being catalogued and satisfy user's subject approach to catalogue. This session discusses the important features of Library of Congress List of Subject Headings.

2.7.3 Glossary

Adjectival Heading	:	In consists of a noun or noun phrase with one or more adjectival modifiers.
Antonyms	:	Words opposite in meaning.
Associate Relationship	:	It indicate headings that are related in concept.
Broader Term	:	The term denoting super-ordinate idea in a hierarchy.
Compound heading	:	It consists of two nouns connected by the word 'and'.
Cross-reference	:	A direction from one heading or entry to another.
Equivalence Relationship	:	In this relation, tow or more terms are regarded, for indexing purposes, as referring to the same concept.
Hierarchical Relationship	:	It indicate topics that are either broader or narrower in scope than the one in question.
Phrase Heading	:	In consists of group of words without a verb forming a short expression.
Synonyms	:	Words with the same meaning.

2.7.4 References :

1. Bakewell, KG.B. (1992), Manual of cataloguing practice. Oxford : Pregamon Press.
2. Chan, LM. (1994), Cataloguing and classification : An introduction. 2nd ed. New York : Mc Graw Hill.
3. Girja Kumar and Krishan Kumar (2000). Theory of cataloguing. 5th rev. ed. New Delhi : Vikas Publishing House.
4. Library of Congress (1997). Subject Headings. 20th ed. Washington D.C. : Library of Congress.
5. Rowley, J. E. (1987). Organising knowledge : An introduction to information retrieval. England : Gower.

6.   Tripathy, S. M. (1982) Modern cataloguing. 2nd rev. ed. Agra : Sham Lal Agrawala.

2.7.5 Self Check Exercise

Note : Write the answer in the space give below the question and check your answer with the answer given at the end.

1.   List various forms of headings used in Library of Congress List of Subject Headings.

----

2.   Various forms of headings which have been used in the list are described below :

Single noun headings : The simplest form of main heading consists of a single noun e.g. Agriculture, Botany, Cataloging.

Phrase Heading : When a subject of concept cannot be properly expressed by a single noun a phrase is used. There are several patterns of phrase heading which are given below :

(a)   Adjectival Phrase Headings e.g. English Drama.

(b)   Conjunctive Phrase Headings e.g. Literature and science.

(c)   Prepositional Phrase Headings e.g. Photography of children.

(d)   Inverted Phrase Headings e.g. Art, Medieval.