Indira Gandhi National Open University (IGNOU)

Bachelor of Library and Information Science (BLIS)

STUDY MATERIALS

Course code: BLI-223
Organizing and Managing Information

JATINDER SINGH
BLIS (JULY-2018)

www.jatinderjyoti.in
jatinderjyoti84@gmail.com
fb/insta: jatinderjyoti.raina

Indira Gandhi
National Open University
School of Social Sciences

BLI-223
Organising and Managing Information

Block

1

LIBRARY CLASSIFICATION

Programme Design Committee

Programme Coordinators

Prof. Jaideep Sharma and
Prof. Neena Talwar Kanungo

Course Coordinator

Prof. Jaideep Sharma

Programme Editor

Prof. Jaideep Sharma

Course Preparation Team

Unit No(s).	Contributor(s)
1-4	Prof. M.P. Satija

Course Editor
Prof. Jaideep Sharma

Internal Faculty:
Prof. Jaideep Sharma
Prof. Neena Talwar Kanungo

Material Production

Mr. Manjit Singh
Section Officer (Pub.)
SOSS, IGNOU

Secretarial Assistance

Ms. Sunita Soni
SOSS
IGNOU

Cover Design

Ms. Ruchi Sethi
Web Designer
E Gyankosh
IGNOU

November, 2013

ISBN :

Further information on Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi-110 068 or visit University's Website http://www.ignou.ac.in.

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi by Director, SOSS.

Lasertypesetted at Graphic Printers, 204, Pankaj Tower, Mayur Vihar, Phase-I, Delhi-110091.

Printed at :

UNIT 1 BASIC CONCEPTS

Structure

1.0 OBJECTIVES

After reading this Unit, you will be able to:

- define classification and know its various meanings;
- understand that it pervades every activity of life;
- explain the process of classification;
- discuss the various manifestations of library classification;
- state how it is vital to library management and services;
- understand classification as foundation study of library management, and also its limitations; and
- explain the basic concepts of classification.

1.1 INTRODUCTION

Classification is a process of making classes. A class is a set or group of entities (both abstract and concrete) having at least one similarity/ commonality. This similarity is called characteristic and is the basis of grouping or sub-grouping of entities. For example, all the students of BLIS class, whether male or female, of any religion and caste, coming from different states or regions, speaking different languages, having different political ideologies, have one characteristic in common, that is all of them are candidates for BLIS degree. A class can be of any size. All human beings make one class called *homo sapiens* by scientists. All Indians make one class. Similarly, all Christians make one class, Roman Catholics make another class, Indian Roman Catholics make yet another class, or Keralitie Roman Catholics may make yet another class. There seems no end to making classes and subclasses of people or of any other entity. A class may be of even a single entity.

1.2 MEANINGS OF CLASSIFICATION

Classification is a process of grouping of similar or like entities. It may be noted that there can be no grouping without division, as there can be no shadow without light, or no parting without meeting. Therefore, grouping implies division. Grouping and division are two sides of the same coin. We add a member to a group by separating it from other in the process of grouping and regrouping. Therefore, grouping and division are the basic processes of classification. But classification is more than endless grouping and sub-grouping. After grouping starts the process of ranking that is arranging the members of the group in a sequence. Even a small family can be divided further by status or age of its members. This is ranking. For example, say all the twenty students of a class may be further arranged in a row by age, height, educational merit or even alphabetically by name. Let us take the case of candidates appearing in IAS examination. The final result by UPSC divides them into two groups: successful and unsuccessful. Successful candidates are further ranked according to the marks obtained. That ranking is very vital. All this is classification. Classification is a systematic and predictable order. A group of chemical elements is a class by itself. Their grouping into Group 0, Group 1,2,3,….8 is further classification. Their further arrangement according to their atomic number is classification and ranking. In the third sense, assigning each ranked entity a code or symbol to preserve their ranking is classification. For example, a class of 25 students may be first arranged according to educational merit then each students may be ranked 1, 2, ….. 25 or A to Y in order of merit for convenience of handling. This allocation of codes will mechanise and fix their ranking and consequent sequence.

Various manifestations of classification

Grouping and division seem primitive or elemental processes of classification. Looking at the bottom, classification is co-relation or discovering relations between entities. All members of a group are related to one another by some commonality. When we admit an entity into a group, a relation between that entity and group is discovered or created. For example, Potassium is not only a member of the class 'inorganic substances', but also bears relation to sodium on its left and copper on its right. Sodium, potassium and copper are related to one another. In a family, which is always a class by itself, all members are related by blood. Hence classification is relation.

1.2.1 Classification and Organisation

Since grouping and inter-group ranking are acts of organisation, thus classification is organisation. In fact classification and organisation are inseparable. Now classification is considered as a tool for organising in every sense of the word. So classification is structuring and mapping. Difference between a heap of bricks and a mansion is classification. In a mansion every brick is positioned in an organised way.

In an organisation all members are related and coordinated, so classification is co-ordination and control. Difference between a disciplined army and a chaotic mob is classification. Army men are coordinated and controlled while a mob is uncontrolled, though both the groups comprise of men.

Classification is matching and pairing which is implied in grouping of entities brought together. When we are ranking and arranging we are sorting and tabulating. So, tabulating and sorting are acts of classification.

1.3 USES OF CLASSIFICATION

Classification is a mental act and logical process of association and relation. It goes on every moment of life knowingly or unknowingly deliberately or unconsciously. Any system be it biological (man), social (government, libraries, institutions) or mechanical (computers, machines) has to classify for successful functioning. All human beings, what ever they do, have to classify in every sense of the word. More sophisticated and intelligent a person, better his/her sense of classification.

A postman classifies postal items for efficient and timely delivery. For quick, efficient and easy delivery, a postal item is sorted (classified) many times at different stages between posting and delivery. A fruit seller sorts his fruits into categories, say, oranges, apples, grapes, and so on. Further each group, say, of apples is further sorted into species say Kashmiri apples, Simla apples, Golden apples, Green apples, etc. An astute vendor may further sort each species by quality and price. At every step of grouping sorter is adding value to the items. Thus classification is value addition.

Record files in an office are arranged in some order, and within each file letters and memos are arranged in some known order. Without such an arrangement the previous record cannot be located and used. Books and other reading material in libraries are arranged, no doubt to increase their usability.

1.4 SCOPE OF CLASSIFICATION

There is no act of life where classification is not used. It is applied everywhere. It is a basic process to learn. Opposite of classification is disorder and chaos. Classification can be done of all objects entities, actions, thoughts and concepts. We can classify people, countries, natural phenomena, plants, flowers, animals, libraries, philosophies, literature, artifacts, automobiles – what not. It is a universal constant. It is the only method to simplify, understand and comprehend a complex universe to discover its structure and impose some order over the otherwise chaotic world.

1.5 PROCESS OF CLASSIFICATION

Classification is a process of co-relation. It is a way of thinking – thinking systematically and purposefully. It is an aid to memory and reasoning power. Nothing can be identified without it. It means to define an entity is to classify it first. For example, a gun is a firearm; a chair is a piece of furniture, a car belongs to the class of vehicles, and so on.

All thought and reasoning contains some process of classification.

English philosopher J.S Mill (1806-1873) says that classification facilitates the operation of the mind in clearly conceiving and retaining in the memory the nature of the entity or phenomena.

Someone has aptly and axiomatically defined empirical science as "a systematic classification of experiences". Therefore classification is training of the mind. It is often said that to learn to classify is itself an education. "Sharpness in thinking, clarity in expression, exactness in communication depend ultimately on classification", says Ranganathan (*Prolegomena*, X B2).

A group is divided or a member is included into a group on the basis of some characteristic. A characteristic is an attribute, quality or property of an entity which relates it with, or separates it from a group. For example, a group of people may be divided into males and females. Here "gender" is the characteristic of division. All the students of a university may be divided into under-graduate, postgraduate and research degree students. Here level of education is the characteristic. Books in a library may be arranged on the basis of their subject content. Thus a characteristic is the basis of division. Successive application of right and relevant characteristics produces deeper and finer classification.

1.5.1 Genus-Species Relation

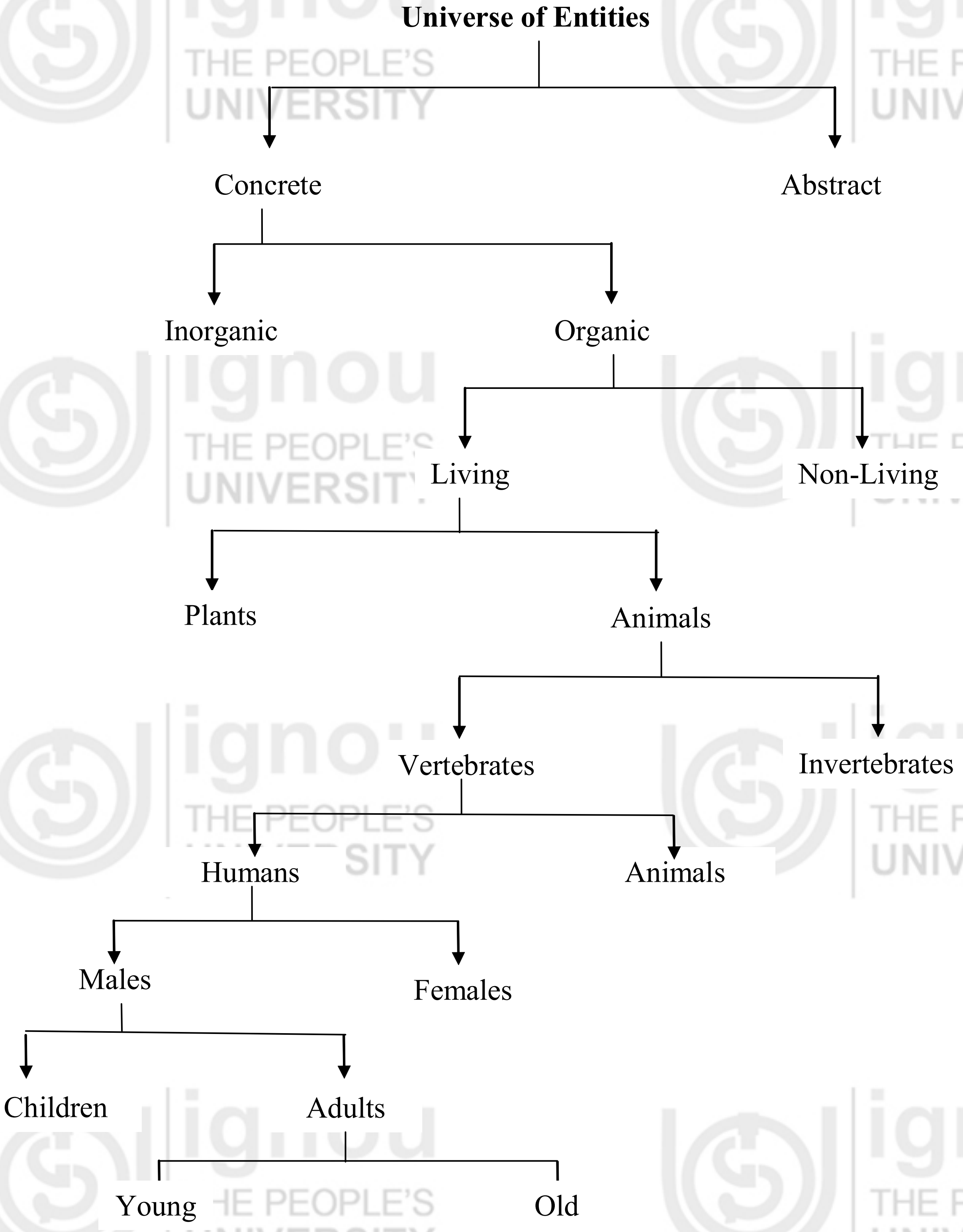
A class or group is logically called a genus, and the characteristic is the difference we add to produce species:

Genus + Difference = Species

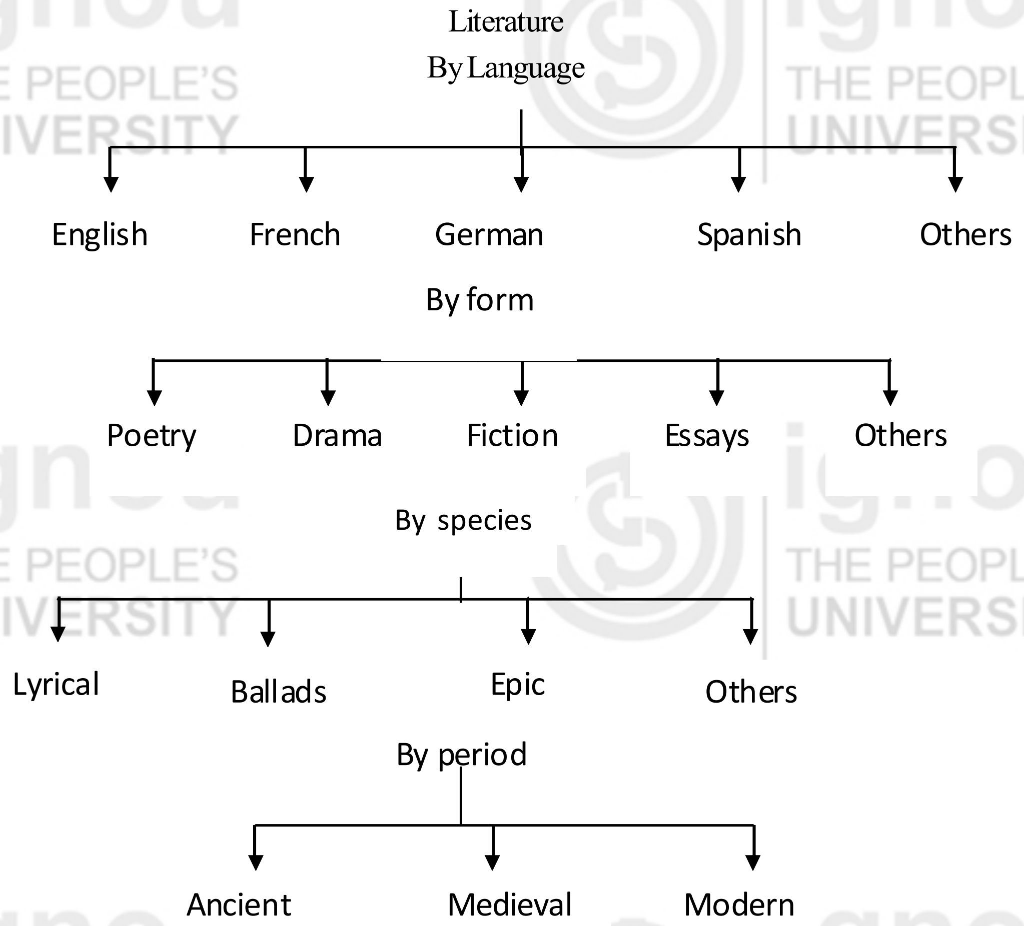
Eg., Tables + Material = Glass tables, Wooden tables, Plastic tables, Metal tables, etc. Here material is the characteristic to divide the universe of tables.

Here table is genus, material is the difference, and glass tables, wooden tables, etc. are species of table. Ancient philosophers applied dichotomous method to divide the universe into two groups at every step. Greek philosopher Porphyry (232-304 AD) used this method and the resulting groups and subgroups are known as tree of porphyry. This method, however is artificial, as every phenomena is not dichotomous in the universe: there are many shades between white and black. Modern method is to divide by genus – species, or by whole-part methods.

Division by Dichotomy



Let us apply genus-species to literature by applying characterstics:



We proceed from broader to narrower classes by applying respectively the characteristics of language, form, species, and period. Nature, quality and mode of application of these characteristics is very important. (You will read more about the characteristics in Unit 3A: Postulation Approach to Classification.)

To classify an entity we must have some knowledge of it. A guitar cannot be classified unless we know it is a stringed musical instrument. To classify cricket (game) we must know it is an out-door game played with a bat and ball.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

1) Explain the scope and methods of classification.

....

....

....

....

1.6 NATURE OF CLASSIFICATION

Logically speaking classificatory groups are not absolute; classifications are relative as something is classified with reference to others. An entity cannot be classified or ranked in itself. It takes two to create a classification. It means a unique entity cannot be classified. It is a class of its own. No classification is absolute also means that classifications are not permanent. Classifications are not real even. No classification exists in nature. All classifications are man-made, and made for a purpose. No classification is good or bad; these are helpful or unhelpful to a varying degree. It all depends upon the purpose of classification. A classification which serves its purpose well is best, whether logical or not. Logic of classification depends upon the characteristic chosen and the order in which these are applied. A large group of persons could be divided by age, gender, religion, race, nationality, mother tongue or colour of skin, hair or eyes, and many more characteristics. Each time it will result in different grouping. Choice of characteristics and the order in which these are applied one after the other will depend upon the purpose of classification. For example, a farmer would place birds, rats, insects and monkeys in one group as enemies of his crop. A scientist may laugh at such a classification. Both are correct, as their purpose is different. A farmer produces food while a scientist produces knowledge. A child or layman thinks that birds, butterflies and bats belong to the same class as all these can fly, while for a scientist all the three belong to different classes, each of its own. Different classifications produce different maps or depict different structure of knowledge depending upon the society and time of its origin. Classifications are not neutral. These are mirrors that reflect their time, place and society. Vedic classification is different from the one produce by Aristotle. Classification of knowledge by English philosopher Francis Bacon (1561-1626) is quite different of the two. No two classifications are similar.

1.6.1 Classification as a Tool

According to Aristotle (384-322 BC) classification is theoretical, practical and productive science. As the saying goes theory is the most applied knowledge, we can arguably say that classifications are always practical and designed for some purpose. As already said, classification is a tool for simplification, understanding and organisation. Without organisation nothing works. All the uses of classification may be summed up as: management; aesthetics and knowledge creation.

Organisation is for better management. Classification organises anything and everything: life to a shoe store; knowledge to libraries. A library, archive, postman, grocery shop all use classification to save time. It ultimately leads to economy of time, money and manpower.

Aesthetics is the science of beauty. An arranged store appeals visually to the visitors. Classification was also described as pairing and matching. So a housewife matching the colour of her room curtains with other items in the room is essentially doing an act of classification. A gentleman matching the colour of his tie with the suit he is wearing is an act of classification, too. Here the purpose is purely aesthetic to feel good and look good.

Philosophers, scientists and researchers classify to study and understand the growth and structure of knowledge. It is to outline and map the vast sea of knowledge. Without this map it will not be possible to navigate and further explore this boundless sea. A formal researcher has to select, tabulate and co-relate data to create new information. All the three stages are acts of classification. Hence classification underlies research also.

Classification is pattern making and pattern recognition. Computer retrieves information by patterns recognition and comparison, hence works by classifying; so does our brain which always works by association, grouping (integration) and pattern recognition.

Selecting a life partner for marriage involves a series of classification acts. Marriage is selection and pairing – both are manifestation of classification. Characteristic selected for marriage are: age, religion, caste, physique, looks, job, financial status, values and attitudes, educational qualification, and many more. In which order you apply these characteristics depends upon your need and values. Marriage is selection, and every selection is classification.

1.7 KNOWLEDGE CLASSIFICATION

As said earlier, classification can be of any object, phenomena, concept or acts. Classification and categorisation of knowledge *per se* is called knowledge classification. From time to time philosophers, scientists, educationists and the likes have formally categorised entire known knowledge to outline its boundaries and show the structure of knowledge. For example, Hindu Vedas (1500 BC) divided knowledge into four categories in the order: Dharam, Arth, Kam, and Moksh. Aristotle (3rd century BC) divided knowledge into three parts: theoretical, practical and productive, and further divided entire knowledge into ten categories. A propedia is classification of knowledge and vice-versa. Knowledge is defined as sum total of ideas, theories, experiences, history, feelings, values, sciences, symbols, arts, facts, fiction, myths conserved by a society. Classification of knowledge is essential for its simplification, understanding and progress. Without its organisation no further growth and progress can be made. For example, a new idea or a discovered fact will not become knowledge until it is related and integrated with the existing knowledge. Therefore, it has aptly been said that all knowledge is classification.

1.8 LIBRARY CLASSIFICATION

Libraries are established to house and preserve books and other documentary heritage of mankind. Books and other information sources are knowledge objects and can also be classified like other physical objects. Since antiquity librarians have classified books to produce convenient groupings and to facilitate their location at the time of need. An unarranged collection is a heap of books, not a library by definition. To find a book from such a library will be like locating a needle from a huge heap of hay. In earlier times books were grouped and arranged on the basis of their language, size, colour of binding, authorship or broad subject categories. Those methods were perfectly useful in those times as the main aim of libraries was to store and preserve documents rather them to serve them to the scholars. Access to knowledge was the preserve of the privileged few.

1.8.1 Modern Library Classification

In the then emerging industrial society of the late nineteenth century there was an attitudinal shift in the values of education and libraries. Importance of literacy was recognised for aware and responsible citizens. In a democracy access to education was democratised and opened to all. “Education for all” became the objective of the welfare state. To meet the needs of the society not only many new libraries were established by law, the doors of libraries were open to all and sundry — scholars, students, neo-literatures, poor, children, housewives, old and challenged, and other marginalised sections of the society without any discrimination. Further to maximise their use books were placed in free stacks and users were allowed open and direct access to the books. That open access policy required new and better arrangement of books for the browsers. Then Melvil Dewey (1851-1931) designed his Decimal Classification which divided knowledge by academic disciplines of study and used decimal notation to denote subjects. Latter provided almost infinite capacity for expansion and insertion of new subjects at proper places. Since then the books are being classified predominantly on the basis of their subject content that is knowledge. Thus library classification is knowledge classification as applied in libraries. In other words library classification is applied knowledge classification. But library classification is lot more than knowledge classification as it has also to take into account the physical aspects of the documents, the way knowledge has been formatted and presented in them, as well as the viewpoint of the author. It includes aspects such as language, media, form, format, viewpoint and many more such things. Formally and traditionally, library classification has been defined as the arrangement of books and other reading material in a way that is helpful to the users. Today's libraries are arranged by subject though different types of collections are arranged in different ways. For example, government documents, patents or standards are arranged by their own official codes. Current periodicals are arranged alphabetically by title. Maps, CDs, pamphlets, photographs indeed require different and separate arrangements. Making of library classification systems is also classification. Designer of a classification scheme is known as a classificationist. Operating a classification system to assign class numbers to documents in a library is also classification; such a person is called a classifier.

1.8.2 Uses of Classification in a Library

Classification is vital to library services. In fact classification is implied in definition of a service library. In supports all library services. Classification is to a library as skeleton is to human body on which all the body organs rest. Classification of a library collection is like map of a city. In a library, classification serves all the functions given above, namely it is a tool of management, aesthetics and knowledge creation. All the Five Laws of Library Science (1931, 1957) formulated by Ranganathan support library classification and have specific implications to design quality classification systems. Without classification a library is an unorganised dump of books. Therefore, without it the full value of a library collection cannot be obtained. However, its specific uses can be broadly listed as:

- It brings together books on the same subject. Thus a user gets all the books at one place which is much more convenient to the users.
- It facilitates the browsing function of a library. Browsing is to look at library collections without any specific purpose – a sort of window shopping. Browsing is a habit with the scholars. It is only possible fruitfully in a library organised by subjects. Browsing always leads to incidental discovery of long needed and valuable information. This accidental discovery is known as serendipity. Not only this, the general to specific order of arrangement with some notational maneouvering has been turned into pedagogical order in schemes like the CC. Ranganathan calls it APUPA pattern on the shelves. It is quite helpful for the self-learners. Thus systematic arrangement of books in open access libraries is helpful in self learning.
- It is a location tool; without it the library catalogue will not be able to function properly. It is also used for preparing shelf lists.
- Classification is the basis of all information retrieval systems and methods both in manual and electronic systems.
- It helps to replace the books at their correct places when the books are returned to the stacks after the home use or use within the library.
- It has been claimed that a library classification has three functions, namely, cognitive, information retrieval and shelf arrangement. Cognitive function is to represent the structure of knowledge and intra-relation of subjects. That is to produce a map of knowledge. Many library classifications, e.g., Ranganathan's CC, Bliss BC and BSO have emphasised this function.
- Many bibliographies, catalogues etc. are classified for better use. UDC was created to arrange entries in a universal bibliography.

- It has been found useful in reference service for facet-analysis of users' questions in reference interviews. It is useful for arrangement of circulation record.
- It helps in building a balanced collection of documents in a library.

Uses in Electronic Environment

Classification can be easily used to arrange and retrieve records in electronic databases. Online Public Access Catalogues (OPACs) function far better when class numbers are provided as another access point. In fact, in the electronic information era it has found so many new uses that it is rightly said that we are witnessing the second golden age of classification. Conventional classification systems such as the DDC, UDC, LCC have been used to organise and search information on the world wide web (WWW) . Search engines like Yahoo, Google, AltaVista use broad classification methods for organising their information. Eccellio (http://science.eccellio.com) is a search engine which uses faceted classification to return precise information. It uses Google database but adds an extra level of classification to refine search. They have defined it as Google^{++}. In the web environment at least seven functions have been identified by Professor Lois Mai Chan. These are location, browsing, hierarchical searching, retrieval, identification, sub-grouping (partitioning) and profiling. It has aptly been called mathematics of librarianship.

There are many day to day routine uses of classification in a library so much so that it will be impossible for a library to function properly and achieve its objectives without a sound classification. It has aptly been said that a book is the foundation of a library but classification is foundation of librarianship. Indeed there is no escape from, nor any substitute to it in libraries, or life.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

2) State the uses of classification in a library.

....

....

....

....

1.9 LIMITATIONS OF CLASSIFICATION

Classification was described as mathematics of librarianship, yet like the value of *x*(pi) it is never exact. Classifications are social, not natural. These do not satisfy the needs of all the library users. Only majority are served while some users with specialised needs require different arrangement. Classification, especially library classification has many limitations and problems. It is a costly process and subjective, too. Two classifiers may differ widely on the correct classification of a given book. Not only this, a classifier may class a given book differently at different times. No classification can comprehensively represent the total subjects dealt in a book. Even a monograph may deal with more than one subject at a time. In classification only the dominant subject is represented. A textbook on algebra and geometry is either placed at algebra, or geometry, not both. Further, let us say a textbook in cataloguing may have a very valuable chapter on history of cataloguing or on the life of C A Cutter. These side topics will not be represented by the class number, and may remain hidden from needy users. Classification by discipline also scatters subjects. For example, a book on "Family life" may be placed in different main classes such as Ethics, Sociology, Anthropology, Social Welfare, and

Medicine. Hence it also results into scattering. It is not incorrect to say that classification suppresses and scatters more than it reveals and collocates. A classification may not satisfy all the users as they have invidual needs. Classifications are not based on the survey of the needs of library users. A nineteenth century English philosopher W.S. Jevons (1835-1882) had criticised classification as a logical absurdity. This is no less true of library classification of which there is no substitute. We have to work with imperfect tools.

1.10 SUMMARY

Classification is a universal constant. It is an activity that goes on every moment of life. It is no exaggeration to say that we live by classifying. Broadly speaking, classification is the process of making classes or set of entities on the basis of their similarities. Grouping also implies separating as selection also implies de-selection or rejection. The criterion or basis of grouping is called a characteristic. Quality of final classifications will depend upon the right choice of characteristics and the right order of their successive application to produce subsequent sub-grouping. Ultimately classification is co-relation between two entities. It has numerous manifestations like sorting, grouping, ordering, arranging, ranking, structuring, coordinating, matching, mapping and pattern making. Classification can be made of all entities under the sun. Philosophers, scientists, librarians, shopkeepers, postmen, housewives all do classification for different purposes. The four broader uses of classification are organisation, economy, aesthetics and productivity. Many philosophers right from Aristotle have done classification of the entire universe of knowledge. Scientists have produced taxonomies of plants, animals and chemical substances. In libraries we apply knowledge classification to organise our books, databases and other reading material both in print and electronic form. In fact in computerised databases and network information searches classification has found new but powerful uses. Classification is so much the basis of all library services that it has been described as foundation study of librarianship. Yet classification has its own limitations.

1.11 ANSWERS TO SELF CHECK EXERCISES

1) Classification is universal. It applies to daily life routines and work. It is method of organisation, and can organise any object, phenomena and entity. It is a logical process of grouping and division. Ultimately it is to co-relate one entity with another. Grouping or division is done on the basis of characteristics i.e. dividing genus into species by adding a difference. Human beings can be divided into males and females by adding the characteristic/difference of gender to the genus of human beings.

2) Classification is basic to a library. A collection of books which is not classified cannot be called a library. It brings at one place all the books on a given narrow topic which helps in easy location and browsing. It arranges books in a pedagogical order which is useful for self learners. It also arranges records in catalogues and bibliographies. It helps in building a balanced collection. Its methods are useful in reference service. It can be very useful in searching electronic data basis and web. It is rightly said that it is the foundation of librarianship.

1.12 KEYWORDS

Characteristic : It is basis or criterion of division or grouping. If a group is divided into Hindus, Muslims, Christians, Sikhs then religion is the characteristic. Quality

		of the characteristic will determine the quality and aptness of classification.
Class	:	A set of entities having at least one characteristic in common.
Classifier	:	A person who classifies books in a library by operating a classification system.
Classificationist	:	A person who designs a classification system.
Classification	:	It is a process of grouping entities on the basis of likeness or some underlying relation. Ultimately, classification is organisation and co-relation. It is grouping, selecting, sorting, ordering, tabulating, ranking, mapping, preparing classification schedules and operating classification systems.
Knowledge Classification	:	The process of outlining, structuring and mapping the entire knowledge or some part of it. It helps to study the nature and growth of knowledge. It is also the basis of modern library classification.
Genus and Species	:	Genus is any original universe to be divided into species by adding some characteristic to the genus. These are relative terms. A father is a genus for the children; when children become father/mother they will be genus for their own children.
Library Classification	:	Arrangement of books and other reading material of a library in a way useful to the users. It is knowledge classification as applied to books and other packages of information.
Porphyry Tree	:	Dichotomous method of classification invented by the Greek philosopher Porphyry (232-304AD). It divides the universe into two antithetical groups at every step of division.

1.13 REFERENCES AND FURTHER READING

Philip, H.W. *A Primer of Book Classification.* 5th ed. London: AAL, 1961. 9-20. Print.

Chan, Lois Mai. *Cataloguing and Classification: An Introduction.* 3rd ed. Lanham, MD: The Scarecrow Press, 2007. 309-314. Print.

Ranganathan, S.R. *Prolegomena to Library Classification*, 3rd ed. Bombay, Asia, 1967. 77-79, 547. Print.

Satija M.P. "Classification: Some Fundamentals Some Myths". *Knowledge Organisation.* 25.1-2(1998): 32-35. Print.

Satija, M.P. "Classification: An Essay in Terminology". *Knowledge Organisation.* 27. 4(2000): 221-229. Print.

Satija, M.P. *A Dictionary of Knowledge Organization*. Amritsar : Guru Nanak Dev University, 2004.35-36. Print.

UNIT 2 TYPES OF CLASSIFICATION

Structure

2.0 OBJECTIVES

After reading this Unit, you will be able to discuss different types of classifications by:

- history, i.e. fixed and relative location systems;
- methodology, i.e. enumerative and faceted Systems;
- areas of application, i.e. general and special systems;
- depth of details, i.e. broader and depth classifications;
- media, i.e. print and electronic versions; and
- environment, i.e. classifications for the web and ontologies.

2.1 INTRODUCTION

Classification of documents and other reading materials is indispensable for any library. Various standard and local methods for arranging library materials, ranging from clay tablets, papyrus rolls, monographs and other print documents, audio-video material, CDs, multimedia and now web sources, have been employed from time to time by librarians to organise their collections. Their classification has varied from home-made or *ad hoc* systems to somewhat adapted from some universal knowledge classification systems. Since the late nineteenth century librarians have developed many standard classification systems pioneered by the Dewey Decimal Classification in 1876.

2.2 FIXED AND RELATIVE LOCATION SYSTEMS

2.2.1 Fixed Location Systems

These systems of Pre-Dewey era assigned a fixed place to a book on the shelves. The "call number" indicated the shelf ground on which the book was stacked. For example, a call number, say, 2.4.6.25 meant that it was the 25^{th} book on the 6^{th} shelf of 4^{th} almirah in the 2^{nd} room of the library. Thus a book could be located even by a blind person. Its advantage was saving of shelf space, as such systems require vacant place only at the end of the shelf or almirah. (Present day systems called Relative Location systems require space for intercalation of new books at every point on the shelves.) In fixed location, maintenance of subject grouping was difficult. New books could be accommodated at the end of the shelf. Whenever the books had to be shifted to another place their call number had also to be changed frequently. That involved lot of labour and wastage of time and money. Melvil Dewey (1876-1931) while working in the library of Amherent College (Massachusetts) could not tolerate this wasteful task of reclassification, too often. He invented a relative classification system to solve this problem. His question was : "How to give permanent call numbers to books in a library?"

2.2.2 Relative Location Systems

In relative systems class number refers to the intrinsic subject of the book rather than the shelf place. Decimal notation used by Melvil Dewey provided a neat technique for dividing knowledge and by denoting each division and subdivision by decimal fraction numbers. Call number indicated the subject rather than any fixed places on the shelves. It was a master stroke which brought a paradigm shift in library classification. The new books could be accommodated at their proper places without disturbing the relative location of the existing books. In the new method, the shelf location of books changes with addition of more books. Books on the shelves keep shifting to the right side, but their relative location of the document remains the same: on its right and left it will always have the same subject as its neighbours. Its great advantage was maintenance of strict and finely divided subject collocations. For example, a new book on geometry could be placed with earlier books on geometry without any problem. Not only this, Geometry could be further divided into Euclidean Geometry, Plane Geometry, Solid Geometry, Trigonometry, etc. Earlier a new book could only be placed at the end of the almirah containing books on mathematics. Fixed location systems are now a dead history as now all library classification systems are relative location systems.

2.3 BY DESIGN METHODOLOGY

2.3.1 Enumerative Systems

A classification is a map of knowledge which lists every subject and its subdivisions in a top-down approach. Each subdivision is given a notational mark to denote it. Known as class number, this notation, or a cluster of digits, is assigned to a document having that topic as its specific subject. These are also known as 'mark and park' systems. Enumerative classifications are pre-defined and frozen lists of subjects of the past, present and of near future. These only provide readymade pigeon holes for documents rather than finally individualising them according to their subjects and their various documentary aspects. Most of the time these prove square holes for round pegs. These are now considered rather old fashioned classification systems.

2.3.2 Faceted Systems

An enumerative system produces systematic but linear lists of subjects. Knowledge is multi-dimensional and growing dynamically. An enumerative classification can represent only one aspect of the specific subject at a time. Many aspects have to be left out. For example, in the earlier editions of the DDC a simple subject like "Anatomy of Dogs" could either be classified as "Zoology of Dogs" or "Anatomy of Animals". Both aspects could not be taken together. Thus such systems fail to classify co-extensively the present day knowledge, not to speak of the subjects to emerge later. By the beginning of the 20th century these enumerative models were not very effective, yet no other model was available, though the UDC (1895+) had made some improvements in the DDC to denote some more auxiliary aspects of a document.

S.R. Ranganathan (1892-1972) after a long study and experimentation in late 1920s developed a different method to classify multidimensional knowledge thrown by the 20th century industrial society. These are now known as faceted systems. A facet is any of the many sides of a cut diamond. Ranganathan used this term in classification to designate different aspects of a specific subject. Instead of making a long list of subjects in some systematic order he divided a subject horizontally into various categories and then vertically into different subdivisions known as facets and isolates respectively. For example, the subject of library science could be divided into 'Kind of Library' Facet, 'Kind of Document' Facet, 'Kind of Operation' Facet, and 'Kind of Service' Facet. The kind of service facet could be detailed as circulation services, reference services, current awareness services, reprographic services, and so on. Space and time facets are kept as common facets applicable to all classes of subjects. Later Ranganathan developed a theory of "Five and only Five Fundamental categories." These categories are Personality, Matter, Energy, Space and Time. His postulate is that any subject comprises of some or all of these categories. A subject is always made of any of these categories. Nothing is beyond them. These facets are converted into digits and then combined in some postulated order to produce unique class numbers to suit specific subject of the document. Thus a class number can be tailored to exactly fit the document instead of assigning a class number to a book from the long list of readymade class numbers as in an enumerative classification. From a small list of facets numerous class numbers can be produced by their combinations and permutations. It started a new revolution in library classification.

This faceted system was later refined and developed into a very dynamic and effective model based on postulates and principles for which Ranganathan developed a theory in his famous book *Prolegomena to Library Classification*, (Madras Library Association, 1937). Its advanced version is called Analytico-synthetic classification. Such classifications have proved useful for the growing universe of knowledge, for information retrieval, and later have proved basis of designing all indexing languages. These are equally efficient at the traditional role of shelf arrangement. Indeed these have become popular methods of modern knowledge organisation. All the new library classification systems are faceted, while old systems like the Dewey Decimal Classification, or the Bibliography Classification (BC) and UDC are getting faceted through revision.

2.3.3 Synthesis Grafted on an Enumerative Base

It is a mix of the two systems. These basically enumerative systems have later developed some special tables of documentary aspects to be combined with the base number. First such example is of the Universal Decimal Classification (UDC) which started simply as enhancement of the DDC in 1895. To the DDC base a multiplicity of auxiliary subdivisions were added to make a class number multifaceted. As survival approach, some additional internal tables are devised and many more provisions of number synthesis have been invented to make them more hospitable and faceted to combat the dynamically growing knowledge. The DDC in its 18th edition (1971) introduced five more tables and made many more provisions for number synthesis through instructions and special tables here and there. Being added quite late such a structure is not regular or uniform. It is an add-on provision which has its own problems.

2.4 KNOWLEDGE CLASSIFICATION AND LIBRARY CLASSIFICATION

Classification is a process, a logical visual method of simplification and understanding. No phenomena or object can be understood without classifying it. It organises all sorts of entities and depicts their due place in the universe. Classification can be both of abstract and concrete entities; of ideas and things. It is essentially a life process of learning, doing and living successfully. Human civilisation progressed as primitive humans learnt to classify the visible phenomena around. Let us say man learnt of edible and non-edible things; divided animals into useful and harmful groups. It was another (more sophisticated) act of classification when man related clouds with rain, and rain with growth of vegetation and life; and related certain herbs with certain diseases. All that became knowledge by and by. It has rightly been said that all knowledge is classification. Knowledge is defined as sum total of facts, beliefs, experience, memories, expressed feelings, arts, sciences, fiction and myths conserved by the society. Thinkers in all ages have tried to categories knowledge to understand its nature, categories, boundaries and growth. That became knowledge classification. Knowledge classification is outlining and mapping to depict its structure and boundaries. It leads to better understanding of its history, nature, kinds, properties, growth and also gaps in it. It becomes guide for the educationists, scientists and librarians. Knowledge classification is both speculative and empirical, and is a province of philosophers and scholars. One example of knowledge classification is Vedic categories (1500 BC) of knowledge in Dharma, Artha, Kam and Moksh.

Library classification is book or document classification. Documents are nothing but carriers of knowledge and information. In modern libraries documents are classified on the basis of their knowledge contents. Therefore, knowledge classifications have become basis of library classifications. But knowledge is abstract and fluid and takes the shape of its container and carrier. Books on the other hand are physical and solid entities to be arranged on shelves. Documented knowledge has some non-subject aspects such as language, viewpoint of the author, format, media, etc. All these have to be accounted for in a library classification. Library classifications are in fact more detailed, have notations/ symbols to be assigned to documents, separate common tables for physical aspects, and have also a detailed index of subjects. Therefore library classifications are more formal and complex than the knowledge classification on which these are based.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

1) Make a comparative study of the features of faceted and enumerative systems of classification.

....

....

....

....

2.5 WEB CLASSIFICATION: ONTOLOGIES

Ontology is a powerful classification tool for the semantic web which describes and represents an area of knowledge. It specifies description for classes, relationships that exists among entities and properties that entities have. Ontology characterises well defined concepts, their taxonomy and many sided relationships. It is a hierarchical collection of concepts arranged in categories combined with multidimensional relations in order to reflect vocabulary of that area of knowledge. It is a web of connections. An ontology which is used for knowledge organisation and retrieval in an electronic environment has the power of traditional hierarchical classifications, subject headings lists and thesauri. It combines elements of all the three. Ontologies are sources of controlled and standardised terms which help to organise information in a more precise and multidimensional ways. They yield better search results by using new search techniques and natural language processing. Some examples:

Biogen Idec: Using semantics in drug discovery research

www.bio-itworld.com/issues/2006/oct/biogen-idec.

MINDSWAP: Using ontologies to aid terrorism intelligence gathering

http://mindswap.org/papers/Terrorntology.final.pdf.

2.6 BY AREAS OF APPLICATIONS

Another division of library classifications can be by the subject area covered. There are general classification systems which cover the entire universe of knowledge; the other categories are special classifications covering a specified subject or limited area of knowledge. These by default are broader and depth classifications, respectively.

2.6.1 Special Classifications

A classification for specific area of knowledge, for example, Economics, even Banking, Occupational Safety, Diamond Technology, Women Studies, Indology, etc. are examples of special classification. Some examples of special classification schemes are:

Uniclass: Unified Classification for the Construction Industry (London: RIBA, 1997)

London Education Classification, University of London, 1974.

London Classification of Business Studies, (London Business School, 2000)

Thesaurus of Psychological Index Terms (Washington: American Psychological Association)

Special Classifications inevitably are depth classifications used for classifying and indexing micro literature in the form of journal articles, research reports, theses, etc. These are eminently useful for information retrieval in special libraries and information centers. Ranganthan calls it depth classification.

Kinds of Special Systems

Special classifications usually do not exist alone. Every special library has documents on its special area and also on related subjects. Evan a nuclear science library may have books on fiction, sociology, management, etc. A special classification such as classification of business studies may require another general classification for classifying documents in other areas such as political science, sociology, psychology, mathematics. For classifying in related and other areas usually a standard system is adopted. On the other hand a special classification may be a mere extension of certain class of a general classification system. For example, in India many local made detailed extensions exist of the DDC class numbers like 954 Indian History, 294 Indic Religions, 181.4 Indian Philosophy to adequately classify such subjects in Indian libraries.

2.6.2 Users' Interest Classification

The ultimate function of a lending library is to serve its users to their satisfaction. A classification is a tool to manage a library. By definition it is an arrangement of information material in a way useful to the majority of the users. In other words it is a rational sequence of maximum utility. Convenience of library users is a weighty consideration, if not the overriding one in a classification. It is always advised to put a book at the most useful place. Practical utility must govern all arrangements, feel many librarians and classifiers.

But this is a utopian thought, or an ivory tower theory. In fact this logical or systematic arrangement is forced on the library users. We have assumed without much research that shelf arrangement is useful, meets the needs of, and matches with the habits of the users. Certainly, it is based on the theory of one-size-fits-all. Here we are certainly living in our make - believe world that the inverted Baconian arrangement or its like are useful and logical.

In fact different user communicates and organisations need different arrangements transcending the traditional division by orthodox disciplines. A user oriented library classification need not be overly logical. The essence of reader interest arrangement is to classify material in a bold and utilitarian way which cuts across traditional groupings. The aim is to group library materials the ways which coincide with the user's thinking, interest, activity and needs. It is not a cognitive classification of a field of knowledge, but a contrived arrangement to serve local needs. A group of housewives may prefer all reading and information material on domestic chores such as child rearing, cookery, home-remedies, laundry work, interior decoration, pet care and personal body care arranged together, close by. In a junior college the students of commerce may need at one place books and other documents on elementary economics, accountancy, office management and mercantile law. In such situations the librarian may well adopt a shelf arrangement influenced by curriculum than by logic. It is called utilitarian arrangement and is not any negation of classification.

For this, we can have a broken order. If using the DDC we can place 400 languages together with 800 Literature. Similarly 320 Political science may be followed by 350 Public Administration. We may use some artificial digit in the call numbers to show this contrived proximity. Some libraries adapt their classification in such ways. If supported by adequate shelf guides it has high potential of satisfying users' needs. Arthur Maltby is of the opinion that the user interest arrangement needs more attention and research. It may provide a real step forward a in shelf arrangement which is waiting for an innovation for long.

2.6.3 General Classifications

A classification of the entire universe of knowledge is known as general classification. These are also known as universal classifications. Dewey Decimal classification (DDC), Universal Decimal Classification (UDC) Colon Classification (CC) and Library of Congress Classification (LCC) are some outstanding examples of this category.

By Levels of Details

Universal classifications are further of two kinds known as full and abridged editions, depending upon their level of details and use in a kind of library.

Full and Abridged Editions

These are the standard editions having full details and generally aim at large general libraries, say a university library, or large public library. Historically speaking till 1980s the UDC was available in three versions of details, namely, Full, Medium and Abridged editions having about 2,30,000, 70,000 (70% of the full), and 20,000 (10%) entries respectively. Abridged edition was meant for small libraries whereas Medium edition was adequate for general libraries. Full edition was available in many small fascicules which were meant for highly special libraries. At present UDC has two official versions: the Standard version of 70,000 entries and the Pocket/Abridged edition meant for teaching and shelf arrangement of small collection in libraries. Similarly, the DDC is also available in full version of four volumes and one volume abridgement. The latter is now in 15th edition (2012). The Abridged Dewey Decimal Classification is useful for a small library of about 20,000 titles. This simplified edition is a good model of a broader classification. It is quite popular in small public and school libraries. This trend of varied versions goes back to the Expansive Classification (1893) by C.A. Cutter (1837-1903) who planned to design his system in a series of seven schedules of successive increasing details. The first version was suitable for, say village libraries, the final version was meant for large libraries of the magnitude of national libraries having huge collection in all areas of knowledge. Abridged versions being comparatively inexpensive are also popular in developing countries. Abridged DDC is also very useful for teaching. It may also be used in conjunction with some special classification to cover remaining general areas of knowledge.

2.7 BY FORM OF LITERATURE

Apart from subject specialisation, special forms of documents such as, official reports, patents, standards, maps, CDs and videos are arranged by different methods. Some official documents such as reports, patents and standards bear some special code number. These are arranged by that official number. Pamphlets are usually arranged by title. Popular fiction is mostly arranged alphabetically by author. There is also what is called user oriented system. These are local even *ad hoc* arrangements to place together all books at one place likely to be required by a single group of users. J. Mills has listed the following sections or form of documents which may require a different type of arrangement in a library:

1) Age and grade of the reader.

2) Books for short loan or those for reference.

3) Current and reserve stock.

4) Size of documents.

5) Other physical considerations, i.e. films, cassettes, CDs , etc.

6) Factual and imaginative literature.

7) Language of the documents.

8) Documents of temporary significance.

9) Value of documents, like manuscripts, rare materials, special editions, etc.

10) Form of presentation, like bound periodicals etc.

11) Date of printing, incunabula.

12) Documents for abnormal readers, such as the blind.

2.8 PRINT AND ELECTRONIC VERSIONS

Traditional classifications are not only being used for classifying electronic and web documents, the IT itself has been used as a tool and medium for designing and publishing classification systems. As a result many classification systems are now available in electronic form. The 21st edition (1996) of the DDC available on a CD was known as Dewey for Windows (DfW) which has many more useful features over and above the print version. The electronic and much enhanced version of the DDC-23 (2011) known as Web Dewey 2.0 is now available only through "OCLC Connexion" on the Internet to the licensed users. Abridged Web Dewey is also available in the same way. Electronic versions are updated monthly. Electronic version of the Library of Congress Classification is available on the www as Classification Web (www.loc.gov/cds/classesweb). The website includes both the LCC number and LC subject headings with links between many of the class numbers and their LCSH equivalent. It also displays co-relations among LCC, LCSH and WebDewey. On the extreme, Broad System of Ordering (BSO, rev. ed. 1991) is now available only as machine readable form on disk and on the web at (www.classbso.demon.co.uk). All these electronic versions are easy to use, versatile with many more features including expert systems for number building, and are easily updated. Unlike the print or CD versions web versions cannot be pirated or duplicated.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

2) Explain the special features of the electronic version of the DDC.

....

....

....

....

2.9 SUMMARY

In the universe of classifications there are three types of them: Classification as such of any entity which also includes the process and methods of classification. It is applicable to all phenomena, objects and entities whether abstract or concrete. We can classify all living things, human beings, mammals, dogs, languages, chemicals, fruits, plants, everything and anything.

The second category is, taxonomies, that is classification of animals and plants to study their evolution and group them in families. It is essential for simplification, understanding and for living with sanity and success. As human beings we are classifying every moment unconsciously. Classified abstract knowledge becomes knowledge classification which is a field of philosophers, educationalists and scientists. These range from Indian Vedas (1500 BC), Greek philosophers (Aristotle, Plato, 300 BC), Muslim scholars of Medieval ages to Conrad Gesner, Francis Bacon, and August Comte.

Third category is of library classifications which are applications or adoption of knowledge classification to classify and arrange books and other reading materials in libraries. For individualised arrangement of documents and to display the internal and external features of the documents (media, viewpoint, form and language) a library classification has some additional features over the knowledge classification. Modern library classifications which originated in late 19th century can be divided into various categories and types. Since the beginning we have traditional enumerative systems which are long systematic lists of past and present subjects along with their class numbers. Also known as "mark and park" systems, these are bit old fashioned and out of favour of the librarians. Though easy to use indeed these are not effective to classify and index dynamically growing multi dimensional knowledge. These have given way to faceted systems which first divide knowledge into traditional main classes and then each main class into various categories, facets and other aspects of the subjects. These facets are combined to tailor the fitting class number for the specific subject of the book. There is nothing readymade. Some advanced faceted systems also termed as analytico-synthetic classifications are quite useful for depth classification and indexing for information retrieval. S R Ranganathan is the father of faceted systems which have been much improved by his disciples and the Classification Research Group, London. By all account faceted system are basis of all information retrieval and have a bright future in knowledge organisation. But to organise and retrieve information on the semantic web we need ontologies which are hierarchical systems showing deep hidden and multiple relations. General classifications are available in various versions of details such as full, medium and abridged editions to suit large and small libraries. General classifications such as the DDC, CC, LCC meant for general libraries cover entire universe of knowledge, while there are numerous special subject classifications delving into much more details suitable for depth classification and information retrieval in a specified area of knowledge. These range from a narrow subject, say Indian History to a multidisciplinary subject like Indology or Women Studies. However, Library of Congress Classification, spread over to 41 volumes, and also being available as an outline with sufficient details, serves both as general and special classification. S R Ranganathan was of the opinion that a general classification like the CC can serve as both. He compared his system to a trunk of an elephant which can pick a heavy log of wood and a light leaf with equal ease. But broad and depth classifications are relative, even subjective, terms. For some even the full versions of UDC and LCC may prove broader, and such libraries may resort to special or super special depth versions to serve their needs. These classification systems are available both in print and electronic versions, latter are now termed as vocabulary management systems. The electronic versions of DDC and LCC are much more enhanced, versatile, multi functional and remain always updated. For researching the semantic web ontologies have been developed which are classifications with myriads of deep links.

2.10 ANSWERS TO THE SELF CHECK EXERCISES

1) Enumerative classifications are old systems which provide readymade class

numbers. Their schedules are long and frozen. They usually fail to provide co-extensive class numbers for the subject of the document. They are also less hospitable to the new subjects. Faceted systems have slim schedules. Here subject of the document is first analysed into facets which are arranged in a citation order. Then a class number is constructed which fits the subject of the document. They provide solutions to the problems of enumerative systems. Faceted systems are also better for database design and retrieval in computerised databases.

2) www.oclc.org/dewey

2.11 KEYWORDS

Enumerative Classification : These are long and systematic lists of subjects of past and present along with their class numbers. Also called as 'Mark and Park Systems', these create pigeon holes for the subjects to fill in which usually prove to be square pegs in round holes and vice-versa.

Facet : A group of entities obtained by applying a single characteristic, e.g., kind of libraries facet in library science.

Faceted classification : A system in which a main class is first divided into various facets and the facets are combined in a specified order to tailor a class number according to the specific subject of the document.

Fixed location systems : The pre-Dewey systems which only indicated shelf ground of a book just like the house number in a street and sector of a city.

Knowledge classification : Systematic outline of knowledge known at a time to study its growth and structure. It is knowledge mapping.

Ontologies : Hierarchical and multi-relational classifications for the semantic web.

Special classification : Depth classification for limited and narrow area of knowledge used for information retrieval and classifications of micro subjects like theses, journal articles, patents, research reports, etc.

Taxonomy : Scientific classification of animals and plants.

2.12 REFERENCES AND FURTHER READING

Hunter, Eric. *Classification Made Simple.* 2nd ed. London: Ashgate, 2002. 1-23. Print.

King, Brand E and Kathy Rienold. *Finding the Concept not Just the Word: A Librarian's Guide to Ontologies and Semantics.* Oxford, UK: Chandos, 2008. 7-14. Print.

Maltby, Arthur. *Classification in the 1970s*. London: Clive Bingley, 1972. 19-23. Print.

Mills, J. *A Modern Outline of Library Classification*. Mumbai : Asia, 1962. 3. Print.

Ranganathan, S.R. *Prolegomena toLibrary Classification*. 3rd ed. Mumbai: Asia, 1967.Chaps CT-CY, 94-110. Print.

Satija, M.P. "Classification: An Essay in Terminology". *Knowledge Organization* 27.4(2000): 221-229. Print.

UNIT 3 POSTULATIONAL APPROACH

Structure

3.0 OBJECTIVES

After reading this Unit, you will be able to:

- discuss the evolution of the theory of library classification;
- describe and differentiate the descriptive and dynamic theories of classification;
- understand the work of designing library classifications in the different planes; and
- explain the applications of canons in these planes of work.

3.1 INTRODUCTION: HISTORICAL PERSPECTIVES

Modern library classification started in the late nineteenth century. Initially, classification systems such as DDC, UDC and the LCC were developed without much theory to guide the technical work. Classification design was the field of a few geniuses who were guided mostly by their intuition and flair. W.C.B. Sayers (1881-1960), a respected teacher in the University College, London derived theory of classification in the form of canons in his famous books. Other two pioneering names are of E.W. Hulme (1859-1951) of UK and E.C. Richardson (1860-1939) of USA. Their work on the theory of classification was mostly descriptive of the existing schemes. It was a descriptive phase of the theory of library classification. Then came H.E. Bliss (1870-1955), *a par* excellence philosopher and designer of classification. He spent his life in designing his classification entitled *Bibliographic Classification* (1940-1953). But prior to that he delved deep and long into the theoretical foundations of library classification based on the empirical foundations of knowledge and its organisation in his two famous books:

- *Organizing of Knowledge and System of Sciences (1929)*
- *Organization of Knowledge and Subject Approach to Books in Libraries (1933, 2nd ed 1939).*

Bliss formulated many canons and principles of classification to guide the designing of classifications. But that was a phase of descriptive theory.

3.2 POSTULATIONAL APPROACH

Postulational approach means going about the work of classification making by a pre-mediated theory in the form of Laws, Canons, Principles and Postulates. When Ranganathan designed his Colon Classification between (1924, 1928-1933), ironically he did it without any formulated theory. Indeed he had learnt some canons of classification in the classification classes of his most revered teacher W.C.B Sayers. After the publication of his CC in 1933 Ranganathan started thinking about the theory behind it that lay in his unconscious mind while designing the CC. That theory was so comprehensive and objective that it became theory of classification in general. He formulated objective even mechanical methods to design a classification system. Apart from the Five (Normative) Laws, he formulated 55 Canons, 22 Principles, 13 Postulates and 10 Devices for synthesis of class numbers and for evaluation of classification systems. For this he divided the whole work into three planes of work, namely, Idea, Verbal plane and Notational plane. He neatly divided the work to be carried out in each plane. Ranganthan thus liberated classification design from the elusive flair and intuition, and raised it to the status of a science.

3.3 IDEA PLANE

Idea plane is a plane of foremost importance where core intellectual work is done. Mostly the work here pertains to the choice of the model, defining the subject, its scope and sources of terms and concepts. Core classificatory work is the choice of characteristics and order of their application to produce categories of concepts called facets and isolates.

Further, it is to arrange the isolates into arrays and chains. Ranganathan has given following five sets of canons for the work in the Idea Plane.

1) Canons of Characteristics

2) Canons for Succession of characteristics.
3) Canons for Formation of Arrays of Classes.
4) Canons for Formulation of Chains of Classes
5) Canons for Filiatory Sequence.

3.3.1 Canons of Characteristics

The terms are collected on small slips from the identified standard sources of the subject. This mass of terms is broken into smaller groups called facets of the subject by choosing suitable characteristics. A characteristic is an attribute to divide a group into smaller group. It is a sort of hammer. It is a basis of division. The choice of characteristics, amongst many attributes, of an entity is vital and momentous. The ultimate quality of the final classification will depend very much on the chosen characteristics. The three Canons of Characteristics are:

- Canon of Differentiation
- Canon of Ascertainability
- Canon of Permanence

The canon of differentiation means that the characteristics that we chose should be able to divide the group into at least two sub-groups. A group of entities cannot be further subdivided or differentiated on the basis of a characteristics common to all (Canon of Differentiation). For example, bicycles cannot be divided on the basis of number of wheels, as all of them have two wheels. But they can be differentiating into subgroups on the basis of kind of rider like gent's bicycles, ladies bicycles, and children bicycles. Further, the differentiated characteristics that we chose should not itself be unverifiable or uncertain (Canon of Ascertainability). For example, a group of living people cannot be divided on the basis of their previous birth, or the date of their death. Date of death of living persons is uncertain, so is our previous birth. But living people may be divided on the basis of gender or nationality or colour of their skin; all these are ascertainable characteristics. Further the chosen characteristics should be of permanent nature, not transitory or changing (Canon of Permanence). Chameleons cannot be grouped on the basis of their colour as it is changeable. Political parties hopper Ganga Charan Rajput cannot be classified on the basis of his party. People should not be classified on the basis of the colour of their clothes, as these can easily change. Canon of Relevance means the division should be relevant to the purpose of classification. For example, a class of foreign language learners should not be divided on the basis of skin colour or gender or body weight or height, as these have no relation with the language learning ability of a person. Similarly, a class of wrestlers or boxers can be divided by age or weight, but not on the basis of their religion, mother tongue or colour of skin.

There may be too many relevant characteristics. Their choice must be on the basis of being most helpful for the purpose of ultimate classification. This question is further linked to the users of classification and their needs. Ranganathan warns that it is not easy to determine the ultimate purpose of classification.

3.3.2 Canons for Succession of Characteristics

There may be many relevant and permanent characteristics for dividing a group. All may be needed to divide and subdivide a group. But the order in which these are applied one after the other is very important for the kind, quality and ultimate value of classification. Therefore, this set of canons is concerned with the sequence in which the various characteristics are applied. These are:

- Canon of Concomitance
- Canon of Relevant Succession
- Canon of Consistent Succession

Concomitant means happening at the same time. Therefore, this canon means that two characteristics applied to a class should not produce the same sub-classes. For example, (in 2009) a class can be divided into two groups by applying the characteristics of below 20 years and above 20 years of age. Then to apply the characteristics born in or before1989, and born after 1989 will result in the same grouping. The 20 years of age or 1989 born are concomitant characteristics. Similarly, if the first characteristic is the bird, then we should not apply the characteristics of having feathers, as all birds have feathers, and it will again produce the same group of animals. It also means that the characteristics applied should be in order of broader to narrower – narrower to broader will not work. For example, a group of people can be divided on the basis of Brahmins and Non-Brahmins. There will be two groups. But to apply the characteristics of "Hindus" to Brahmins will make no further divisions as all Brahmins are Hindus. *Relevant succession* means that the sequence in which these characteristics are applied should be relevant to the purpose. To cite an example, literature has four relevant characteristics, namely, Language, Form, Period /Author, and Name of the literary work. These characteristics can be applied in different sequences:

- Literature – Language – Form – Author – Work
- Literature – Form – Language – Author – Work
- Literature – Period – Form – Language – Work

All of them are different classifications. The suitability of any order of citation will depend upon the purpose of classification. But which sequence should be chosen? It depends upon the need of library users. One sequence of facets may not be useful to all, as library users have individual needs. Scholars who are interested in one language literature may prefer the first order, while scholars who are interested in poetry irrespective of the language may prefer the second sequence, and the third will be useful to historians of literature. Similarly, in a library devoted to area studies the first division should be by area rather than the subject. Lastly, the Canon of Consistent Succession only advises that the order of application of characteristics once chosen should be followed consistently, until unless the purpose of classification itself changes.

As a result of successive applications of characteristics we obtain numerous individual concepts. These concepts have to be arranged in what is called Arrays and Chains.

3.4 CANONS FOR ARRAYS

An array is a long line of entities of equal rank arranged in some systematic order. For example, all the children of a father make an array. In the same vein, all the continents of earth make an array. States of India make an the array, and further all the district towns of a state make another array. Ranganathan has prescribed the following Canons for their formation:

- Canon of Exhaustiveness
- Canon of Exclusiveness
- Canon of Helpful Sequence
- Canon of Consistent Sequence

Exhaustiveness means, an array should be all inclusive of its eligible members. See the following array of men by colour:

- Whites
- Blacks
- Brown

This array is not exhaustive as it does not include other colours such as pale yellow, fair, wheatish. Hence while forming an array every member should be included, otherwise classification will not be comprehensive. On the other hand, Exclusiveness means that a entity should belong to one and only one array; in other words a member should not be included into two groups at the same time. For example, in classification of dogs either they should be kept under mammals or under pets, not under both, as it will result in what is called cross classification. But in computerised databases and OPACs cross classification is a boon as it provides an extra access points and increases the probability of retrieval.

Helpful sequence demands that entities should be arranged in some logical predictable and helpful order. For example, all the children of a father can be arranged by age. As another example, all the students of a class can be arranged by merit or alphabetically by name. There are numerous ways for the systematic and helpful arrangement of entities in an array. Lastly, consistent sequence means that if a set of entities occurs at different places then their arrangement should be the same everywhere. For example, the terms male, female and child occur in main classes Psychology, Education, and even Law. Their sequence should be the same in all these classes. As another example, names of countries occur in the main classes Geography and History. According to this canon the sequence of countries in the above main classes should be the same.

3.4.1 Principles of Helpful Sequence

As already said, an array is essentially a systematically ranked and arranged group of equal entities. There are many ways the members of a group may be arranged. Librarians have to chose their sequence which is helpful to the majority of the users and also logical. Entities in an array may be arranged in a chronological or historical sequence. A group of boys and girls may be arranged by age; Kings of a country may be arranged according to their period of rule. Indian Prime ministers may be arranged in the order: Nehru, Shashtri, Indira Gandhi, Morarji Desai… Vajpayee, Manmohan Singh. Extending this analogy a bit, a queue waiting for a bus or before a booking window is also according to this order. It is an order which may be called “First come-first-served”. Related principle is of “Later in Evolution.” Some entities can be arranged as they have evolved: animals can be arranged from amoeba to mammals; plants are arranged from Thallophyta to Dicotyledons. Society can be arranged: Hunting society, Agriculture Society, Industrial Society, and Information society. Another Principle is of geographical proximity or Spatial Contiguity. If the entities exist in space their arrangement should be near to one another. We can arrange Indian states in the order: J & K, Punjab Himachal, Haryana, Delhi, UP, and so on. Planets in space may be arranged like: Mercury, Venus, Mars, Jupiter, and Pluto. Principle of Quantitative Measure means that if entities are associated with some quantity, then these may be arranged in the order of their increasing quantity. For example, Indian currency notes may be arranged in the order: one rupee, two rupee, five rupee, twenty rupee, fifty rupee, hundred rupee, five hundred rupee and one thousand rupee.

In Town Planning we can arrange like: Village Planning, Town Planning, City Planning, and Metropolitan Planning. Principle of Increasing Complexity means the entities may be arranged in the order of their increasing complexity, e.g., Linguistic elements can be arranged as: syllable, word, phrase, clause, sentence, paragraph, and so on. Algebraic equations can be arranged as 1st order, 2nd order, 3rd order equations, etc. Further, entities can be arranged according to their popularity of use, e.g., we can arrange food seeds in the order: Rice, Wheat, and Rye. Further if there is any traditional sequence of entities then it may be followed as we traditionally say Radhe-Shaam, Sita-Ram, Algebra-Geometry and so on. We never say Ram-Sita. It is known as Principle of Canonical Sequence. Lastly if no other principle applies then the entities may be arranged in alphabetical order. All the UN member states are arranged in alphabetical order. Long list of names of persons is better arranged in alphabetical order. Words in a dictionary are arranged in alphabetical order – also called dictionary order. Alphabetically arranged entities can be easily located. Ranganathan advices to use this as a least preferred method. He even thought it as opposite of classification, yet many systems prefer it for ease of its use, and later for easy operation. In nutshell, for arrangement of members in an array use any systematic predictable method that you think will be helpful to majority of the library users.

3.5 CANONS FOR CHAIN OF CLASSES

As a result of successive application of characteristics to a group not all members will be of equal rank. This could be wholes, their parts, kinds and various species and subspecies of an entity. Such whole and their parts should be kept together to form a chain of classes. A Chain is a sequence of entities in successive subordination. For example, Grandfather, father, sons, grandchildren make a chain. There are two canons for arranging entities in a chain:

- Canon of Decreasing Extension
- Canon of Modulation

The first canon means that the entities should be arranged in a broader to narrower or general to specific, or whole to parts order. For example, Asia, South Asia, India, North India, Punjab, Amritsar makes a chain of classes in decreasing extension. Social Sciences – Economics – Financial Economic – Money – Banks make another chain of broader to narrow classes. Modulation means that no intervening link should be missed in the classes arranged in decreasing extensions: no jumping; no snapping. In the first example, we should not directly jump from North India to Amritsar omitting Punjab. Though this snapped chain will satisfy the canon of decreasing extension, but will violate the canon of modulation.

In this way in the Idea plane we will have a network of discrete facets and isolates arranged and laid out in arrays and chain. But these are still concepts without specific and **pucca** names. An idea without a **pucca** name cannot be classified properly.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

1) State the Principles of Helpful Sequence in an Array.

....

....

....

....

3.6 VERBAL PLANE

In this phase of classification work we give standardised names to the concepts arranged in arrays and chains in the Idea plane. Ideas are something airy in the mind. They cannot be well expressed and communicated without having a proper name. In this phase we dress the airy ideas in proper and visible garments of language to make them standard terms to handle. It is important as ideas need some media for expression and communication. It is rightly said an idea in words is worth thousands in the mind. No idea can be communicated without apt words. In the field of sciences and all other technical subjects like accountancy or literary criticism these names for concepts are known as "terms". And science of assigning terms to ideas is called terminology. Verbal plane deals with terminology.

3.6.1 Nature of Language

No doubt language is a vehicle of thought; without it there could be no civilisation, no progress, perhaps no art or science. Without a developed language we will not be able to think even. Without language there will be only animal living. But any language spoken by any community is both rich and poor. It is rich in the sense that usually it has many words for a single concept, e.g., wages, pay, salary mean the same thing. Similarly, in Indian languages Suraj, Surya, Shashi, Dinkar, Prabhakar, Aditya all mean the sun. Lord Krishna has many names from Gopal to Nandlal, Kanhiya to Sham. Different words having the same meaning are called Synonyms. These are synonymous words.

It is poor in the sense that one word may denote more than one meaning : cricket is a popular game and also an insect. Bridge is a construction over river, and also the name of game. Beas is a river and also a town in Punjab. A word having more than one meaning is a called a homonym. In English there is hardly a word having only one meaning. It is said that the word "order" has 250 meanings in the English language. Ranganathan finds five meanings of the word "classification". Homonyms and synonyms are vitally useful for the literary writers. They make the writings delicious, imaginative and rich with aesthetic pleasure.

3.6.2 Homonyms and Synonyms

But synonyms and homonyms are a great barrier in communication; even harmful for legal and business transactions, science and research communication. In science communication we should be precise to the extent that one word should have one and only one meaning and vice versa. It means an ideal situation of one to one correspondence between concepts and terms. That is language of any science and academic field of study and research should be strictly free of homonyms and synonyms. It has already been achieved to a great extent in natural sciences. Social sciences, including library and information science, are struggling for the standardisation of terminologies in their subjects. Maturity of a discipline of knowledge can easily be measured by its standardised terminology. S.R. Ranganathan always urged librarians to learn and use technical terminology of library science for research and communication in our field. He also advised librarians to learn the technical terminology of other subjects for better information retrieval and to provide satisfactory reference service to scholars of various subjects. Ranganathan himself has developed good number of terms across the length and breadth of library science. It was his method and style to define the terms adequately in the beginning of his research articles. He always used well defined technical terms in all his writings and has left a treasure trove of technical terms as a legacy for us.

Language, including technical language, being a living and dynamic entity, keeps changing in many ways. New terms are coined to convey new concepts and theories, some old terms are deleted as they become obsolete with time and some of the terms even change their meaning. In our discipline old terms such as open access, browsing, networks, etc have acquired different meanings in the electronic environment.

3.6.3 Canons for Terminology

Ranganathan not only himself coined many a new terms, he also formulated many principles for coining and display of new terms. As said earlier, the first and foremost quality of any terminology is to be free of homonyms and synonyms. Following are the canons he formulated for the purpose:

- Canon of Context
- Canon of Enumeration
- Canon of Currency
- Canon of Reticence

Canon of Context lays down that the terms in the classification schedules be written and read in the context of the upper class for example, instead of writing

Simple salts

Double salts

Complex salts

We should write:

Salts

  Single

  Double

  Complex

Similarly in Psychology under the persons facet:

- Child
- Male
- Female

mean Child psychology, Male psychology, Female psychology respectively. For the brevity and simplicity of the schedules it is not necessary to repeat the upper link. This canon is meant both for the classificationist and the classifier. It also applies in our daily routines.

Canon of Enumeration: Scope of terms or disciplines is not universally settled. For example, subject "Dynamics" is considered a part of mathematics by some, and of physics by others. As another example "Documentation" is part of library science in Colon Classification whole in the Universal Decimal classification (UDC) it is an independent class. Further constitutional history is not part of political science or law but of history in the CC. In such cases no one is wrong – as it is debatable. In such circumstances it is safe to delineate the scope by listing i.e. enumerating its subdivisions. It is a operational and pragmatic way to define the scope of a subjects. This is the advise of this canon.

Canon of Currency: As already said terminology in any discipline is not static or frozen. It is always dynamic and moving. Even the old concepts get new meanings and an old concept may get a new term. Old terms may denote new concepts. A classification system must always update its terminology using current terms. New editions of classification systems not only include new knowledge but also use current terminology. For example, in the DDC library science has become "Library and information science"; Home science has become "Home management". It will not only help in efficient information retrieval but will also make our classification systems respectable in the eyes of scholars and subject specialists. This canon lays down: "Terms used to denote a class in a scheme for classification should be one current among those specializing in the subject field covered by the scheme". Lastly the Canon of Reticence means that the terms that we use in the schedules of classification should neither be judgemental nor critical. The DDC 14th edition used the term "humbugs" for parapsychologists. For example, one should not use the terms major or minor authors while dividing literary writers. It is not proper for librarians to categories authors so or be contemptuous of a subject believed by others. Literary historians can describe authors so. Even then they could be on slippery grounds. The literary reputation of an author is never constant. It is changing even after his/her death. Hence a classification system should use current and neutral terms.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

1) Briefly explain the work in the Verbal Plane.

....

....

....

....

3.7 NOTATIONAL PLANE

Though the last plane of work in designing a library classification system, yet it is the most visible face of the final product. Being visible, some people mistake it for classification itself. In fact notation is a series of symbols or codes to represent the subjects which were concepts and names in the idea and verbal planes, respectively. It is a plane which faithfully implements the findings or decisions of the idea plane. Ranganathan called it as servant of the idea plane. It is better to call it as the Executive Plane. Its position can be better understood by using an analogy of motion picture. In a movie, story and plot make the idea plane, screenplay and dialogue make the verbal plane, while actors are the notational plane. The function of notation in library classification is clear, but its status is ambiguous, at best. It is rightly said that a bad notation may mar a good work in the idea plane, but a good notation cannot improve the bad work of the idea plane. (Idea plane may be treated conceptual as classification).

3.7.1 Definition

Notation is a system of short hand symbols to denote subjects and their subdivisions by ordinal digits. A digit is an element of notational systems. These digits have only ordinal value i.e. they show only order. These digits are not cardinal or quantitative. To explain, in a notational system comprising of 1,2,3,…9 etc., it means value of 2 is not greater than1, but will only come after it. Similarly, if it comprises of symbol A,B….Z, it means C is not greater than A or B, but will fall between B and D. That is notation conveys only the order not value or weight.

3.7.2 Need and Purpose

In the Idea and Verbal planes every concept is given a due place in a classification system, but these terms howsoever, cannot be assigned as index terms to the documents in a library. This is mostly for the following:

- Names/terms for subjects in the verbal plane are too long to be written as labels. Hence we require short symbols to denote subjects.
- Names of subjects are different in different languages whereas in a library we have books in many languages. Mathematics is called "Ganit" in Hindi-then where should be place a book of mathematics in Hindi? : under "G" or "M".?
- Names keep changing. For example Economics was once called Political Economy. Education Ministry is now called Ministry of Human Resources Development. Sri Lanka was once called Ceylon.
- If names are used for arrangement of documents then there would be many problems. The subjects will get scattered, e.g. Mathematics will come under "M", whereas Algebra will go to Geometry to "G" and Trigonometry to "T". A family will thus get dispersed.
- Even if these names are labeled on the documents how these will preserve the systematic order of subjects (in arrays and chains) arrived in the idea plane.? Words in themselves can only be arranged alphabetically. Therefore, to preserve the decided sequence of subjects and their subdivisions a series of ordinal symbols are assigned to the concepts and terms. For example, in botany the arrangement of various parts of plants decided in the idea plane is

Plant	I
Root	I,3
Stem	I,4
Leaf	I,5
Flower	I,6
Fruit	I,7
Seed	I,8

(The above arrangement is according to the principle of spatial contiguity – we proceed part by part from bottom to top. This sequence also conforms to the principle of "Later in Time"). We have assigned a number to each subdivision to preserve their arrangement and use it mechanically at some later time. This symbols-complex, called class numbers, are not only short but will also help to shelve or arrange document at proper places. These class numbered documents can be taken out, read and placed again at their proper places without any difficulty. Notation offers a self-evident order. Obviously, it is much easier to use these shorthand symbols for arranging documents than to use the names of subjects.

3.7.3 Other Uses

- It is an indispensable component of library classification – which is usually not required in knowledge classifications.
- Apart from arranging documents on the shelves it arranges entries in classified catalogues and shelf lists.
- Shows relations of subjects in the over all scheme of the mapping of knowledge.
- In a faceted classification it makes the structure of the subject transparent.
- It is essential for Chain Indexing. (You will learn about Chain Indexing in the unit on cataloguing).
- It may be essential for arrangement of (books) circulation record in a library.

Notation is so essential to classification that Palmer and Wells have defined library classification as “representation of an infinite series of subjects by a finite series of symbols”. It is aptly said by W.H. Philips that if classification is foundation study of librarianship, then notation is the basis of practical book classification.”

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

1) Describe the need and purpose of notation.

....

....

....

....

3.8 CANONS OF NOTATION

Now we pass on to the *how* of notation. To design a qualitative and effective notation Ranganathan has formulated and prescribed some canons for the choice and design of a notational system. These are:

1) Canon of Homonym and Synonym
2) Canon of Relativity and Uniformity
3) Canon of Hierarchy
4) Canon of Mixed Notation
5) Canon of Faceted Notation
6) Canon of Co-extensiveness

Ranganathan though described notation as servant of the idea plane, yet it is more than a mere series of symbols. He expected much from notation. He always treated it as a device to translate the subject of a document into an artificial language of ordinal numbers. And he wanted to endow it with qualities of a language. In the same vein his Canon of Homonyms and Synonyms prescribes that a class number should denote one and only one subject, and conversely a subject should be denoted by one and only one class number. In fact, it is too much to expect from notation which is no more than an ordering device. Ranganathan has visualised only an ideal and perfect notation which is far from reality and much advanced of its present needs. Even half of its envisioned efficacy will be achieved at an unaffordable price in terms of complexity. It is possible to denote a subject by one and only one class number, but in reality a class number denotes a group of subjects. For example, in the DDC 610 always means Medical Science and 611 always means Human Anatomy. On the other hand Bengali language has only one number 491.44, but 491.49 Other Indian languages stands for Awadhi, Bagheli, Chhatisgarhi, Eastern Hindi, Kafiri and Pahari. So many languages share one class number. A classification like Rider's International Classification (1961) which always uses three alphabets to denote a subject is likely to be full of homonyms –a class number denoting many subjects at a time.

Any classification at best makes broader groups than to faithfully and precisely translate the subject of the document into notation. In fact, more homonymous a class numbers more hospital a classification system. Also many classification system such as Bibliographic Classification (2nd ed, BC-2), even the DDC offer many alternatives (optional class number). For example, in BC-2 Religion can either be placed at P or Z. therefore, this canon is difficult to be observed in reality – even Ranganathan CC violates this canon at many places. Canon of Relativity and Uniformity means that length of a class number (i.e. total number of digits in it) is usually the indication of the breadth and depth of the subject, it denotes. Breath and depth are technically known as extension and intension of the subject respectively. For example:

02	Library Science
025	Library Operations
025.3	Bibliographic analysis and control
025.32	Descriptive cataloging
025.322	Choice of entry
025.3222	Authority files

In the above example as the subject becomes more and more specialised the number (quantity) of digits goes on increasing. It also means that the hierarchy of a subject is depicted through the increasing length of notation. Theoretically, it is quite possible to go down the hierarchy to a great depth. In practical library classification this lengthening of chain cannot go on for ever. A practical library classification has to stop somewhere to keep the length of the class number in check. Therefore, for brevity and simplicity some classifications do not show the hierarchy of notation beyond some point, though order of subjects is maintained. In the National Library of Medicine Classification (USA) we have,

FAI	Great Britain
FE5	England
FG9	Guernsey

Here three subjects of decreasing extension (and increasing intension) are denoted by equal number of digits, though their order on shelf will be from general to specific or broader to narrower. Canon of Hierarchy is implied in the Canon of Relativity. It means that every characteristic used in the division of a universe of entities must be represented by a digit. In others words the class number must and faithfully depict the deepening hierarchy of subjects, e.g.

5	Sciences
51	Mathematics
516	Geometry
516.3	Analytic geometry
516.35	Algebraic geometry
516.352	Theory of curves

Many a scheme violates this canon to secure brief numbers. We do not think the depiction of hierarchy through notation is useful beyond a point.

Canon of Mixed Notation: Usually there are two types of notations: Pure and Mixed. A pure notation is the one comprising of single species of digits, say only A/Z as in RIC, or 0/9 Arabic numerals as in the DDC. There was a time when purity was considered a virtue due to ease of use. Mixed notation comprises of a mix of two or more species as in Library of Congress Classification which uses mix of alphabets and numerals to denote subjects. UDC which mainly uses 0/9 notation with host of punctuation marks and mathematical signs has also a mixed notation. In the present times a classification has a very difficult task of mapping and structuring the expansive, complex and turbulently growing multidimensional universe of knowledge. A pure notation cannot work effectively. Hence classifications, of necessity, have resorted to mixed notation. It is rightly said that many of the problems of the DDC are due to its pure notation. A mixed notation has a wider base that is has more number of digits in it. A wider base is able to give shorter class numbers. For example, a notation of Indo-Arabic numerals 0/9 is able to produce one thousand 001 to 999 three digited class numbers, while a system using A/Z notation will have (26^3) 17576 subjects using three digits. Though a mixed notation is bit difficult to handle, yet it is a necessary evil. Then the question is how much mixed? Moderately mixed, as in BC-2 or Library of Congress is ideal. But highly mixed notation with many unfamiliar symbols may be disastrous for library classification. If some problems of the DDC are due to its purity of notation, on the other hand highly mixed notation comprising of 74 digits in Ranganathan's has played a role in its destruction and unpopularity.

Decimal Fractions and Arithmetical Numerals: Further, Indo-Arabic numerals may be used arithmetically or as decimal fractions. Library of Congress uses them arithmetically while all other systems, including the CC, use Arabic numerals as decimal fractions. The DDC pioneered the use of decimal notation and is now considered naturally convenient to denote subjects and their subdivisions. Decimal notation has many obvious advantages over the arithmetical numbers.

Faceted Notations: Notation must be structurally transparent to show various facets or elements of a class number. For example, in UDC, 82 Shak-2 denotes plays of Shakespeare. Here 8 is literature while 2 is English literature, Shak stands for Shakespeare, and -2 is drama. Even in the DDC 822.33 means the same thing. It is structured internally, as 8 is literature, 22 English drama, 822.3 is Elizabethan plays, 822.33 is Shakespeare. But in this case though the notation is faceted or structured but it is not transparent. Every hierarchical classification is structured but may not be transparent, whereas the faceted notation of CC is both structured and crystal transparent, e.g. O111, 2J64, H means Literature-English-Drama-Shakespeare-Hamlet.

Canon of Co-extensiveness

It means that every aspect of the subject should be indicated by a digit. That is a class number should be totally comprehensive of the characteristics used in dividing a subject. It again brings us to the question of hierarchy and relativity :

Indian History	954
Mughal History	954.02
Akbar	954.02
Court of Akbar	954.02

This classification violates the Canon of Co-extensiveness. So is the case with the following numbers from the Rider's IC:

Diseases of Stomach	UJK
Gastritis	UJK
Gastric Disorders	UJK
Gastric Ulcer	UJK

In the above two examples the increasing depth (intension) of the subject has not been represented by correspondingly lengthening the class number. It means 954.02 and UJK are homonymous class numbers. Ultimately, it results in broader classification. Some classification thinkers arguably believe that it is superfluous to aspire or try for co-extensive class numbers. Moderately depth classification is quite sufficient for shelf arrangement. H.E Bliss (1870-1955) is said to have said "Be minute, be minute, be not too minute". To this Ranganathan replied "Be minute, be minute, be too minute". Debate is endless. It again brings us to the question: Is our library classification sophisticated enough to carry the entire burden thrust on it? We must not expect from library classification, especially its notation, what it inherently is not capable of.

3.8.1 Qualities of Notation

Apart from all the above mandatory qualities (as Ranganathan terms them as canons) other qualities of notation can be of three types:

1) Optional or desirable

2) Essential and vital

3) Scientific

Above all the notation should be user friendly. Among the desirable qualities, a notation should be brief and not highly mixed, so that digits convey a self-evident order. For example, order of mathematical symbols or punctuation marks is not self evident. Digits should be easy to write, in fact should be available on the computer key board. Class number should be easy to pronounce and remember for a short while. Brevity of class numbers in the print environment was a necessity as a class number had to be written on the book spine, which has a very limited breadth. Lengthy class numbers also pose difficulties in arrangement on the shelves. But in the OPACs the length of a class number does not matter. Hierarchical and faceted notation with highly recall and relevant ratio is best for information retrieval. Mixedness and length of notation do not matter in an automated library whereas in a print or manual library these are weighty and influential considerations.

How of Brevity: Proportionate or equitable allocation of digits to subjects will result in brief numbers. It means that static subjects like philosophy or religion should be allocated a small slice of the notational cake whereas dynamic subjects like science and technology, computers should be given a larger slice. As said earlier, wider base of mixed notation will also turn out brief numbers. Above all broader classification results in brief numbers.

Mnemonics

Another desirable quality of notation is it being mnemonic. A mnemical notation denotes same or similar recurring concept by the same digits. The DDC and the CC are highly mnemonic systems.

Language	Literature	Linguistics	History
English	820	420	942
German	830	430	943
French	840	440	944
Hindi	891.43	491.43	-

English language, Literature and History are always denoted by "2" and German by "3", so on. Similarly, take the case of CC:

Main class	Anatomy	Physiology	Diseases
G Biology	G :2	G :3	G : 4
I Botany	I : 2	I : 3	I : 4
K Zoology	K : 2	K : 3	K : 4
L Medicine	L : 2	L : 3	L : 4

Anatomy wherever it occurs has been denoted by "2" and disease by "4". Ranganathan identifies three kinds of mnemonics, namely, alphabetical, schedule and seminal. Schedule mnemonics have been explained above. In alphabetical mnemonics an entity is denote by its name using its initial alphabet, e.g., J381B means Basmati Rice, while D5125H means Hero Bicycle, D5133M means Maruti Car. Library of Congress, (LCC) and UDC use alphabetical mnemonics to a large extent. In the LCC we have:

A General works

AC General Collections

AE General Encyclopedias

AS General Societies

Seminal mnemonics means to denote a concept by its inherent number: 1 denotes unity, God, Parents, World; 2 denotes two dimensions, anatomy, constitution. Further, Social pathology, Torts and Diseases will be denoted by 4.

As another form of seminal mnemonics, seminally equivalent entities, e.g., feed, food, fuel, should be denoted by same digit wherever they occur. Repair of machines, treatment of diseases, and alleviation of social ills also get the same number in respective main classes. Mnemonics though they bring sort of symmetry in the classification yet at many places they may conflict with the helpful sequence. Nevertheless, mnemonics are optional.

3.9 HOSPITALITY IN ARRAY

Hospitality is the most essential, rather vital, quality of any notational system. It is the capacity of notation to accommodate new subjects at their proper places without disturbing the existing sequence. Non-technically, it is also known as flexibility or resilience of classification. It is essential as knowledge is simultaneously growing exponentially in multiple directions. Therefore, any living and practical classification must have the capacity to give place to the new subjects at their *proper places* – latter point is essential. That is why library classifications are revised from time to time to include new subjects at their natural places. For example, the DDC (1876) of 44 pages has grown gradually to more than 3000 pages in the DDC-22(2003). Hospitality can be at two levels.

1) Hospitality by classifiers

2) Hospitality by classificationists in new editions.

3.9.1 Hospitality by Classifiers

Almost all standard library classification systems recommend that the classifiers should not tinker with the schedules. They should not make local numbers in case a number for a new subject is not available in the classification. Usually in such situations it is advised that a classifier should wait for the new edition and temporarily may place the new subject with its broader class. But Ranganathan has made provision for the classifiers to synthesise a class number for a topic not explicitly listed in the schedules. These are devises for hospitality and number building by the classifiers. These are namely:

Subject Device

Chronological Device

Geographical Device

Super-imposition Device

Alphabetical Device

This armory of devices at the disposal of a classifier keeps the work of classification going. Using these devices judiciously a classifier can make class numbers for new subjects. Perhaps on the dint of these devices Ranganathan claimed his system to be "self-perpetuating" – that is which is able to classify new knowledge without immediate intervention of the classificationist. Nevertheless, it is doubtful if a classification, howsoever hospitable, can be self-perpetuating. (A full discussion on the use of these devices is beyond the scope of this unit).

3.9.2 Hospitality by Classificationists in New Editions

Every system designer is aware of making provisions to properly accommodate new subjects at later stages. Therefore, in every system some conceptual, structural and mechanical provisions are made to accommodate new subjects at proper places (sometimes at not so proper places). Sophistication of a notational system can be measured by its instant hospitality to new subjects. Let us make a case study of Ranganathan's Colon Classification for hospitality where notation consisting of six species is highly mixed:

1	A/Z	Roman Caps	26
2	Δ	Greek Delta	01
3	0/9	Indo-Arabic Decimal number	10
4	a/z	(Excluding I,L,o) i.e.o	23
5	* ” ←	Indicator digits with anteriorising value	03
6	& ‘ . : ; , - = + ( ) →	Ordinary indicator digits	11=74

Ranganthan has resolved hospitality at two levels namely, in arrays and chains.

3.9.3 Gaps in Arrays

An array is a sequence of co-ordinate classes, Gap device a method to accommodate future subjects, is used at every level of arrays in almost all classifications. Gap device leaves some vacant numbers here and there to be filled in with future and unborn subjects. In the CC, e.g.,

U1 Mathematical Geography

U2 Physical Geography

U3 [Vacant]

U4 Anthropogeography

U5 Political Geography

U6 Economic Geography

U7 [Vacant]

U8 Travels

In the above array U2 and U7 are vacant positions which can be filled with new topics of Geography. The DDC uses this method to a great extent. In 500 in the Third Summary we have many vacant position in DDC -22 : 504, 517, 524, 544, 545, 574, 589. At lower levels there are numerous such vacancies. These can be used in the future to accommodate new subjects.

Limitations

Though used by almost every classification, gap device is not the real solution. This method does not ensure a rightful place for the new subjects. There may not be any vacant place available at a needed place; on the other hand many vacant places remain unfilled for long as no new subjects are emerging there. In the DDC many new subjects are misplaced as due to lack of space at the right place they are allotted a vacant place nearby. It distorts the structure or mapping of knowledge. There are no gaps left in classes like Technologies where new subjects are popping everyday. Therefore, gaps are not the solution but alibis to postpone the crisis.

3.9.4 Sectorising Digits

These are devices to accommodate a subject at a proper place in an array even if no vacant place is there. For this Ranganathan invented an ingenious method of empty digits, now also called Sectorising Digits. He sets aside 0,9,z and Z as empty digits. These digits are never used alone but used as repeater digits to extend an array. For example 1,2,3,…8, 91, 92… 99, 991, 992…993……999 are all co-ordinate classes. Similarly, we can have an array extended like A, B,C….Y ZA, ZB,ZC….ZY, ZZA…….ZZX and so on. Hero Z has no semantic value, only ordinal value. It means semantically it is empty. Though the DDC has no such provisions nor terms like this, it often uses "9 others" to dump en masse other remaining subjects which could not be accommodated in the array 1/8. A good example is the main class 900 History, Geography, Biography. The three subjects have been clubbed together as there is no other place in the decimal notation beyond 900.

3.9.5 Emptying Digits

Digits T, V and X are set aside as emptying digits as these empty a preceding digit of its meaning but allow it to retain its ordinal value. For example, in the CC

K Zoology

L Medicine

No space is left between the two to insert a new subject Animal husbandry. Using

X as an emptying digit KX is given to Animal husbandry. Here K no more denotes Zoology, and KX may be treated a single digit which means Animal husbandry and is arranged between K and L. Hence it is a very clever device to accommodate new subjects at proper place in an array. Similarly, we have

44 India

44T Nepal

Emptying digits are just like the King Bali in the Ramayana who was blessed with sacking the power of his enemies who faced him in any battle. That is why Lord Rama had to kill him while hiding himself in a bush – though then it was not the norm to kill an enemy from a hidden position.

3.9.6 Empty-Emptying Digits

To make further rather almost unlimited interpolation in the Colon Classification U, W and Y have been postulated as Empty-Empty digits. It means not only these digits are empty of any meaning these also make other digits empty to which these are attached, e.g., in the CC-7.

Y	Sociology	YYT	Sociometry
YT	Demography	YYU	Socio-Cybernetics
YUA	Cyber Culture	Z	Law
YUG	Bio-Sociology		
YX	Social Work		

In this way, any number of new co-ordinate subjects can be interpolated at their proper places in an array of classes.

3.10 HOSPITALITY IN CHAIN

Chain is a sequence of classes of successively decreasing extension. The DDC provides infinite hospitality in chain by the use of decimal fraction. New subjects can be added at the end of a chain by a decimal fraction. This method is now used almost by all the classifications. Use of decimal numerals is almost a norm while designing classification systems:

328	Legislative Process
328.3	Parliaments
328.33	Members of Parliament
328.334	Basis of Membership
328.3345	Election Constituencies
328.33455	Gerrymandering
328.334552	Reserve constituency*

* The last number has been added by us to show how the new subjects can be added by lengthening the chain. We can also give many such examples from the CC. Hierarchy showing relations of sub-ordination and co-ordination, and relativity are scientific and logical qualities of any notational system.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

2) Describe the Qualities of Notation.

....

....

....

....

3.11 PROBLEMS OF NOTATION

There cannot be any library classification without notation. For many library users notation is the library classification. In fact it is one of the three forms to designate and denote concepts, the other two being terms and definitions. But notation is only skin deep, and should not be over burdened with tasks. Apart from subject designation, it preserves and mechanises the chosen order of subjects and documents. Though indispensable it has many problems. In many systems notation has become too complex. Lengthy class numbers are inevitable in depth classification which pose many problems. They are not user friendly. The order of digits in a mixed notation may not be comprehensible to the ordinary library users. Not only this, in the UDC, punctuation marks are posing a problem in automatic arrangement as their value is not compatible with the ASCII. Late French classification thinker Eric de Grolier (1911-1998) regrets that we have not applied to library classification the progress the notation has made in other field such as mathematics, logic and chemistry. One hindrance is our library users who are laypersons and may not understand very advanced or sophisticated notation. Coming of digital libraries may provide a virgin field to realise the full potential of notation in library classification.

3.12 SUMMARY

Since the late nineteenth century we have reached a stage where we have developed a well rounded theory of classification. In the Pre-World War II era the theory of classification was only descriptive, that is what could be inferred from the already published classification systems such as the DDC, LCC. But the theorists like C.A. Cutter, W.C.B, Sayers, E.C. Richardson, H.E. Bliss, S.R. Ranganathan and groups like CRG, London, FID/CR, or DRTC Bangalore have contributed considerably towards a pre-mediated theory of classification systems design and evaluation. Ranganathan rather delved much deeper than others and formed a detailed and minute theory on every aspect of classification. He divided the whole work into three Planes which he called Idea, Verbal and Notational planes. In this lesson we have studied the importance of the Idea and Verbal planes and the various objective canons and principles that Ranganathan formulated for work in these phases. In the Idea plane basic subject constituents are sorted into discrete concepts by selection and successive application of characteristics. Sorted out discrete elements called isolates by Ranganathan are arranged in arrays and chains by use of exclusive canons for them. Principles of Helpful Sequence guide us in placing equally ranked entities in an array in some systematic and helpful order. Verbal plane gives concrete names to the airy concepts in the mind by clothing them in standardised terminology. A standardised, free of homonyms and synonyms terminology is vital for any science communication. In a schedule of classification terms should be current, and written and read in the context of the upper link. Their definition and scope can be best determined by enumeratiion i.e. by listing the subtopics. The terms used should be current but not judgemental or critical.

Notation is more than a series of short hand ordinal symbols to denote subjects. It is a system; an artificial language to translate the subject of the document into ordinal numbers for arrangement. Its other functions are to preserve and mechanise the order of subject decided in the idea plane. It helps in number synthesis, shows hierarchy of subjects, is an essential component of classified catalogues and a location tool in OPACs and shelf lists. Notation is essential for Chain Indexing that is a process of deriving subject headings from the class number. Combined with verbal terms notation can be very effective for high precision and high recall in any retrieval system. Notations are usually of two types, pure and mixed. Pure notation comprises of a single species of digits such as 0/9 as in DDC, or A/Z as in RIC, mixed notation uses mix of many species such as 0/9, A/Z, a/z, even mathematical marks such as = + and punctuation marks such as : ; - ( ) and many more. A mixed notation is a necessity in face of complexity of knowledge and its organisation, yet it should be kept as simple as possible. For example, BC-2 or LCC use 0/9 and A/Z only. The CC notation comprising of six species of 74 digits in all is too complex by all accounts, and it is one of the reason for low use of the CC. Among the qualities of notation are its brevity, simplicity, easy pronunciation, and familiarity. In brief it should be user friendly. Digits should be equitably distributed among the subjects. Symbols used should be familiar and their ordinal value should be obvious or very clear. Now the digits should be available on the computer key board. Notation should show hierarchical relations and be free of homonyms and synonyms. It should be mnemonical which is easy to remember. It means same or analogous concepts occurring at different places should be denoted by the same digits. For example, food, animal feed and fuel occurring in indifferent main classes should get the same number. Not only this; diseases, mechanical disorders and social ills should be denoted by same digits in medicine, mechanical engineering and sociology respectively. Physiology is always denoted by 3 wherever it occurs in the CC e.g., G: 3 General Physiology

I: 3 Plant Physiology

K: 3 Animal Physiology

L: 3 Human Physiology

But the most vital quality of notation is its hospitality. It is defined as capacity to accommodate new subjects at their proper places without disturbing the existing ones. DDC does it mostly by leaving gaps here and there, and by decimal fraction at the end of chain. Gap device though popular is not a scientific solution of problem. Ranganathan invented the use of sector notation, Empty, emptying digits to interpolate and extrapolate new subjects in an array. Devices for making new isolate numbers or specifying the existing vague ones are availed of by the classifiers. These devices are geographical device, chronological device, subject device, superimposition and alphabetical device. More the number of devices for hospitality or specificity of subjects more complex a notation. Need is being felt to use in library science advances made by notation in subjects like mathematics or chemistry.

3.13 ANSWERS TO THE SELF CHECK EXERCISES

1) An array is a line of entities of equal rank. These entities should be arranged in some systematic and predictable sequence. Some guiding principles for the arrangement of members in an array in helpful order are : Historical sequence, Evolutionary Sequence, Geographical or Spatial contiguity sequence, increasing quantity or complexity order, canonical sequence, and lastly alphabetical order.

2) In the verbal plane we assign standard terms to the concepts of idea plane. Standardisation of terminology is very important in any science any serious communications. The terms in any science should be free of homonyms and synonyms. These should be current and uncritical. Terms should be read in context of upper link in a classification schedule.

3) Notation is an essential adjunct of library classification, though in knowledge classification it may be dispensed with. It is more than a code or series of shorthand symbols to denote subjects. It implements the decisions of the Idea plane. Notation preserves the order of subjects decided in the idea plane, and also machanises their arrangement when replacing documents on the shelves after use. It may be easily described as engine of library classification. It is essential by a location device and useful as access point in OPACs and ordering device in shelflisting. Classified catalogues cannot be constructed without notation. It also helps in number synthesis and chain indexing.

4) Notation is a system to represent subjects, and its qualities have been categorised as optional and vital. Accordingly its two major qualities are its user-friendliness and hospitality, respectively. Former refers to its ease of use. A brief, familiar and moderately mixed notation is useful and liked by librarians and users alike. It makes a classification popular. Hospitality is the capacity to give place to new subjects at proper places without disturbing the existing ones. A classification without adequate provisions for hospitality will become dated, senile and die. Hierarchy is its scientific quality.

3.14 KEYWORDS

Array	:	Group of entities of equal rank arranged in some definite and helpful order.
Base of Notation	:	Total number of digits in a notational system. Base of notation in DDC is only 10, while in CC it is 74. In the LCC it is 36.
Canons	:	Normative principles applicable to a branch of a subject, e.g. Canons of cataloging, Canons of classification.
Chain	:	A group of unequal but independent entities arranged in general to specific or broader to narrower order. A family lineage forms a chain.
Characteristics	:	An attribute or property which forms the basis of division into subgroups. "Sex" is a characteristic in dividing a class of boys and girls on the basis of gender.
Digit	:	A single character in a notation say 9, A, or +: etc.
Empty – Emptying Digit	:	A digit which is both empty and emptying U, W and Y are Empty-Emptying digits in the CC.
Empty Digit	:	A digit having only ordinal value and without any semantic value. In CC 0, 9, z and Z are empty digits in lower order arrays.
Empty Digit	:	A digit which takes away the semantic power of the preceding digit but allows it to retain its ordinal value. T, V and X are Emptying Digits. For example in KX, K has nothing to do with Zoology, but KX will file between K and L.
Planes of Work	:	Three successive phases of work to divide the work of classification in three distinct Sectors called Idea, Verbal and Notational planes, respectively.
Postulation Approach	:	Work of designing classification systems based on a pre-mediated theory in the form of normative principles, Canons, Principles, and Postulates.
Terminology	:	A system of standardised names given to concepts and entities for unambiguous communication.

3.15 REFERENCES AND FURTHER READING

Parkhi, R.S. (1972). *Library Classification: Evolution of a Dynamic Theory*, Delhi: Vikas, pp. 52-101. Print.

Ranganathan, S.R. (1967) *Prolegomena to Library Classification*, 3rd ed. Mumbai: Asia, pp. 143-216. Print.

Satija, M.P. (2004) *A Dictionary of Knowledge Organization*, Amritsar: Guru Nanak Dev University, pp. 248. Print.

Satija, M.P. (1993) "Ranganathan's contribution to library and information science terminology : a historical perspective" *Lucknow Librarian* 25 (3/4) : 64-74. Print.

Foskett, A.C. *The Subject Approach to Information.* 5th ed. London: LA Publishing, 1996. 183-189. Print.

Hunter, Eric. *Classification Made Simple*. Aldershot, UK : Ashgate, 2002. 70-80. Print.

Palmer B.I. and A.J.Wells. *Fundamentals of Library Classification.* London : Geroge Allen, 1951. 60-81. Print.

Philips, W.H. *A Primer of Book Classification.* 5th ed. London : Association of Assistant Librarians, 1961.40-41. Print.

Ranganathan, S.R. *Prolegomena to Library Classification*. 3rd ed. Bombay: Asia, 1967. 219-326. Print.

Satija M.P. *Dictionary of Knowledge Organization*. Amritsar : Guru Nanak Dev University, 2004. 94-96, 147-150,157-158. Print.

UNIT 4 COMPARATIVE STUDY OF SCHEMES OF CLASSIFICATION

Structure

4.0 OBJECTIVES

After reading this Unit, you will be able to:

- compare and contrast the salient features of the major living library classifications, namely, the DDC, UDC, LCC and CC;
- discuss the strengths and weaknesses of these systems of classifications; and
- make a right choice of classification suitable to your library.

4.1 COMPARATIVE LIBRARIANSHIP

Comparative librarianship is a branch of international librarianship to study and compare the library systems of different countries and organisations. This methodology can equally be applied to various systems and services offered by different libraries. Quite often comparative study of classification systems and their elements is made by comparison of the features of various systems. A model library classification can be designed by borrowing the best features of each. S.R. Ranganathan's *Prolegomena* (1937, 1957, 1967) uses this comparative method to evolve a theory of library classification. It may be noted that comparison is done between two similar things. It will be no use comparing a horse with a cow or even car with a cart, but will be gainful to compare two different brands of cars.

4.2 INTRODUCTION TO MAJOR SCHEMES OF CLASSIFICATION

Modern history of library classification began in 1876 with the publication of Dewey's system. Its use spread very quickly as it neatly and instantly solved many of the problems being faced by the librarians in shelf arrangement and display of books. In its wake many new general classification systems emerged mostly to improve upon it or explore some alternative approaches. Noticeable among these systems are:

Dewey Decimal Classification (DDC) (1876+)

C.A. Cutter's Expansive Classification (1893)

Universal Decimal Classification (UDC) (1895+)

Library of Congress Classification (LCC) (1903+)

J.D. Brown's Subject Classification (SC) (1906)

Ranganathan's Colon Classification (CC) (1933+)

Bliss' Bibliographic Classification (BC) (1940-1953)

Rider's International Classification (IC) (1961)

Broad System of Ordering (BSO) (1978)

Bibliographic Classification, 2nd ed. (BC)(1977+)

Systems by Cutter, Brown, Bliss (1st ed.) and Rider are no more in use. Bibliographic Classification, 2nd ed., being revised by J. Mills is still not complete, though considered as one of the best classification of the present times. DDC, UDC and LCC are considered the three major living classification systems, and are highly popular, and at time in competition with one another. Colon Classification of S.R. Ranganathan is the most scientifically designed system. It rather brought a paradigm shift due to its revolutionary method of facet analysis and postulational approach. Though CC itself is in danger, or may not even survive for long, but its methods have already pervaded the science, terminology and technology of classification design. It is a class apart, though not used highly. It is a sort of mother to later day classification systems. All other systems like DDC, UDC, BC-2 and BSO have borrowed its methods to the extent possible for their revision and modernisation. Here is a comparative study of these systems.

4.3 DISCIPLINE AND MAIN CLASS

All the library classifications are first divided by discipline. A discipline is a fundamental field of teaching and learning. A discipline is a major chunk of knowledge characterised by the similarity of objects of study or use of a common research methodology. Disciplines are academic in nature and are ways of looking at the world by academicians. Three classic and traditional disciplines in order are Sciences, Humanities and Social sciences. Now many more disciplines have emerged such as Physical Sciences, Biosciences, Behavioral Sciences, though some call them sub-disciplines. Number of such disciplines keeps growing as the knowledge grows. Obverse of a discipline is isolated object or phenomena e.g., copper, child, tree, school which could be studied in context of any discipline.

4.3.1 Main Class

A discipline or sub discipline is further divided on the same basis into smaller chunks called main classes. All current classifications are based on main classes which makes the primary or basic facet. The main classes in any system form the first and mutually exclusive array of the division of the universe of knowledge. A main class may be defined as a homogenous, coherent and interrelated area of knowledge within the comprehension of an ordinary intellectual being. The scope and number of main classes vary from system to system and from time to time. For example, Astronomy is a part of Mathematics in CC (1963), while in DDC it is an independent class. Many smaller topics in the CC sixth edition (1963) got the status of a main class in the seventh edition (1987). Nevertheless, the number, scope and order of main classes form the core of any library classification system.

DDC

As constrained by its decimal notation in the DDC the discipline based main classes are numbered 1/9 and the Generalia class denoted by 0 precedes them. The main classes of the DDC as denoted by a minimum of three digits are :

000 Generalia

100 Philosophy, Psychology

200 Religion

300 Social Sciences

400 Linguistics

500 Natural Sciences

600 Technology (Applied Sciences)

700 Arts (Fine)

800 Literature

900 Geography, Biography and History

It may be noted that in the last main class Geography and History have been clubbed together as there is no more space available after 9 in a decimal notation. The (MCs) 100/600 are sciences based upon reason; 700/800 are imaginative works, while 900 pertains to memory. This is based upon the three faculties of mind as proposed by an English Philosopher Francis Bacon (1561-1626). Dewey in his arrangement of main classes inverted the three Baconian mental faculties of memory, imagination and reason. Each of the main class is further divided into ten divisions and each of the division is further divided into ten sections :

500 Sciences

510 Mathematics

520 Astronomy

: : :

590 Zoology

Each of the 100 Division ending with one zero is divided into ten sections

510 Mathematic

511 General Principles

512 Algebra

516 Geometry

519 Probability

This division can be carried to any extent by putting a dot after the third digit:

511.1 Finite mathematics

511.2 Mathematical logic

511.3 Approximation

511.4 Mathematical models

This division can be further carried:

511.3 Mathematical logic

511.32 Sets

511.322 Set theory

511.3223 Fuzzy sets

It may be noted that apart from being rooted in 17th century philosophy of Francis Bacon there are many flaws in the structure. The theory of three faculties of mind namely, reason, imagination and memory is no more scientifically valid. The matrix of dividing by 10 at every stage is artificial. One can easily see the unjustified separation of 400 Linguistics from 800 Literature. There is no justification for keeping Philosophy 100 and Psychology 150 together. History 900 has been separated from Social Sciences 300. There are many more such flaws at lower levels. Yet the DDC is credited to be first discipline based classification and is the most popular system today. Disarming its critics it makes no pretense of being a true map of knowledge. It is a practical shelf arrangement system and tries to give every significant topic a place in an overall scheme of subjects.

UDC

DDC is the base of UDC. Therefore, all the above criticism equally applies here. Yet some rectifications have been done by merging 4 Linguistics with 8 Literature. The main class 4 has been kept vacant and is likely to be filled with the newly developed faceted class Medicine. It will vacate 61 of its current class Medicine to expand 620 Engineering. Its auxiliaries and special tables are a sort of cosmetic surgery over the DDC to improve its structure and efficiency in classifying micro literature.

LCC

LCC, developed during 1898-1910, consists of 21 main classes denoted by A/Z, and are based somewhat on the Expansive Classification (1893) of C.A. Cutter (1837-1903):

A General Works

B	Philosophy, Psychology, Religion
C/I	History
G	Geography and Anthropology
H, J/L	Social Schemes, Law and Education
M/N	Music and Fine Arts
P	Languages and Literature
Q/T	Science, Medicine, Agriculture and Technology
U/V	Military/Navy
Z	Bibliography and Library Science

I,O,W,X,Y have been kept vacant. In the above evolutionary arrangement theory precedes practice. Though each class is independent, the whole schedules have been expanded to 51 volumes in depth of details. It has been described as a general classification comprising of a series of depth schedules. It is a best example of an enumerative system. The main classes are further divided by a second seemed alphabet:

Q	Science
QA	Mathematics
QB	Astronomy
QC	Physics
QD	Chemistry

Double digit subdivisions are further sub divided by numerals. It is the only classification which is now using arithmetic numerals in face of the trend of decimal notation:

QD	Chemistry
1-65	General topics
71-142	Analytical chemistry
146-197	Inorganic chemistry
241-441	Organic chemistry
901-991	Crystallography

As a shelf classification it is quite successful.

CC

S.R. Ranganathan (1892-1972) was a great thinker and theoretician. Despite this he had not developed any theory of classification when he conceived, designed and published CC between 1924-1933. Though he believed in Vedic classification, yet outlook of his main disciplines and main classes is traditionally Western; the first division of knowledge in CC is in traditional disciplines in the order of their evolution, i.e. Sciences, Humanities and Social sciences. Within each discipline the CC has a well thought out order of main classes based on clearly stated principles.

A/B	Science/Mathematics

C/D	Physics/ Engineering
E/F	Chemistry/Technology
G/H	Biology/Geology
I/J/K	Botany/Agriculture /Zoology
L	Medicine
M/N	Useful/Fine Arts
O/P	Literature/Languages
Q/R	Religion/Philosophy
S/T	Psychology/Education
U/V	Geography/History
W/X	Political Science/Economics
Y/Z	Sociology/Law.

Between M and N a unique main class Δ Mysticism has been interposed. Generalia or form classes such on Bibliography, Encyclopedias, etc. have been denoted by a/z, while the newly emerging main classes such as Library Science, Mass communication have been denoted by 1/9. A/M Sciences and Technology have been arranged in order of their increasing concreteness, as M useful Arts, an assortment of applied arts and crafts, is the most concrete in the group. Within A/Z, as shown in the above pairings, theory is followed by applications, e.g., I Botany precedes J Agriculture. This arrangement known as the Principle of dependency was first proposed by August Comte (1798-1857), the father of Sociology. In N/S Humanities the arrangement is in order of increasing richness of subject contents. The T/Z Social Sciences are in the order of the increasing artificiality of their laws. Z Civil Laws are considered purely artificial and frequently changing. Mysticism is at the confluence of sciences and humanities and is considered highest knowledge in Hindu tradition. Coupled with the form of documents the arrangement of books on shelves is in a form, what Ranganathan calls APUPA pattern. This order is pedagogical. No other classification comes near to such a fine and systematic order of subjects as that of the CC.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

1) Define a main class (MC). Make a critical study of the MCs of the DDC.

....

....

....

....

4.4 NOTATION

Notation, an adjunct of classification, is the most visible feature of a library classification system. Its efficiency and user-friendliness mostly determine the quality and popularity of any system. A notation should consist of familiar digits which convey their order obviously, should show synthesis and hierarchy of subjects. Further it should be brief, mnemonic and must be hospitable to new subjects.

4.4.1 DDC

DDC invented the use of Indo-Arabic decimal notation to denote subjects and their further subdivisions. A dot is put after the third digit only as a partitioning devise, which has no mathematical or semantic value:

001/999 Universe of knowledge

700 Fine arts

780 Music

787 String music

787.8 Plectra lute

787.87 Guitar

Its notation is pure, hierarchical and mnemonical. It hospitality is poor as it has only gaps and decimal fraction devise for this purpose. Its allocation of notation among the subjects is faulty due to historical reasons. For example, the highly dynamic and rich class 600 Technology gets the same space as given to a static subject like 200 Religions. But its strength lies in its simplicity and internationally used Indo-Arabic numerals.

4.4.2 UDC

UDC based on DDC denotes its main classes decimally 0/9 which can be further divided hierarchically as in the DDC. But more powerful is its kit of synthesis and auxiliaries for number building.

In addition there are special auxiliaries applicable only to a given small area. Thus the notation of UDC is mixed, quite hospitable to new subjects by way of hierarchy, gaps and by use of alphabets and auxiliaries.

Symbol	Function	Examples
+	Coordination	02+07 Library Science and Journalism
/	Consecutive extension	5/6 Science & Technology
:	Simple relation	02:07 Relation between Library Science & Journalism
[…]	Subgroup	[1+2]03 Dictionary of Philosophy & Religion
::	Order-fixing	02::07 Library Science & Journalism (order fixed)
=	Languages	02=161.1 Library Science in Russian
(0…)	Form	7(091) History of Art
(1/9)	Place	7(540) Indian Art
(=…)	Ethnic grouping and nationality	7(=72) Australian Tribal Art
'…'	Time	02 '20' Library Science in 21st Century.
.00	Point of View	7.00028 The Christian views on art
-03	Materials	645.13-037.87 Linoleum floor coverings
-05	Persons	7-053.2 Children's art

Problems of UDC Notation

The ordinal value of the symbols is fixed and clear in manual arrangement, but there is problem in computer aided arrangement as the ordinal value of the punctuation marks comes in conflict with the ASCII. Though the notation seems complex, yet this complexity seems inevitable and acceptable in view of its aim of being a bibliographic classification to be used in bibliographies and information centres.

4.4.3 LCC

As said earlier, the LCC uses two Roman capitals for its main classes. Then each of the two digit alphabet is further divided by arithmetical notation. The notation being moderately mixed, and the large base from A/Z, and further divisions like AA to VZ and Z give it enormous capacity for future expansions. Moreover, letters I,O,W,X and Y are still vacant. In the arithmetical notation many gaps have been left which can be filled. Where there are no gaps, of late, it has started using decimal extension for inserting new subjects. Use of alphabetical subdivisions provides endless hospitality at a point.

QD 149	Inorganic chemistry	
QD 149.5	General works	
QD 149.7	By region or country	
QD 149.7	A-Z	By country
QD 149.7	In	From India
QD 149.7	Jap	From Japan

The notation is not mnemonic except for the alphabetical subdivisions

A full class number may also include cutter number for the author and the year of publication:

Economic way of thinking by P.T. Heyne, 2003

HB 717.5 H 46 2003

Here H46 is cutter number for Heyne, and 2003 is the year of publication.

4.4.4 CC

Notation of CC is a high water mark of library notation. It is a comprehensive system in itself, and is bred on systematic canons and devices. Only problem is its complexity and frightening class numbers. But Ranganathan was not daunted by its or his own criticism on this account. Notation of CC comprises of 74 digits belonging to six species of digits:

1.	A/Z Main classes	26
2.	Δ Greek letter	01
3.	0/9 Decimal Notation for isolate numbers	10
4.	a/z (except i,l,o) Common isolates	23
5.	Special Indicator digits * ←	03
6.	Ordinary indicator digits.	
	& ‘ . : ; , - = → + ( )	11

The notational base of the CC is the widest ever in any library classification. Therefore, it has the largest room to accommodate new subjects at their proper places. Ranganathan devised many methods for hospitality of notation though faceting itself is a great hospitability mode. Apart from conventional hospitality devices of decimal fraction and gaps, he devised sector notation, empty and emptying digits for interpolation and extrapolation of new subjects in arrays. Also the notation is extremely mnemonic:

Class	Anatomy	Physiology	Diseases
G Biology	G : 2	G : 3	G : 4
I Botany	I : 2	I : 3	I : 4
K Zoology	K : 2	K : 3	K : 4
L Medicine	L : 2	L : 3	L : 4

To explain, General anatomy, Plant anatomy, Animal anatomy and Human anatomy have everywhere been denoted by ":2" .

4.5 EXTENT OF USE AND POPULARITY

Library classifications are designed for practical use; some were even designed for use in a specific library. But soon their use extended outside their specific institutions. Extent of their use determines their survival and consequent their teaching, research and published literature on them. Popularity of a system depends on many factors like, time of its origin, inherent technical qualities, ease of use, use in centralised cataloging such as MARC, constant revision, support services, institutional backing, and its marketing.

4.5.1 DDC

It is the pioneer system and also the most popular one. Used in about 2 lacs libraries in 140 countries across the globe, it is said the sun never sets on it. Apart from this, it has been translated in about 34 languages of the world including Hindi, Arabic and Vietnamese. Some sixty national bibliographies use this system to arrange their contents. Its Internet accessible version, known as WebDewey, is available through "OCLC connection." It is also used in organising and searching some search engines. Some such examples are: Webrary www.webrary.org/reb/weblinksmenu. html, and the UK Web Library<www.scit.wlre.ac.uk/wwlib/. Webrary is a service provided by the Morton Grove Public Library in USA. It links to the useful references are organised by the DDC class numbers. The U.K. Web Library (WW Lib) is provided by the University of Wolverhampton School of Computing and Information Technology.

4.5.2 UDC

By birth UDC was not designed for shelf arrangement of books. It aimed at documentation and information centers. Now in terms of its applications it is the most diversely used tool from organising libraries, websites. bibliographies to artifacts and realia. It is used in 125 countries. In 34 countries it is the main classification system used across national information centres. In 45 countries it is used in certain kinds of libraries. Its translations exist in 39 languages. Nationals Information Services and Systems (NISS), UK, now called Intuit, provides information for Education. It uses UDC to organise its directory of Networked Resources. For example, selecting a specific class number say 34 will list all the resources on Law, while 343 will only list resources on Criminal law.

4.5.3 LCC

Though the LCC was designed only for the Library itself yet it is being used in about 60% of the US large public, academic and research libraries. LCC numbers appear on MARC records which are used by many libraries for copy cataloging throughout the world. Even some national bibliographies of Asia and Europe are using this system. Due to depth of details, constant revision and institutional backing it has a very bright future. Its online version known as Classification Plus also includes LCSHs, and is much more versatile. In the web environment the potential of using LCC as a tool for organising Internet resources has proved quite successful. It is used in Cyberstacks (sm) on the Iowa State University website. The Cyberstacks is a collection of important www and other Internet resource in the selected fields of science and technology categorized by the LCC, e.g,

G Geography, Anthropology

H Social sciences

J Political science

These are further divided by the LC class numbers, e.g, TL 787-4050 will provide a resource on the NASA astronauts biographies. For each resource a brief annotation is also provided.

4.5.4 CC

CC is an influential system though not a highly used one. It is claimed that in 1960s about 20 university libraries used this system along with many public and college libraries in India. The CC numbers are given as an element in the INB entries though its main arrangement is by the DDC. No new library is opting for this system due its dated schedules and lack of any support or backup service. But it is still being taught in library schools of India.

4.6 HISTORICAL CONTRIBUTION

Many library classifications both dead and living, have contributed individually to the classification theory and practice. History of library classification is exciting.

4.6.1 DDC

Contributions of DDC are many and everlasting. Use of decimal notation was an ingenious stroke of discovery. Though simple and efficient to represent subjects and place new subjects, it has many inherent limitations to portray a true structure of knowledge. Later classifications have used decimal notation in one way or the other. Even the LCC is now resorting to this method for hospitality of new subjects. Its other contributions are division by discipline, depiction of hierarchy, and invention of the relative index. It is 135 years old and going strong and getting popular day by day. The lesson is: well governed schemes remain rejuvenated and are trusted by librarians.

4.6.2 UDC

UDC, credited to be the first bibliographic classification, introduced for the first time in any classification powerful synthetic equipment in the form of auxiliaries. It performs the jobs of shelf arrangement and information retrieval with equal ease. Some also credit it as the first faceted scheme which heralded the Ranganathan methods. It was also the first classification to be tested for usefulness of classification in computerised databases.

Also known as European Dewey it was the first classification sponsored by a professional organisation which ultimately became International Federation for Information and Documentation (FID). Further, it was the first classification available in three official languages namely, French, German and English. Its translations are available in 24 languages. It again is the first classification to be owned and managed by a consortia of publishers (UDCC) spread across the globe. Its contributions to classification are both technical and organisational.

4.6.3 LCC

It has the distinction of being the producer and the consumer at the same time — a prosumer, Alvin Toffler would say. It is a general classification with a series of depth schedules and can be used alike both in general and research libraries. Its great support base from the Library of Congress and use in centralised and cooperative cataloguing services cover many of its technical drawbacks. Support of world's largest and greatest library and its use in its excellent bibliographic services impose greatness on this system despite its so many technical faults and weaknesses.

4.6.4 CC

It is a pioneer faceted scheme which brought a revolution in classification thinking and practice. Its methods of facet and phase analysis have become the general theory of classification. It has devised an objective mechanism for designing and evaluating library classification systems by dividing the entire work in Idea, Verbal and Notational planes. By finely formulating canons, principles and postulates for the process of classification he raised classification work to the status of science. Now many schemes, both general and special, have sprung up using facet analyses. BSO is one such example. Many old systems like DDC and BC-2 have used its methods for their revision. It started a new paradigm which is the basis of all indexing languages useful for print and electronic environments.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

1) Explain the contribution of the CC to classification.

....

....

....

....

4.7 SUMMARY

We have discussed the salient features of four great library classifications namely, the DDC, UDC, the LCC and CC. We discussed their main classes, their arrangement, notation, and their methods for hospitality of new subjects and their popularity or extent of use. DDC is a pioneering modern classification. It invented decimal fraction notation which is hierarchical and provides endless hospitability in chain. Its constant revision and use in OCLC WebCat and in about 2 lac libraries around the world makes it the most popular system. New libraries are going for it as a natural choice. Its other contribution of division by discipline, invention of the relative index and well formed and oiled sound machinery for governance, maintenance and marketing make it one of the outstanding classification among the public and the professionals. UDC is the first faceted bibliographic classification sponsored by an organization which later became FID (closed in 2000). Over the DDC base it superimposed a powerful synthetic equipment. That has made it an efficient scheme for information retrieval in computerised databases. It is used in more than one lakh libraries and information centres. Its official availability in French, German and English also makes it to serve as an inter-indexing switching language. The Library of Congress system with 21 main classes comprising of 29 parts and 51 volumes runs to about 11000 pages. It serves both as a general and depth classification. It is the only living and thriving enumerative classification today. Though its organisation is faulty yet it is among the big three library classifications due to strong organisational back up and use in MARC records. Being is producer and consumer keeps it ever updated to accommodate new subjects. The CC though not a highly used classification is the one that has brought a revolution in classification theory and practice. Designed as the first truly faceted classification Ranganathan further refined it into a analytico synthetic classification. He formulated a wealth of canons and principles for the science of classification which now form the general theory of classification and are helpful in designing other such systems. Though its own future in bleak, yet the methods it has developed will live long.

4.8 ANSWERS TO SELF CHECK EXERCISES

1) In modern library classification disciplines of knowledge are first divided into main classes. A main class is a homogenous area of knowledge whose length and breadth is within the comprehension of a normal scholar. Number and scope of main classes varies from classification to classification system and from time to time. In the DDC there are 100/900 main classes preceded by a Generalia class 000. The number of MCs is ten only because there are ten places in a decimal system. It is an artificial division, indeed. Further DDC main classes are based on the inverted Baconian System.

Faculty	**Subject**	**Main Classes**
Reason	Sciences & Technology	100-600
Imagination	Art & Literature	700-800
Memory	History	900

This theory is now outdated. Also it separates sciences from Technology and History from Social Sciences. In all, divisions by ten are unnatural.

2) The CC designed by S R Ranganathan (1892-1972) between 1928-1933 was first published in 1933 by the Madras Library Association. Now it is in its 7th edition published in 1987. It was the first faceted classification and later it refined and upgraded itself into an analytico–synthetic classification based totally on postulational approach. Ranganathan solved many problems of the enumerative systems such as of DDC and LCC. His system provides individualising class numbers to the documents and provides infinite hospitality for new subjects. On the theory and methods of the CC many new faceted schemes have been developed. Faceted systems are now quite useful for information retrieval and searching the Web.

4.9 KEYWORDS

Apupa Pattern : The CC is able to arrange documents in a pedagogical order on the shelves through the use of two types of common isolates.

Comparative Classification : Use of comparative methods to classification systems to identify the best practices and elements to design an ideal classification. It is a part of international librarianship.

Discipline : A large area of knowledge having similar objects of study or a common research methodology. The traditional disciplines are Natural science, Humanities and Social Sciences.

Main Class : A traditional area of coherent knowledge whose length and breadth falls with in the comprehension in of a normal scholar. Also it is the first array division of a discipline. For example, sciences are divided into main classes mathematics, physics, chemistry, zoology, etc.

MARC Record : Machine readable catalogue produced by the Library of Congress for online use and for distribution of cataloguing records. The main feature of such records is internationality recognized numerical tags assigned to each field known as MARC 21 to identify each field in a catalogue entry.

Mixed Notation : A notation comprising of two or more species of digits, e.g., combined use of alphabets and numerals as in the LCC, or numerals and punctuation marks as in the UDC. Notation of the CC is the most mixed.

Notation : Series of short hand symbols to represent subjects and to mechanically fix their order decided in the idea plane. That is why Ranganathan called it as servant of the idea plane. It is also helpful in synthesis of numbers and to mechanise the arrangement of documents on the shelves or entries in a classified catalogue.

Pure Notation : A notation comprising of single species of digit, e.g. the DDC has only 0/9 used decimally, or Rider's International Classification comprising of alphabets uses only A/Z.

4.10 REFERENCES AND FURTHER READING

Chan, Lois, Mai. (2007) *Cataloguing and Classification: An Introduction,* 3rd ed. Lanham, MD: Scarecrow Press, pp 320-443. Print.

Dewey, Melvil (2003) *Dewey Decimal Classification and the Relative Index*, 22nd ed/ ed by Joan Mitchell. Albany, N.Y: The OCLC, 4v. Print.

Hunter, Eric J. (2002) *Classification Made Simple*, 2nd ed. Aldershot U.K: Ashgate, pp. 120-130. Print.

Ranganathan, S.R. (1987) *Colon Classification*, 7th ed./ed by MA Gopineth Bangalore: SRELS, Vol. 1. Print.

Ranganathan, S.R. (1967) *Prolegomena to Library Classification*, 3rd ed Bombay: Asia. Print.

Rowley, Jennifer and Hartley, Richard (2008) *Organizing Knowledge : An introduction to Managing Access to Information,* 4th ed. Farnham, England : Ashgate, pp. 214-217. Print.

Satija, MP. (2004) *A Dictionary of Knowledge Organization*, Amritsar: Guru Nanak Dev University. Print.

Satija, M.P. (2008) "Universal Decimal Classification: Past and Present" *DESIDOC Jl. of Info Tech. 28(6)* : 3-10. Print.

Satija M.P. (1986) "Use of Colon Classification" *Int. Classification* 13 (2) : 88-92. Print.

BLOCK 2 LIBRARY CATALOGUING

Introduction

Library cataloguing shares a symbiotic relation with library classification. The latter provides access to the resources in a library by class number, enables browsing to select a book, provides a panoramic view of the resources whereas the former provides all bibliographic details and subject of resources at one place. Library classification provides a linear view of subject relations but the catalogue takes care of the multi-dimensional relations. There have been significant changes in library cataloguing due to the arrival of non-book material and use of computers for processing information. ICT has enabled sharing of information for which paradigmatic changes have been brought about in cataloguing. The four units in this Block are devoted to the concepts of cataloguing and developments in it.

Unit 5 introduces the concept of cataloguing. It is titled **Basic Concepts**. After defining library catalogue, it discusses its need, objectives, functions and qualities. There are some other bibliographic records in a library that are related to a library catalogue. These include shelf list and accession register that have been discussed in relation to a library catalogue. **Types and Forms of Catalogues** is the title of Unit 6**.** There are two forms of a catalogue, viz. internal and external forms. Internal forms relate to the internal arrangement of catalogue entries whereas the different physical forms are referred to as external forms. Different types of both the forms have been discussed in detail in the Unit.

Unit 7, **Formats and Standards**, discusses International Standards Bibliographic Description (ISBD) and MARC formats. Standardisation in cataloguing is needed to bring in uniformity in cataloguing procedures. Rules for cataloguing are examples of standardisation. These help to identify, present and display bibliographic elements in catalogue entries. Formats have been given to codify and present bibliographic elements for processing by computers. Different MARC formats and their characteristic features have been discussed in the Unit.

Unit 8 is **Cataloguing of Non-Book Material**. Non Book Materials (NBM) are those materials which do not come within the definition of a conventional book, periodical or pamphlet. Audio-visual materials, microforms, computer files, electronic resources, etc are some examples of these. NBM require special treatment in terms of their bibliographic description for access and searching information. The Unit discusses different NBM giving examples and rules for their cataloguing in AACR2R.

Blank Page

UNIT 5 BASIC CONCEPTS

Structure

5.0 OBJECTIVES

The library catalogue is an index to the contents of a library, which helps the reader in discovering and locating documents in a library. It is an important tool of a modern and well equipped library. It is very essential for the proper use of library resources, for easy and quick location of information contained in the library holdings. It is a key to unlock the resources of a library. In this Unit, you will be acquainted with the basic concepts of library catalogue and cataloguing. After reading this Unit, you will be able to:

- discuss the objectives and functions of a library catalogue;
- differentiate between a library catalogue and other library records;
- discuss the role of technology in cataloguing process; and
- explain symbiosis (mutual relationship) between classification and cataloguing.

5.1 INTRODUCTION

Library catalogue, said to be the mirror, reflecting the holdings of a library, helps the user to locate and access the material required by her/him. It identifies to the user the bibliographical details of books for the purpose of selection and study. The catalogue groups the books in the library according to author, subject and collaborator etc. By consulting the catalogue, a user can know the availability of all types of reading materials, both print and non-print in the library. It is a time saving device both for the user and the library staff. A library catalogue is expected to satisfy every kind of bibliography i.e., enquiry but it is obviously limited to the collection of a particular libraries.

5.2 LIBRARY CATALOGUE

An ideal and functional library facilitates its readers in many ways for optimum utilisation of the resources by adopting certain techniques and operations like classification and cataloguing etc. The extent of the use of resources greatly depends on an efficient, effective and updated catalogue so as identify, locate and access the collection easily and quickly. A comprehensive catalogue enhances the image and reputation of a library. In this section, we shall discuss in detail the definition, purpose, objectives and functions of a library catalogue.

5.2.1 Definitions

The word 'catalogue' has been derived from the Greek phrase 'Katalogs'. It means a list, register or complete enumeration of things. 'Kata' means, by or according to, where as 'logos' means word, order, reason. Hence catalogue can be explained as the work in which contents are arranged in a reasonable way or in a particular order or according to a set plan.

The library catalogue can be defined as, "a list of documents of a particular library or group of libraries arranged according to a systematic or logical order providing bibliographical information along with a location mark for easy identification and quick access".

The Terminology Group at the International Conference on Cataloguing Principle (ICCP) held at Paris in 1961 defined a catalogue as a, "Comprehensive list of a collection or collections of books, documents or similar materials". Ranganathan defined it as "a list of documents in a library or collection forming a portion of it. It is a methodically arranged record of information about bibliographical resources".

According to C.A. Cutter, a library catalogue is a, "list of documents which is arranged in some definite plan. As distinguished from a bibliography, it is a list of books in some library or collections".

The Oxford English Dictionary defines a catalogue as "usually distinguished from a mere list or enumerated by systematic or methodic arrangement, alphabetical or other order and often by the addition of brief particulars, descriptive or aiding identification indicative of locality, position, date, price or the like".

Hence a library catalogue can be defined as, "a list of documents (both books and other reading materials) in the holdings of a particular library or group of libraries arranged according to a set plan or recognised order and containing specified items of bibliographical information for the purposes of identification and location of the material catalogued".

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

1) State the definition of a library catalogue.

....

....

....

....

5.2.2 Need

Libraries generally acquire reading and reference materials both in print and non-print forms for the benefit of readers for the purpose of study, reference and research. All the time the desired documents may not be available physically on the book shelves as these are issued and if the users go to the shelves directly without consulting the catalogue, they will think that these documents are not acquired by the library though actually these are possessed by the library. Besides, the non book materials cannot be browsed which are likely to miss the notice of readers. Further, the reader may waste her/his valuable time in locating the needed document without taking the call number from the catalogue. Because of these reasons, it is absolutely necessary that a library must prepare an effective catalogue ensuring the users for quick identification, location and access to the reading materials.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

2) State and explain the need for a library catalogue.

....

....

....

....

5.2.3 Objectives

Library catalogues are prepared to accomplish certain objectives as mentioned by the legend and genius in the field of cataloguing namely Cutter, Ranganathan and Lubetzky. In the opinion of Cutter, a catalogue should:

"1) enable a person to find a book, of which

a) the author, or

b) the title, or

c) the subject is known.

2) show what the library has

d) by a given author,

e) by a given subject,

f) in a given kind of literature.

3) assist in the choice of a book

g) as to its edition by describing the work adequately for easy identification, or

h) as to its character"

However, according to Ranganathan, "if a reader is interested in a subject which takes him to the library, his wants will be better served if the catalogue can spread before him a full, connected, panorama of all the materials on his desired subject, all its sub-divisions and all broader subjects of which it is itself a sub-division" (CCC, p. 81).

Further, Ranganathan has stated in the Library Manual about the objectives of a catalogue which can answer the following questions:

1) Is there a book in the library by such and such author? What are all the books in the library by her/him ?

2) Is there a book in the library with such title?

3) Is there a book in the library by a particular editor, translator, reviser, compiler?

4) Is there a book on a specific subject and its sub-divisions?

5) Is there a book in a publisher's series in the library?

Moreover, Ranganathan has analysed the objectives in the light of the Five Laws of Library Science and stated that the catalogue should be designed so as to:

1) disclose to every reader his or her documents;

2) secure for every document its reader;

3) save the time of the reader; and for this purpose

4) save the time of the staff.

However, Lubetzky is of the view that a library catalogue should serve "First to help for quick location of a particular publication, i.e. of a particular edition of a work which is available with library. Secondly to display the relevant and related documents of a given author".

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

3) Write objectives of library catalogue as stated by Cutter.

....

....

....

....

5.2.4 Functions

The basic function of the library catalogue is to let know an enquirer whether or not a particular publication is in the collection of the library and if so where it can be found. It also reveals what material the library has on a given subject, author etc. Catalogue is a tool designed to enable the enquirer to find a particular book of which some or all the details are known; to survey the entire stock or sections of it; and to select the books which will best serve her/his purpose. The larger libraries and particularly national libraries can assume the bibliographical function appropriately. A library catalogue is expected to answer the following questions: Is a particular book available in the library?

Which books by a particular author are available in the library? Which editions of this particular book are available in the library? Which books by a particular author are in the library? Is there a book in the library with such and such collaborator – editor, translator reviser, compiler etc.? What are all the books in the library with him as collaborator? What are all the books in the library in that publisher's series? Is there a book in the library on such and such subject? What are all the books in the library on that subject?

Shera, in emphasising the importance of a subject catalogue as opposed to a dictionary catalogue, enumerates the following functions of the subject catalogue : (1) to provide access by subject to all relevant material; (2) to provide subject access to materials according to the principles of subject organisation; (3) to bring together material of same subject regardless of the disparities in terminology etc.; (4) to show affiliations among subject fields; (5) to provide entry to any subject field from most general to most specific; (6) to provide entry through any vocabulary common to groups of users; (7) to provide a formal description of the subject content of any bibliographic unit; and (8) to provide means for the user to make selection from among items in any particular category.

In the International Conference on Cataloging Principles, Paris it has been unanimously decided that the functions of the catalogue should be, an efficient instrument for ascertaining:

i) whether the library contains a particular book specified by

   a) its author and title or

   b) if the author is not named with book, its title alone, or

   c) if author and title are inappropriate or insufficient for identification, a suitable substitute for the title, and

ii) which works by a particular author and which editions of a particular work are in a library.

   Hence the functions of a library catalogue can be summarised as:

   1) It is the most important finding tool for staff reference but its primary task is to enable the readers to ascertain what books a library possesses.

   2) It provides records in an order which helps the reader to know what those materials are, where they are, and to determine their character and suitability for that purpose.

   3) It serves as an exhaustive bibliography on a subject.

   4) It helps the library staff in ordering new books and to avoid duplication of the materials by checking the catalogue.

   5) It displays the record of library's resources with a view to making them easily accessible for study and reference.

   6) It serves as a dependable tool for communication of ideas and subjects dealt within the books to readers who use the library.

   7) It bridges the wide gap between the resources of rich collections of a library on one side and the users on the other side who are looking anxiously to satisfy their thirst for information by getting their desired documents without any loss of time.

5.2.5 Qualities

The requisites of a good catalogue should be considered for preparing a functional library catalogue which are enumerated as under:

i) **Accuracy:** Errors of spelling and style or format may be avoided. Inaccuracy of any type may lead to misunderstanding, loss of time and even dislocation of a card leading to disappointment to a reader. Cards should be carefully checked before filing.

ii) **Consistency:** The rules of cataloguing should be strictly adhered to so that consistency can be maintained. If an amendment in rules is required to be made due to some practical exigencies, it may be made only after a careful consideration of all the pros and cons. But once an amendment has been adopted, it should be followed consistently and there should not be after thoughts and frequent changes. All changes should be recorded in a file or on the cataloguing code for future reference and guidance.

iii) **Needs of Users:** A librarian should always keep the needs of the users in mind while deciding about the various entries in the catalogue. He should keep in mind the type of queries frequently made by the users. The catalogue should be so designed that it is able to satisfy the needs of majority of the users.

iv) **Arrangement:** The arrangement of content in an entry should be simple, uniform and easy to understand. It should not confuse the reader but give clear cut guidelines about the nature of the book and its location.

v) **Up-to-dateness:** There should be no time gap between acquisition and processing of books and between preparation of catalogue cards and their filing in the catalogue cabinet.

vi) **Multi-pronged approach:** The cataloguer should try to provide at least a few added entries in order to enable the users to locate a document from various angles i.e. author, title, editor, subject, joint author. The chances of retrieval of documents on demand increase with this multi-pronged approach.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

4) Mention the functions of a library catalogue as decided in the International Conference on Cataloging Principles, Paris.

....

....

....

....

5.3 LAWS OF LIBRARY SCIENCE AND LIBRARY CATALOGUE

All the activities of the library revolve around the Five Laws of Library Science because these are the guiding principles for effective scientific planning of a library. For preparing an effective catalogue, a number of guidelines are implied in these Five Laws enunciated by Ranganathan.

'Books are for Use', the First Law implies the need for organising the collections of a library for optimum utilisation by providing a number of physical facilities and providing reader services.

The physical form of a catalogue should be such that it is flexible, can be kept up-to-date, easy to consult and allows speedy search. It allows for addition of entries for new documents as and when added to the library. The users should know the scope, contents and coverage of documents by means of annotation and useful notes in a catalogue so as to make appropriate choice among various documents. The catalogue is the only medium to bring together the users and the collections leading to maximum use and satisfy the thirst for knowledge.

The Second Law states "Every Reader Her/His Book" which implies that various approaches of the users i.e. search by author, subject and title etc. should be satisfied. Many books are of composite nature. Neither the title nor the main entry may disclose the contents. This law requires that hidden contents should be brought to the notice of the readers by preparing subject analytical entries. So also the Third Law "Every Book its Readers" requires for preparation of analytical entries and cross-reference entries as and when necessitated, especially in case of books published under editorial direction (where a number of contributors contribute papers). Seminar/Conference proceedings and journals in which some of the articles/papers are highly needed by the readers are likely to miss the notice of the readers if only added entries are prepared. Instead they will be helped if analytical entries are prepared.

The Fourth Law "Save the Time of the Reader" ensures to save the precious time of the users. A catalogue should not only be simple in its design and construction but also up-to-date, communicable with elaborate guides. For effective use of the catalogue, user education/orientation programmes should be organised for the users.

'Library is a Growing Organism' is the Fifth Law which implies that catalogue also grows because of addition of books to the library. Keeping in view the changes and growth in nature and variety of publications, other forms of documents, needs of users and the advent of ICT has enormous impact on changes both in the physical form and internal structure of the catalogue. Change is a must and the library should prepare to face such challenges. Hence, while selecting a physical form decision should be taken on the basis of features like longevity, durability, space, simplification, portability, selectivity, flexibility and cost.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

5) Explain the implications of the Laws of Library Science in designing a catalogue.

....

....

....

....

5.4 LIBRARY CATALOGUE VIS-a-VIS OTHER LIBRARY RECORDS

So far we have discussed the definition, purpose, and functions of a library catalogue. In a library certain other records are necessary, primarily for administrative and stock taking purposes.

5.4.1 Library Catalogue, Index, Shelf-List and Accession Register: A Comparison

Let us discuss such records and tools to know their differences in scope and utility.

i) Shelf list, Accession Register and Index

ii) Bibliographies

iii) Publishers Catalogues

Index means 'to indicate' or 'to point out' and we all know that at the end of a book there is an exhaustive index showing the list of terms along with the relevant page numbers for instant reference for easy retrieval of the information. Very often the index satisfies the subject approach.

		Library Catalogue	Index
i)	Arrangement	A catalogue can be arranged alphabetically or in a classified sequence.	An index is always arranged alphabetically.
ii)	Entry	Includes some descriptive specification of a document containing a subject.	Index entry only specifies the subject.
iii)	Flexibility	New entries can be inserted in a systematic order at any time.	New subjects can be inserted any time.
iv)	Purpose	It is a record describing the documents acquired by a library.	Provides access to any of bibliographical entries of the catalogue through author, title or subject.
v)	Entries	Analytical entries may be prepared depending on the nature of the document.	No such entries are required in case of an index.

Shelf List

It is an inventory which records bibliographical data of items arranged exactly the same way as arranged on the shelves. Each title is represented by a card mentioning the author, title, edition, numbers of volumes, number of copies, call number and such other items as necessary. The call number determines the arrangement of the cards in the shelf list in the exact order of the arrangement of books on the shelves and a reader can easily obtain the needed book from the shelf by this number. A shelf list thus serves principally for inventory and control of the collection.

More or less the shelf list card is a duplicate of the main entry excepting the exclusion of notes, contents and tracing. To some extent, it resembles a classified catalogue i.e. a subject catalogue.

		Library Catalogue	Shelf List
i)	Users	It is an indispensable tool for readers as well as library staff.	It is consulted only by staff of the library.
ii)	Degree of use	Used frequently.	Used usually at the time of book selection and stock verification.
iii)	Arrangement	Arranged either alphabetically or in a classified sequence depending on the choice of the inner form of catalogue.	Always arranged in classified order.
iv)	Form	May be in various forms like card, register or computerised.	Generally in card form.
v)	Added and Analytical entries	Needs added entries to satisfy various approaches of the readers.	No such entries found in shelf list.

Accession Register

It is a very important administrative record in which books either purchased or received through gift and exchange are listed in chronological order date wise and assigned a number sequentially. The record contains the information about author, title, edition, date of publication, publisher, price, source of supply, sometimes the call number and remarks. As it is a date wise record of the collections and arranged according to serial number of the items, it cannot serve as a finding tool about the availability of documents of a library. In many libraries it also serves as a stock register. To get information from this register, one needs to know the accession number of the document.

Library Catalogue	Aecession Register
i) **Purpose**: Helps in identifying and locating a book in a library and serves as a retrieval tool.	Indicates the total collection of the library. Retrieval of books is not easy rather a time consuming
ii) **Use**: Extensively used and brings all the like subjects together.	process. Date-wise collections Used by library staff only and fails
iii) **Items of information**: Detailed and useful Information about the document is given.	to bring together the similar subjects together. Bibliographical details are much
iv) **Degree of use**: Since it is an indispensable tool, it is used heavily both by the users and library staff.	More less as compared to library catalogue. Used some times by staff only,
v) **Arrangement**: Arrangement is done systematically either alphabetically or by call	Arranged chromo logically according to accession or serial number of the item.
vi) **Form**: Form may be in register/ book, card or computerized form self check exercise	Usually in a book/register

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

6) State briefly the purpose of shelf list and accession register.

....

....

....

....

5.4.2 Library Catalogue, Bibliography and Publishers' Catalogue : A Comparison

Library Catalogue	Bibliography	Publisher's Catalogue
1) Purpose: To serve the users and satisfy the laws of library Science.	Depends upon kind of bibliography. To bring to the notice of the readers select list of documents relevant to the study	Promotion for sale of books.
2) Function: Serves as retrieval tool.	Inventory functions	Inventory functions
3) Scope: Collection confined to a particular or group of libraries. However it is not restricted to subject or to only language or geographical area and time etc.	Normally it is restricted to a particular subject region, person or language. Importantly it is not confined to any library.	Resticated to the books published by a particular publishers or a group of publishers.
4) Arrangement:Either alphabetically or acceding to call number.	Usually it is arranged alphabetically.	Subject wise alphabetically and again the list of document of particular subject arranged according to authors alphabetically.
5) Users: Large number of uses including research scholars and library staff.	Mostly the scholars and library staff.	Librarians, faculties and scholars.
6) Physical form: Found generally in card, register and computerized form.	Usually found in the form of book:	Occurs in the form of book (both printed or mimeographed)
7) Bibliographical information: Maximum information about the document along with availability and location mark is provided.	Likely to have useful information with an annotation for selecting the best document	Information is inadequate. Highlights the special feature and revision, if any about the document.
8) Preparation of entries: With the help of catalogue codes containing the rules for preparation of entries made basing on the needs of users.	Prepared acceding to some citation standards (e'g BIS: 2378 or 2381	Prepared from the documents them selves Does not follow any cataloguing codes

Some examples of publishers' catalogues are:

- Books in print
- Ulrich's International Beriodical Directory,
- Catalogues of publishers like Mcegraw-Hill, SpringerVerlag, Van-Nostrand, Academic press, Elsievier Concept, UBS, ESS, Oxford & IBH, Vikash Publishing company etc.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

7) What is the distinet characteristic of library cataloguer that differentiates it from a bibliography?

....

....

....

....

5.5 CATALOUGING AND THE ROLE OF TECHNOLOGY

As we know varieties of entries are prepared for the document to satisfy the various approaches of readers as the needs and requirements differ from one user to another. Hence the main entry and added entries are usually prepared for this purpose. Main entry provides maximum relevant and useful information about the documents. Where as, the added entries are prepared under a number of access points such as joint authors, editors, translators, title and series of the document consisting of brief information for easy location.

Cataloguing is the process of preparing entries for a catalogue.

The steps involved for making entries are enumerated as under:

i) Choice of access point / determination of heading and rendering for the above entries.

ii) Recording of information under various sections / areas of deseription of the entries.

iii) Determination of the style of writing and punctuation marks etc, as per international standard bibliographic deseription(ISBD).

iv) Writing call number in all the entries.

v) Filing of catalogue cards.

vi) Preparation of guide/ reference card, proper maintenance and updating of entries in the catalogue.

All this processes and procedures are involved in the act of cataloguing. Provision of detailed information facilitates a reader for selecting the needed documents. To help the cataloguer for cataloguing properly keeping in mind the expected approaches of the readers, the catalogue codes have been developed, designed and revised. The Classified Catalogue Code (CCC) was formulated by S. R. Ranganathan in 1934 is a user oriented code that provides rules for subjects entries and exhaustive rules for choice and rendering of bibliographic items for descriptive cataloguing. Likewise Anglo- American Cataloguing Rules (AACR) edition 1and 2, 1988 revised edition, amendments and regularly incorporation of new rules for new items like"electronics resources" proves to be the most popular catalogue code. It follows International Standard Bibliographic Description (ISBD) format for the purpose of cataloguing which achieve uniformity, consistency and standardisation for preparing catalogue entries.

Besides, it is amenable for computerisation of library activities and leads to international library corporation, resources sharing and global networking.

Guidelines to the Cataloguers

i) He must be thorough and well versed with the rules of the catalogue code used for preparing entries. The information provided in the entry should be accurate and, most importantly, the choice of access point otherwise the needed document may not be located and the reader will be appropriate about the accessibility of that document.

ii) The title page is the main source of information to a cataloguer. However, some time the information contained in preface, table of contents and the body of the text etc. is also used for cataloguing non book material.The cataloguer is expected to run the different materials like CD's, micro film, video-cassettes and web resources.

iii) The information contained in the catalogue entries should be sufficient enough to provide access points to satisfy the multifarious approaches of the users.

iv) All the entries for a book should be filed immediately before there are sent to circulation section. The entries of all the lost, damaged and with drawn books be removed from the catalogue to ensure that the' catalogue is kept live and update. Proper maintenance, use of guide cards and training to the users on how to use the catalogueimportant for optimum utilisation of resources of the library.

Due to unprecedented growth of publications, it is difficult to know the location, availability of publication with the help of traditional / manual methods. But with the advent of IT and Information communication Technology (ICT), it is possible to ascertain what information exists and where . Many libraries are computerized and networked so that they can have access to the resources of other libraries through on line access i.e. the catalogue entries are stored and searched for display as desired by the user.

In the context of cataloguing process, there is no scope for computer to create the basic entry as it involves technical skill and expertise in subject field. But once the data is entered and suitable software package is in use, then the computer is capable enough for preparing highly complicated type of catalogue for control of information. It can manipulate the entries and the list of books of be displayed easily either under other, title, subject even under a particular publisher and in a specific year. We have to grab this opportunity offered by IT so that maximum benefit can be given to the users for best exploitation of the resources of other libraries.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

8) Name four important catalogue codes known to you.

....

....

....

....

5.6 SYMBIOSIS

Classification helps to arrange documents on the shelf in a logical order bringing together documents on like subjects and separate those on different subjects. Library catalogue interprets the collection of the library and helps the user to select and locate the desired document. These are two interrelated processes which help the readers in laying hands on their requisite information. Classification and cataloguing satisfy the laws of library science and work for their mutual help. Where classification fails to satisfy the need of the users, catalogue, comes to the rescue by preparing the suitable entries and vice versa. This mutual relation between classification and cataloguing is termed as symbiosis. This mutual cooperation is necessary for providing effective and expeditious services to the clientele.

Now let us discuss the possible areas where the classifier comes to the aid of cataloguer:

i) Use of classification scheme automatically brings a systematic and logical order of the documents on shelf as well as the entries in the classified part of a catalogue. All the books on a subject and its subdivisions are arranged in close vicinity due to classification.

ii) According to the Canon of Pre-potence, the call number is an individualising number, a part of the leading section of the main entry. As such the place of main entry among various entries is totally concentrated in the leading section.

iii) The class number assigned by the classifier helps the cataloguer in assigning subject headings to a document.

iv) The class index entries are derived from the class number.

Similarly, the cataloguer helps the classifier, as mentioned below:

i) Some books deal with more than one subject or are interdisciplinary in nature but are classified in one place only. To bring to the notice of the readers, the cataloguer prepares more than one subject entry providing access from different subjects dealt with therein.

ii) In case of books bound together/ printed together in one cover, another main entry is to be prepared for the second book as the book will have one place only on the shelf as per the classification number assigned. Hence by preparing cross reference entries the location of second part of the book can be shown to the users. So also for seminar and conference proceeding and edited books etc., analytical entries are prepared under the specific articles/papers contained in the above documents.

iii) The books acquired by the library may not be available all time on the shelves as these might have been issued or misplaced making the reader to think that these books are not available in the library. A library catalogue will inform the reader about the availability of such books in the library.

Hence, both classification and cataloguing act as complementary and supplementary to each other for effective organisation of collection and providing best services to the users.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

9) Explain briefly complementary nature of classification and cataloguing.

....

....

....

....

5.7 SUMMARY

In this Unit we have studied some of the basic concepts of the library catalogue and cataloguing. The definition, objective/ purpose, and the different functions of a library catalogue are explained with reference to the various approaches of readers in using the collection of a library. The implications of the Five Laws of Library Science are explained in relation to a library catalogue and cataloguing. The Five Laws not only provide useful guidelines for the preparation of library catalogue but also emphasise the need for saving the time of reader in the use of the library. The catalogue vis-à-vis other library records such as accession register, shelf list, bibliographies, and trade catalogues are highlighted with illustrative examples. Cataloguing as a process in organising a library collection has been briefly explained along with provision of some guidelines to cataloguers. The complimentary and supplementary nature of the twin library operations of cataloguing and classification are indicated. The usefulness of a library catalogue as a reference tool is also briefly mentioned.

5.8 ANSWERS TO SELF CHECK EXERCISES

1) Library catalogue may be defined as a list of documents and other materials organised in a systematic order available in a library.

2) The basic purpose of a library catalogue is to serve as a guide to the collection of materials acquired for the library. Primarily the library catalogue reveals to users of a library the document in a library and helps the person in finding out whether documents of the person's interest are available in the library or not. It also serves users as a retrieval tool.

3) According to C.A. Cutter, a library catalogue should:

   i) enables a person to find out a document of which

      a. The author, or

      b. The title, or

      c. The subject is known

   ii) Show to user what the library has

      a. by a given author

      b. on a given subject

      c. in a kinds of literature

iii) Assist users in the choice of a document

a) as to its edition (bibliographically)

b) as to its character (literary or topical)

4) In the I.C.C.P. Conference it has been on unanimously decided that the function of the catalogue should be an efficient instrument for ascertaining

i) Whether the library contains a particular book specified by

a) its author and title or

b) If the author is not named with book, its title alone, or

c) If author and title are inappropriate or insufficient for identification, a suitable substitute for the title.

5) Laws of library science/implication for a library catalogue

a) Books are for use

Accessibility to library collection through a well designed physical form of a catalogue as well as the inner structure of a catalogue entry and their organisation.

b) Every reader his book.

The physical form as well as it is internal structure of library should be designed to help every type or category of users, such as a children, specialists or physically handicapped persons.

c) Every book its reader

Provision of analytical entries for document to reveal the hidden contents of documents.

d) Save the time of reader

All possible approaches to the catalogue through author, title, and subject indexes, cross references entries, and guidelines in the use of the catalogue, all these provide speed of service.

e) Library is growing organism

Provision to adopt a library catalogue to changes in publications (both paper-print and electronic) and changing needs of users and other environment factors.

6) An accession register and a shelf list are essential management records of a library and are not meant for public use. Accession register is a data-wise record of reading materials acquired by the library, giving all details about the documents, arranged in the serial order of documents. A shelf list reflects the arrangements of documents on the shelf .Their purposes are different and hence they cannot be used as a substitute for a library catalogue.

7) A bibliography is an organised list of reading materials of a particular author or particular subject or a particular geographical region . It is not limited to collection of any particular library. A library catalogue is list reading and reference materials. Acquired by a library. This feature differentiated a library catalogue from bibliography.

8) The important cataloguing codes are as follows:

   a) Anglo- American Cataloguing Rules,1908.

   b) American Library Association Cataloguing Rules, 1949.

   c) Anglo-American Cataloguing Rules, edition 2.1978.

   d) Classified Catalogue Code of Dr. Ranganatham, edition 5.1964.

9) Catalogue and Classification are two interrelated processes that are performed in a library. The basic purpose of these two operations is to aid users in locating and selecting appropriate reading materials required for study, research or any other organisation of documents on the shelves of library while a catalogue serves the purpose of a retrieval tool and also project the contents possessed by a library. All these mechanisms are to a large extent complementary to each other, and together they serve the users in an effective manner to use the library collection. Between themselves, they rectify the shortcoming or limitations of each other.

5.9 KEYWORDS

Analytical Entries : Entries that go into a catalogue for parts or chapters of a document.

Bibliographical Data : Data comprising author, title, editor, publisher, place of publishing, year, number of pages, illustrations, etc.

Canon of Prepotence : The potency to decide the position of an entry among various entries in a catalogue should if possible be concentrated totally in the leading section and the highest potency i.e. power/ strength lies in the entry element that is on the class number and further if total concentration is not possible in the leading section then it flows to the next section that is heading section and the least potency lies on the last section that is the accession number.

Cross Reference Entries : Directing elements connecting two related entries, one of which leads to the other, also called see and see also entries.

Index : To indicate or to point out .Provides access to any of the bibliographical entries of the catalogue through author, title or subject index.

Inventory : A detailed, often descriptive list of something.

Linear Arrangement : Arrangement one after another, as in a line on library shelves.

Machine-Readable Form : A form of document, readable only by a machine such as a computer or microform reader.

Standard Format : Recognised and accepted organisation of bibliographic data of a document.

Symbiosis : The mutually beneficial partnership between classification and cataloguing.

5.10 REFERENCES AND FURTHER READING

Bowman, J.H. *Essential Cataloguing*. London: Neal-Schume, 2002. Print.

Krishan Kumar. *Cataloguing: Theory and Practice*. New Delhi: Har- Anand, 1993. 9-23. Print.

P.S.G Kumar and Muhammad Riaz . Cataloguing: Theory and Practice. New Delhi, S.Chand , 1999. 17. Print.

UNIT 6 TYPES AND FORMS OF CATALOGUES

Structure

6.0 OBJECTIVES

In Unit 5, you have learnt about the definition, objectives and functions of a library catalogue. In this Unit you will know about the forms of a library catalogue. After reading this Unit, you will be able to:

- discuss the different forms of a catalogue;
- compare their relative advantages and disadvantages;
- explain the criteria to select the best form of a catalogue; and
- discuss the features of a computerised catalogue.

6.1 INTRODUCTION

A library catalogue has a long history of one and a half century. It has evolved over the period of time. The evolution has taken place in the form, format as well as contents of a catalogue. Developments in document production and changing user needs have affected the library catalogue. Information Communication Technology (ICT) has also played a role in the developments in catalogues and cataloguing. A catalogue is discussed from the point of view of its external form/ physical appearance and internal form. The different forms have been discussed with their relative advantages and disadvantages in this Unit.

6.2 INNER FORMS OF A CATALOGUE

The inner form of a library catalogue refers to the arrangement of a catalogue entry in a logical and systematic order to fall into a helpful sequence for storage and retrieval. The chart below depicts the various inner forms of a catalogue.

INNER FORMS OF CATALOGUES

- Alphabetical
  - Author
  - Name
  - Title
  - Subject
  - Dictionary
- Classified
- Alphabetico-classed

There are three types of inner forms of a catalogue, viz. alphabetical, classified and alphabetico-classed. Author, Name, Title, Subject and Dictionary catalogue fall in the category of an alphabetical catalogue. A Classified Catalogue is so named because it is arranged in a classified order. Classified order is an artificial order which may be difficult to use. Hence a classified catalogue has two parts the other part being alphabetical which is easy to use and leads the user to the classified part. Alphabetico-classed catalogue is a variation of the classified catalogue. It overcomes the shortcoming of the artificiality of a classified arrangement by making it classified alphabetically.

Self Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

1) Define the inner forms of a library catalogue.

....

....

....

....

2) Enumerate different categories of inner forms of a library catalogue.

....

....

....

....

6.2.1 Author Catalogue

Author is the person who is chiefly responsible for the intellectual thought content of a work. In simple words, author is the creator of a work. In an author catalogue, the entries of documents are under author's name and arranged alphabetically. It provides access to documents by the names of their authors. In other words, the leading section of an author catalogue would comprise the name of an author. An author is generally a person or a corporate body who is responsible for the thought contents of the document brought out under the name. Listing of personal names of author varies greatly because of the cultural traditions in the naming of person in different regions of the world. For example, names of persons in Western countries, Indic names, Muslim names, Chinese and Japanese names differ as per their own traditions.Which part of the names should be taken as the lead in a catalogue has been set by cataloguing codes and there are established practices. There are also a variety of corporate bodies under whose names documents are published. Although we are not concerned very much with rendering of names in catalogue entries here, it is important to note that their filing position is determined by the rendering of names. Inaccuracies in the rendering of names would seriously affect the alphabetical arrangement of entries in the catalogue.

The catalogues of the British Library, the Library of Congress, the National Library of India are some of the fine examples of author catalogues.

Libraries may have author catalogues arranged in three different ways: a) There may be an exclusive author catalogue without entries mixing it with any other entries such as titles, subjects, series, etc. b) Author entries may form part and parcel of a dictionary catalogue. c) Author entries may form part of the alphabetical index of a classified catalogue. Irrespective of the form in which an author catalogue exists, it provides an important approach to a document. If the user approaches the catalogue with the correct name of an author, the catalogue immediately gives the person all the documents by the author. To help a user, the other variants of the name of an author are usually provided as cross-references in a catalogue.

Advantages

1) It brings together the titles of books of the same author at one place in the catalogue

2) It helps a user to obtain at a glance what books are available in the library by a given author. This function, can however be fulfilled by author entries in other inner forms of a library catalogue. In a classified catalogue, this function is performed by the alphabetical index of dictionary part.

3) It ensures that there will be no scattering of works of the same author through the catalogue.

Disadvantages

1) Approaches of readers under subject, collaborators, distinctive titles etc. cannot be satisfied.

However it requires that the reader must know the exact name of the author and title. In case the author is known by different forms of name or pseudonym, the reader may be helped with the use of see references.

6.2.2 Name Catalogue

A name catalogue is a variation and extension of an author catalogue. It contains entries for works of one author and also for books written on him. All entries are arranged alphabetically by the name of the author. In other words, a name catalogue is a compound or mixed type of catalogue which combines the author and subject entries (the subject entries representing the author as a subject) into one alphabetical sequence. In this type of catalogue autobiography, biography and other critical studies, memoirs and diary of an author, are arranged along with his original work. The author entries include:

- corporate authors, both as an author as well as a subject
- name series
- place name forming part of an author heading

The following examples, exemplify these points:

Person as an author	Nehru, Jawaharlal: Discovery of India
	Nehru, Jawaharlal: Glimpses of World History
	Nehru, Jawaharlal: Towards Freedom
	Autobiography of Jawaharlal Nehru
Person as subject	Jawaharlal Nehru, Political Biography by F. Moraces
	Jawaharlal Nehru: Biography by S. Gopal
	Nehru: the Making of a Nation by M.J. Akbar
Name Series	Oxford Historical Series
	Madras University
	Sir C.P. Ramaswami Aiyar Endownment Lecture Series
Place name	Bombay University
	Calcutta University

Name catalogue serves as an author catalogue and also as a subject catalogue. Any reader interested on the works by or on an author may find this type of catalogue very useful to find specific material of his interest. In this type of catalogue, one can find the works of Rabindra Nath Tagore as well as works on him arranged in a single alphabetical order under Tagore.

Name catalogue seems to be almost confined to Great Britain. Such a catalogue outside Great Britain appears to be rare. The catalogue of the British library is near name catalogue which includes references from all names that occur in titles.

Advantages

1) It is useful and valuable for collection of books of an author and books on her/him.

Disadvantages

1) Subject entries are limited to personal and corporate names only.
2) There is no title entry in this form of catalogue so also under collaborator and series.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

2) State the essential difference between an author catalogue and a name catalogue in two lines.

....

....

....

....

6.2.3 Title Catalogue

In a title catalogue, the titles of documents occupy the leading section of entries, which are arranged in an alphabetical order. Queries of readers who remember only the exact title can be answered with the help of a title catalogue. However, it is noticed that many of the readers do not spell out a title exactly the way it appears on the title page, particularly non-fiction titles. To fulfill title approach of readers, entries can be selectively provided in catalogues of public libraries for fiction and for those that are well-known by their titles. It is a part of dictionary catalogue or part of author-title index of classified catalogue. For books in classical language particularly Sanskrit and Pali, this catalogue is somewhat preferred.

Advantages

1) Satisfies the approaches of readers especially in cases for classical languages and for fiction, drama, poetry and for the class literature.

Disadvantages

1) It cannot alone satisfy all approaches of the readers except the title approach.

6.2.4 Alphabetical Subject Catalogue

It is a list of books in a collection, each entered under the name of the specific subject as a heading, the entries being arranged alphabetically. Several books on the same subject will be brought together in the catalogue. When there is more than one book on one subject, the author of the book is taken to determine the alphabetical order. If the author's name is common for several books, then the title of the book is taken in to account for arrangement of entries. The sample example of this type of catalogue is given below:

CHEMISTRY ORGANIC

Bahal, B S

Text book of Chemistry

DOCUMENTATION

Mukherjee, A K

Fundamentals of Special Librarianship and Documentation

ECONOMICS- INDIA

Ghosh, Alok

Indian Economics

This form of subject catalogue is called an alphabetical subject catalogue as the specific subjects are arranged alphabetically. If a reader approaches the catalogue with a view to find book or books on a specific subject it provides excellent service as s/he has to refer to it like dictionary. However if at all reader wants to see all the materials on a field of knowledge systematically, the catalogue fails to respond to his query i.e. a book on BOTANY will be separated from books on ZOOLOGY. Similarly books on CHEMISTRY will be separated from books on PHYSICS. Hence a reader has to look under separate headings at different places in the alphabet if all the materials available on a given field of knowledge are to be gathered.

Features

i) The primary function of this catalogue is to know what books are there on a particular subject in the library.

ii) In this catalogue, entry is to be made under specific subject term which represents the specific subject matter of the book.

iii) This catalogue is made under specific subjects and again arranged in alphabetical order that is why it is named as alphabetical subject catalogue.

iv) This catalogue is prepared with the help of a Standard List of Subject Headings like Sears List of Subject Headings or Library of Congress List of Subjects Headings for consistency, uniformity and standardisation.

Advantages

i) With the help of this catalogue, the specific subject approach of the readers is satisfied.

Disadvantages

i) The related subjects are scattered throughout the catalogue owing to their alphabetical arrangement.

e.g. Money and Banking, Electricity and Magnetism, Astronomy and Planet etc.

ii) The Standard List of Subject Headings fail to cope with the ever emergence of new subjects and the new editions have to be brought out regularly; otherwise the cataloguer is unable to assign exact subject headings for newly published books.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

4) Give ten example of subject headings. Illustrate how an alphabetical subject catalogue scatters related subjects.

....

....

....

....

6.2.5 Dictionary Catalogue

It is a catalogue in which all the entries (authors, title, subject, series etc.) and their related references are arranged together in one alphabetical reference. It resembles arrangement of entries in a dictionary where the words are strictly arranged alphabetically. The various entries of this catalogue are arranged just like a dictionary that is why it is called dictionary catalogue.

Features

i) It is a mixed type of catalogue, where we find two distinct approaches merged together. It is a combination of two distinct and different approaches, the author and title approach for the reader who knows the book by the author s/he wants, the subject approach of the reader who does not know either the name of author or title of the book but wants material on some definite subject.

ii) As a result of this type of arrangement quite unrelated headings will come together and related headings will be dispersed (e.g. classification and cataloguing; flowers and rose will be dispersed.)

iii) It takes its name from its arrangement which follows the simple alphabetical order of entries.

iv) So far the subject entries are concerned, it follows the principle of specific subject heading. It should be entered under the specific subject.

v) Dictionary catalogue provides an elaborate scheme of cross reference to bring together scatted related subjects and to correlate and unify the entries in order to bring systematic and logical order.

vi) A dictionary catalogue with cross references is called syndetic or connective catalogue.

vii) A dictionary catalogue is divided into two files; one for author, title, series and collaborator entries and another for subject entries.

viii) Different types of entries in this catalogue are:

Personal

a) Main entry under- Author

Corporate

and

- Title

b) Added entries under – Joint author, Translator, Editor, Compilor, Title, Series and

c) Subject Analytical entries – In case of collections of papers like conference and seminar proceedings article of journals etc. Author, Title and Subject Analytical entries are to be prepared.

d) Cross- reference – These are of two types, viz. (1) 'see' reference and (2) 'see also' reference from unused to used headings and from specific subject to general subject from narrower term to broader term(e.g. Roses see also Flowers) respectively.

Advantages

i) It is the most popular form of catalogue used in public, school and college libraries, and even in university libraries. Its popularity is due to its arrangement like a dictionary i.e. in alphabetical order.

ii) It can satisfy the different specific needs of the readers.

iii) The users need not be aware of class numbers to refer this catalogue.

iv) With the help of cross references and general references, readers are guided from one heading to another heading.

v) Direct approach on any specific subject can be satisfied easily e.g. if a reader wants books on " CRICKET" he will refer the catalogue directly under this term and at once know all the books available on " CRICKET" in the library.

Disadvantages

i) For readers who need information on a subject with all its ramifications, the dictionary catalogue is most difficult to use. It is slower in yielding information and less satisfactory in its result than the classified form.

ii) Extensive use of cross references to bring together the related subjects together results in the catalogue becoming bulky. Its maintenance is more difficult. Moreover the cross references often proves to be tire some.

iii) It is a tedious and time consuming affair on the part of the readers to find out information on the various aspects of a particular subject and its related subjects from this catalogue.

iv) Dictionary catalogue suffers from all the consequences of excessive dependence on verbalisation.

v) Replacing the old subject terms by new subject terms is a tedious and time consuming process.

The sample of Main entry of Dictionary Catalogue is given below

Call number

541

Pk4 Palit, S R

Text book on physical chemistry/ by S.R. Palit . – 2nd ed. – Calcutta: Asia Publishing House, 1964.

Acc no XX, 245p. : ill. ; 22x12cm.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

5) Write six examples of your own, illustrate how the distributed relatives get collected in one place in a dictionary catalogue.

....

....

....

....

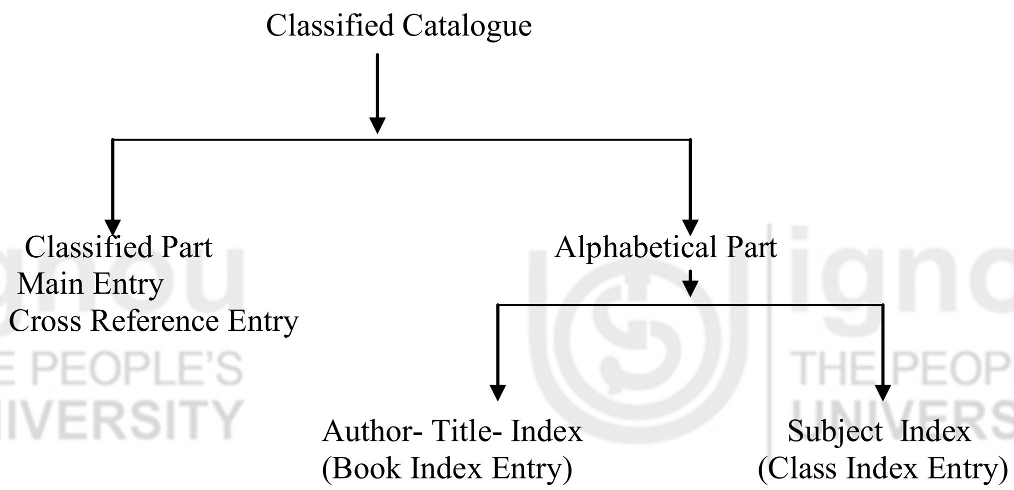
6.2.6 Classified Catalogue

Classified catalogue provides access to information by subject. It is different from an alphabetical catalogue in that the arrangement is by class numbers. It brings in an artificiality as the user is expected to be aware of the scheme of classification. To help the user, a classified catalogue is in two parts, the other part being alphabetical in nature that guides the user in navigating through the classified part and reaching the desired information.

Definition

According to Encyclopaedia of Librarianship, it is " a catalogue of subject entries arranged in the systematic order according to a scheme of classification."

Dr. Rangnatham defines it as, " a catalogue in which some are numbered entries and some are word entries". Consequently a classified catalogue consists of two parts (i) the classified part and (ii) the alphabetical part.



Classified Part

It is the arrangement of entries of documents by class number that gives the systematic and structured display of subjects in a classified catalogue. It maps out the subordinate and coordinate subject divisions and puts them in a logical sequence. This systematic arrangement to a certain degree reflects the logical thinking of specialists in different subjects and gets their appreciation.

While displaying the entries in a classified catalogue merely giving the class numbers would not be very helpful to users of the catalogue. The class numbers are not intelligible to persons who have no knowledge of the classification scheme chosen for the library. Therefore, while displaying the class numbers in guide cards, for entries under them, it is absolutely necessary to give their verbal equivalents of the divisions of class numbers. These verbal equivalents provided for class numbers in a classified catalogue are known as 'Feature Headings'.

Alphabetical Part

The alphabetical index to a classified catalogue, consisting of author, title(wherever necessary), subject entries and other entries for collaborators, series, editors of series and a host of cross references, is meant to support the classified part of the catalogue. It can fulfill all the functions of a dictionary catalogue i.e. collect the works of an author together, bring all the different editions of a title, cross reference for subjects, etc. With the classified part bringing all the related subjects together, and the alphabetical order bringing together all the distributed relatives, the classified catalogue can fulfill all the functions of a library catalogue.

Features

1) It is a subject catalogue where entry is made under the class number which represents the subject matter of the book.

2) It consists of two parts, namely the classified part (the number entries) and the alphabetical Part(the word entries) where the entries are made and arranged according to alphabetical order.

3) The success of this catalogue depends upon largely on the soundness of the classification scheme chosen and the extent to which subjects are collocated in that scheme.

4) As the classified catalogue is based on some classification scheme, all the entries are arranged from general to specific at all levels in this catalogue.

5) Any shortcoming in the scheme of classification reflects in the catalogue when the subject entries are prepared by the chain indexing from the class number of the document.

Advantages

1) In classified catalogue, the main entry not only falls under the subject with which a book deals, but it falls in close proximity with related subjects.

2) The logical or systematic arrangement of subjects is assumed because the catalogue is based on a recognised system of classification.

3) The close identity of arrangement between catalogue entries and books on shelves tends to increase the readers' familiarity with a large number of books.

4) It discloses the strength and weakness of a library by subjects, broad and specific.

5) It is easy to bring out printed parts of a classified catalogue, class by class with a consolidated index at the end of the volume when the parts are printed.

6) Specific sections of classes of the classified catalogue in a large library with a fairly balanced and representative active collection can be used as relative subject bibliographies.

7) The arrangement of entries according to notation of the classification scheme makes it independent of natural language which may lead to international standardisation and ultimately international cooperation.

8) A classified catalogue is very helpful for exhaustive literature search as it offers all possible approaches. The classified part satisfies the linear approach. In alphabetical part it serves multi-dimensional verbal approach.

9) More number of readers can consult the catalogue at a time because of the facility of separate catalogue cabinet both in classified and alphabetical part.

10) Changes in subject terminology do not seriously affect the classified catalogue. Only the subject index cards in the alphabetical portion needs to be re-filed according to changed terminology.

11) The index entries of a classified catalogue are easier to use and consult.

12) Arrangement of entries in the classified part is the same as that of books on shelves. Arrangement in the alphabetical part is like a dictionary, which is easy to understand. Arrangements of multilingual entries do not present difficulties in the classified part but in the alphabetical part, multilingual entries can present difficulties.

Disadvantages

1) In a classified catalogue, the reader has to consult the alphabetical part first then taking the call number he has to again go to classified part for detailed information about the document.

2) The success of this catalogue depends on the classification scheme chosen. Because if there is any drawback in the classification scheme, it will be reflected in the catalogue.

3) Revision of classification scheme necessitates revision of the relevant part of the subject catalogue.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

1) Bring out the differences between a dictionary catalogue and a classified catalogue with reference to the structure and provision for adding new subjects. Give your answer in a tabular statement.

....

....

....

....

6.2.7 Dictionary Catalogue and Classified Catalogue : A Comparison

No 1.	Features/Aspects 2	Dictionary Catalogue 3	Classified Catalogue 4
1.	Arrangement	One single, alphabetical sequence of author, title, subject, etc.	Two parts i) classified part arranged according to class numbers and ii) alphabetical part. Arrangement of entries in the classified part is exactly the same as that of books on shelves and in the alphabetical part is just like a dictionary which is easy to understand and to use.
2.	Author and title filiations	Works of the same author, different editions of the same work, different translations of a work kept together.	The alphabetical index gets this filiations of authors and works.
3.	Subject filiations.	Related subjects get scattered because of alphabetical arrange-ment but can be connected by a system of cross references.	Presents a structured and organized arrangement of subjects, displaying their relations.
4.	Distributed relatives	All distributed relatives are collected and brought together.	The alphabetical index can perform this function.
5.	New subject	Can be inserted into the catalogue without any difficulty.	To be fitted properly in the subject filiation. There may be difficulties to position a new subject into the classification scheme without the permission of the designers of the classification scheme. It may also introduce an element of inconsistency.
6.	Specific subject.	Provides direct approach to users if he searches under a particular subject and need not required to know the classification scheme.	A reader has to consult alphabetical part first, then taking the class number, he has to refer the classified part then only he can get all the books on the subject.

7.	Classification as a basis.	Does not depend on any scheme of classification.	Success of this catalogue depends solely on the qualities of scheme of classification.
8.	Use of cross references.	Extensive use of cross references solely depends for its success. Nut ultimately becomes more bulky which poses problems for best use.	Number of cross references is less due to automatic presence of classified arrangement which brings related subjects together.
9.	Ease of use.	Simple, ease and straight forward, a direct search is sufficient to find any kind of information.	Two step search is necessary. It does not require compete knowledge but by providing elaborate guide cards will help for best use.
10.	International cooperation.	Difficult to achieve.	Possible if UDC is used with its multilingual indexes which facilitates literature searches for foreign language literature.
11.	Printing.	Cannot be done in parts.	Printing of subject is possible, as a result a number of subject catalogues can be produced.
12.	Compilation of bibliographies.	Not very convenient to compile subject bibliographies	Very convenient for the preparation of subject bibliographies.
13.	Strength of collection.	Not possible to know the exact strength or weakness of a particular subject.	Can be easily assessed the strength of a subject and act as a book selection aid for book selector.

6.2.8 Alphabetico-Classed Catalogue

The alphabetico-classed catalogue may be considered as a combination or mixture of the best points associated with dictionary and classified catalogue. In other-words, it represents an attempt to combine some of the advantages of a classified catalogue with the directness and ease of consultation of the alphabetical catalogue. A catalogue with entries under broad subjects alphabetically arranged and sub-divided by topics in an alphabetical order. It is out and out a subject catalogue. It does not stand independently by itself as in case of dictionary catalogue. An alphabetico-classed catalogue provides alphabetically arranged broad subjects. Topics subordinate to the main subjects are listed as sub-divisions. The references needed from specific headings e.g.

MNEMONICS

<u>See</u>

PHILOSOPHY—PSYCHOLOGY-MEMORY—MNEMONICS

"A catalogue of this kind consists of an alphabetical sequence of mutually exclusive broad subject headings, under each of which appears of further alphabetical sequence of sub-headings, the process of alphabetical divisions can be carried down to the degree of minuteness required by the material being catalogue" (E.J. Coates). The rearrangement of D.D.C. division of 600 Technology in an alphabetical order would yield the following:

Agriculture and related technologies (630)

Building (690)

Chemical and related technologies (660)

Engineering and allied operations (620)

Home economics and family living (640)

Management and auxiliary services (650)

Manufactures (670)

Manufacture of product for specific uses (680)

Medicines (610)

Diseases (616)

of blood- forming, lymphatic, glandular system(614.4)

cardiovascular system (616.1)

digestive system (616.36)

ofbillary tract (616.36)

of mouth and throat (616.31)

This type of catalogue must have alphabetical index of specific subjects for its successful operation.

As will be seen from the above examples, the alphabetico-classed catalogue gets related subjects together in a way, but their alphabetical arrangement takes away the filiations of subjects. There are few examples of such catalogues today, because compromises are never entirely successful.

Parts of the original museum subject index use the alphabetico-classed principle.

Advantages

1) In this type of catalogue there is an attempt to incorporate the advantages of dictionary and classified catalogue with the simplicity of an alphabetical arrangement broad subject headings where again sub-divisions are arranged alphabetically.

Disadvantages

1) It is a complex type of catalogue and see references are made from the specific heading to the complex heading.

2) Subject headings are complex creating confusion for the reader .

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

7) List any five criteria for choosing an inner form of a library catalogue for a library.

....

....

....

....

6.3 CHOICE OF INNER FORMS OF CATALOGUE

The essential requirements in the choice of an inner form of catalogue are:

i) A good library catalogue should satisfy the different approaches of the readers.

ii) Works of same author or collaborator or on the same subject should come together.

iii) The catalogue should be logically constructed so that closely related classes should be brought together.

iv) It should be prepared in such a way that a reader can easily understand without much trouble.

v) The catalogue should be selected keeping in mind the type of user community, their information needs and nature of documents.

vi) The library must provide a user- oriented catalogue.

vii) Technical soundness of particular inner form; open or closed access system adopted; the range of services to be planned and financial position of the library.

6.4 OUTER/PHYSICAL FORMS OF CATALOGUE

In Section 6.2 you have learnt about the various inner form of catalogues, their features, advantages and disadvantages. Also you must be aware of the factors to be considered in selecting an inner form of catalogue for a library.

However, now let us discuss about the different types of physical forms of catalogue including their relative requirements i.e. the criteria for selecting the best possible catalogue which can be used conveniently by the maximum number of readers leading to least cost for its presentation and occupying less space.

In due course, you will be able to know the importance of some of the catalogues in the libraries of 21st century. You can also understand the relative merits and demerits of such catalogues and the reasons for choosing a specific one for a particular type of library. Most of these forms have historical importance since the scenario has changed with use of computers in cataloguing.

6.4.1 Criteria for Selection of the Best Physical Form of Catalogue

In the recent past, due to advent of IT, most of the libraries are in the process of computerisation and developing computerised catalogues. However, card catalogue still exists and in operation especially in colleges and university libraries. Whatever physical form a catalogue may have, it should possess all the features of the most suitable catalogue, if not all, the selected physical form should have most of the following criteria for its choice.

- Should be possible to keep it up to date which implies that one should be able to insert or withdraw entries easily as and when required. It means the catalogue should be flexible in nature.
- To bring like entries together i.e. on the same author and same subject.
- It must be easy to use and should be possible to consult it inside or outside a library. This is related to portability.
- Should be economical to produce and maintain and should have facility for making number of copies.
- Must be compact in size, so that it does not occupy much space.
- Should provided multiple access points and save the time of the readers so that at a time more number of readers can use this catalogue.

6.4.2 Bound Register Form

In this form, the entries of documents of a library are written in hand in a bound register or ledger. The information about each document like author, edition, accession number, number of copies and class number is provided and separate registers for author, title and subject can also be prepared.

Advantages

i) The ease with which readers can use this catalogue is unquestionable

ii) Xerox copies of the catalogue can be placed at different locations in libraries facilitating the readers to consult the catalogue in any corner of the library so that at a time more number of readers can use it because it has the quality of portability.

iii) Several entries that can a reader see on a page at a time without the necessity of turning one card after another is definitely an added advantage which saves precious time of the readers. Some times a card may be skipped up while consulting hurriedly.

iv) It needs neither much space nor special equipment for its display due to its compactness.

v) Since the readers are in the habit of reading and consulting the books from childhood, they feel it much more convenient to consult being like a book.

Disadvantages

i) It does not possess the quality of flexibility hence when new books are added, entries cannot be filed in their appropriate places as the left out space is filled soon. Hence such books will be entered at the end of the register or a supplementary catalogue is prepared which is a delaying process. Addition of books is an inherent feature of a library this catalogue is unable to keep good company as one is fast, the other is slow.

ii) It cannot be kept up to date.

iii) The quality of paper is not thick as the card catalogue; hence its durability is less and cannot with stand constant use by the readers.

iv) In case of stolen, damaged, tornout, mutilitated and outdated books are to be discarded, then immediately the relevant entries are to be deleted from the catalogue which looks confusing and indecent.

6.4.3 Printed Book Form

It is a catalogue of books and other reading materials available in a library bound in a volume or volumes where entries are printed on pages. It is also known as printed page catalogue or bound book catalogue. These are prepared in conformity with standard principles and rules of cataloguing. Some of the big libraries like British Library, the Library of Congress, National Library, Kolkata, Glasgow and Liverpool Catalogues printed their catalogues in the book form. To keep the catalogue updated, they regularly issued supplements to these catalogues.

Advantages

a) Since this catalogue resembles the printed reading materials, it creates less psychological barrier to the readers for use as they are familiar with the conventional printed books.

b) Speed of search is fast as compared to other form of catalogues.

c) Multiple copies of this catalogue can be made on demand and can be sold so that readers can consult the catalogue at home and hostels comfortably.

d) No thoughtless and careless reader can make any alteration in the pagination or in the order of arrangement of entries. The volumes of it can be kept on a table without any special equipment and easy to handle. In this context, Gellar has said that, "the printed book catalogue are an active, positive salesman for its service, an effective display window for its merchandise and as an instrument by which it has been able to improve professional service to the public. Experience shows that the printed book catalogue is attractive psychologically to the public, and that it is easy, simple and conventional to use. There are no long trays to pullout and pull back………."

e) In small libraries, it is considered economical to use this catalogue as in such libraries, books are added in small numbers. Subsequent, supplementary catalogues can be prepared.

f) Easy to consult as a reader can have a glance of many entries on one page.

g) It does not occupy much space.

h) Can be segmented to various sections and can also be issued subject wise to satisfy the needs of different subject groups of users.

Disadvantages

i) Insertions and withdrawal of entries or deletion is not possible as it is not flexible.

ii) It is high expensive for its production, issue of supplements takes much time hence it cannot be kept up to date.

iii) Printing of the catalogue consumes a lot of time and in the mean time new books are acquired by the library whose entries cannot be included, as a result, the information about such books cannot be brought to the notice of the readers. As such the catalogue cannot indicate the complete collection.

6.4.4 Sheaf Form

Sheaf form of library catalogue is also known as loose-leaf form. A sheaf form of catalogue is one in which slips of paper are put into a loose-leaf binder and bound by some mechanical device into a volume. This is a loose-leaf binder format, which provides the convenience of handling a book. In the sheaf form, each entry is made on a separate slip. But, there may also be more than one entry on each slip or page. The entries are either handwritten or typed. New slips can be inserted in appropriate places without disturbing the existing order of arrangement of entries. It is also possible to remove entries for specific documents in case such documents are withdrawn from the library stock. Roughly, each volume of a sheaf catalogue may contain about 500 to 600 leaves. The volumes so constituted may be displayed on special shelves with appropriate labels on their spines, indicating the order (either alphabetical or classified) of arrangement. At one time this form of catalogue became somewhat popular in countries like England and other European countries.

Advantages

i) It combines certain advantages of book form catalogue like portability, familiarity and certain good features of card catalogue like up-to-dateness, infinite expending capacity and freedom of manipulation of entries.

ii) It is possible to bring like entries together i.e. books by the same author and on a specific subject.

iii) Easy to consult and can be referred inside and outside the library.

iv) Most economical to produce and can be kept in a small wooden/ steel rack resulting into occupying less space due to its compact nature.

v) A mobile library collection may be entered in a sheaf catalogue and it can be carried in a van.

vi) Duplicate entries can be made easily.

Disadvantages

- Since the size of clip is 7- ¾ by/4 inches/ the particulars of one document may be recorded resulting into wastage of space.
- The entries be struck off as and when the relevant document is lost, damaged or withdrawn. Otherwise, again the whole slip is to be re-written or retyped on the entire page.
- The filing and fixing of sheaves into loose leaf binder is a time consuming and tedious process as it involves unlocking and locking time and again. Hence it creates problems in proper maintance of the catalogue.
- Those libraries that follow this catalogue cannot conveniently participate in any cooperative venture where unit cards are used for entries.

6.4.5 Card Form

Library catalogue in the card form is by far the most popular physical form. It is widely prevalent in libraries throughout the world including India.

In this form the bibliographical elements of every document are recorded on a single card. This method of representing every document on a single card is known as the unit card principle. These cards stand in card-trays or cabinets with a punched hole of about half a centimeter from the bottom for inserting a locking rod. This locking system keeps the cards from falling out and also prevents unauthorised persons from removing any card from the tray. Because of its wide usage all over the world, many aspects pertaining to the card catalogue are standardised. For example, 12.5×7.5cm or 5×3inchs is the universally adopted size for a catalogue card. Similarly, the sizes of cabinets, trays for a card catalogue are all of uniform standard. Consequently, most of these items of furniture could be obtained readily from commercial vendors.

Advantages

Some of salient features, which made the card catalogue quite popular are:

a) It is flexible in keeping it constantly updated with the quality of expansion and withdrawal of entries.

b) The users and the library staff can handle it with ease.

c) Possible to bring together entries with the same handing.

d) The cards are single, self-contained units. This feature permits additional approach points and cross reference in the catalogue.

e) The entries for lost books can be withdrawn and like entries can be filed together.

f) The library using the card catalogue can participate in central and cooperative cataloguing scheme. This reduces the burden of the staff.

g) The cards are hard and tough, so its durability is longer than that of other catalogues.

In fact, the unit card principle is one of the most beneficial outcomes of the card catalogue.

This principle paved the way for centralised cataloguing of documents at central place. The printed unit cards can be multiplied and distributed to other libraries at a fairly low price. The Library of Congress, USA, was the pioneer in starting this card service and many libraries in the United States and other countries use this service. The well-known commercial bibliographic publisher, H.W Wilson and Co. also provides printed catalogue card service to libraries, for selected items.

Cooperation in compiling bibliographic records is another extension of centralised cataloguing.

Libraries participating in cooperative cataloguing provide catalogue entries to the Library of Congress of those documents that are not available at the Library of Congress. The Library of Congress gets them printed for distribution, as part of its catalogue card service.

The British National Bibliographic (now incorporated with the British Library) from its very inception in 1950, assumed responsibility as a national cataloguing agency. Besides, its printed weekly and monthly edition and other cumulations, a printed card service was also made available on subscription to individual libraries. Now new records can be downloaded from its site.

Cataloguing-in-Publication(CIP)

Centralised cataloguing led to the logical step of publishing bibliographic entries of documents in the publication itself, on the reverse side of the title page. Ranganathan called it "prenatal cataloguing". This facility enabled libraries to use the bibliographical data available on the book for the preparation of their catalogues. The Library of Congress was the pioneer in this enterprise. The British National Bibliography has also been participating in the CIP Programme from January 1977. Such facilities are yet to develop in countries like India.

Disadvantages

Though the card from of library catalogue is universally accepted. It also suffers from some disadvantages. Which stated as under:

a) The card catalogue occupies large space in libraries. The problem of space is very acute in large libraries located in metropolitan cities. Libraries with a massive collection running into lakhs of documents with an annual addition of 10,000 volumes would need enormous space for its card catalogue. The cost of space is prohibitive in metropolitan cities and this has been a deterrent against continuing with the card catalogue in such libraries.

b) The growth and complexity associated with the card catalogue call for greater maintenance cost and administratively unmanageable. Difficult to detet the lost card taken out by some careless and thoughtless reader .

c) Due to inherent defects associate with the structure of a card catalogue, it is possible for a single person to monopolise a considerable number of trays at a time, precluding its use simultaneously by other users.

d) It is not portable and hence the users has to go to the library for consulting it . This naturally leads to wastage of time of time, particularly if the library does not possess the documents of his interest.

e) Speed of search is slower as compared with book form and sheaf form of catalogue.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

2) Enumerate the physical forms of a library catalogue.

....

....

....

....

3) State two reasons for a card catalogue to be popular than printed book forms.

....

....

....

....

6.4.6 Computer-produced Book Form

This type of book catalogue was produced with the help of a computer. The structure, extent of information about the document, typography etc. varied with that of conventional physical form of catalogues. The print out available used to be in the form of line printer output. Many institutes of repute like Library of Congress used to provide services like abstracting and indexing services including the production of catalogues by use of a computer.

Advantages

1) The cost of production of multiple copies was less because it is produced first by creating a master copy from which additional copies can be reproduced.

2) The maintenance cost was low since the whole catalogue display is recreated at internals which facilities the complete catalogue which can be reorganized as and when necessary.

3) At a time one entry could be referred in a card catalogue, where as this catalogue presents an entire page displaying a number of entries.

4) The computer produced book catalogue representing both the bibliographic information of a catalogue and organisation.

Disadvantages

Problems may arise in displaying current additions and changes in the cataloguing database.

6.4.7 Microform Catalogue

In microform catalogue, entries were greatly reduced and printed upon a film or fiche.

The microform cannot be read by naked eyes. A suitable microform reader, magnifies the reduced images on the film or fiche and projects them on to a screen is necessary consulting a microform catalogue.

As mentioned earlier, microform catalogue may either be in the form of a microfilm or in the form of microfiche. Microfilm can be on a single reel, but generally it is housed in a cassette containing two reel so that the film can be wound forwards or backwards within the container at the time of consultation. Microfilm readers usually allow the rotation of images through 90 degree.

On the other hand, microfiche is a transparent card type format. A reduction of nearly 42 × 42 indicates that the width and height of the microimage is 1/40 of that of the original. Microimage area is correspondingly 1/1600 of that of the original would give 200 frames per card. Microfiche has the advantage of direct access to a particular frame whereas microfilm requires a serial number though the film to locate a required entry, and serial search is a time consuming process compared to the direct access method afforded by microfiche.

Microform catalogues became popular with the development of Computer-output microform/COM. The COM method converts the digital information contained on the computer generated magnetic tape into print displayed on microform.

Microform catalogues are the output forms computerised cataloguing system. These forms were used extensively in the production of library catalogue during 1970s.

Advantages

Some of the advantages of microform catalogue are:

i) Microform catalogues are compact and occupy less space in libraries. The space requirements of microform catalogues arise mainly because of space for microfilm or microfiche readers.

ii) At a time several records can be viewed.

iii) They are portable and accessible to users depending upon the number of copies of catalogues and machines available. Multiple copies of these catalogues can be prepared easily and inexpensively.

iv) These catalogues are very easy to use and maintain. COM catalogues are much more cheaper than printed book catalogue produced by computer.

Disadvantages:

1) It is not easily updated and rearrangement of information is difficult on live catalogue.

2) Unless the file is updated, the deleted material still remains.

3) Psychologically the users do not feel comfortable in operating the equipments.

4) Cannot be used without microform reader.

5) They also required special care and protection

6) Medium and small libraries may not be benefited by this catalogue as microform readers/ printers are costly.

6.4.8 MARC and Online Catalogue

There has been a sea change in cataloguing practice with the increasing use of computers, information communication technology and network communications. Cataloguing standards have been rationalised to suit the changing cataloging practice, which have been universally accepted and adopted.

In a machine-readable catalogue, entries are rendered in a format which permits input and storage on magnetic tape or magnetic disc for manipulation in a computer. MARC format, UNIMARC, Common Communication Format (CCF) are standard formats. There are communication formats that are different from physical formats. Access to the catalogue entries may be ‘off- line’ or ‘on-line’. ‘Off-line’ means that the computer can be used only at certain times. At the available computer time search must be made for a collection or batch of enquires. On-line systems, however, are linked directly to the computers which can be used immediately or at any time for processing enquiries and searching.

There are three major computer produced physical forms of library catalogues which are relevant to off-line access. These are:

- Printed form: Entries are printed in a conventional book format and are available in multiple copies.
- Card form: Each entry is transferred to one or more cards of standard catalogue card and are filed just as in a conventional card catalogue.
- Microform entries are transferred to microfiche and are used with appropriate microform readers.

In these forms, they are no more than the conventional catalogue forms. The only difference being their mode of production.

In an on-line catalogue, the entries are held in computer files and can be projected on the screen or printouts obtained. It is an organised accumulation of machine- readable accumulation of bibliographic records which are maintained on computer storage media facilitating for easy retrieval by library users and staff. It also helps keyword searching of title names and series names. The online catalogues are designed to access the information online about the library materials by the users with varying background, age, subject interests and computer literacy etc.

Library of Congress defined online catalogue as, "An online catalogue is an access tool and resource guide to the collections of a library or libraries, which contains interrelated sets of bibliographic data in machine-readable form and which can be searched interactively on a terminal by users."

The computer configuration needed for this catalogue is:

- A computer with keyboard and display units.
- Secondary storage facility
- Terminals wherever necessary.

Focus on the User

Online catalogues can be looked at from several perspectives. Many issues pertain to their design and using libraries, ranging from the number of terminals required to serve a given number of users, to the ergonomics of terminals placement, to the types and amount of information to be included in a catalogue, as well as its arrangement on a screen display. But in examining any from of library catalogue, online or other, the distinction between its implication for the internal operations of the library and its implications for library users must be kept in mind. Even though two perspectives do act on one another in the creation and use of the final product, the distinction is important. The internal-operations perspective has to do primarily with the production and maintenance of a catalogue; the library- user perspective has to with the presentation and use of it.

An online catalogue has the following characteristics:

- It is meant to be used by end-users with or without training in online searching.
- The database records are usually in the MARC format or derived from MARC format.
- The records are brief bibliographic descriptions enriched by a small number of controlled subject descriptor from Sears list or LC Subject Headings and classification number either from DDC or LC.

The online catalogues consist of information about the books like author (s), title pagination/volumes, publisher, year of publication, series, ISBN, subject descriptors and class number etc. Where as in case of periodicals this catalogue generally describes periodical as a whole. The contents of a periodical and the description of the article are recorded in the full-text databases.

Online catalogues are basically menu-driven and designed with little search options to help the novice searches. Some online catalogue, however operate with command language mode for more exhaustive searching. It can be done by specific item searching and subject searching.

The MARC record format was designed by the Library of Congress and the British Library also adopted it. The aim was to construct bibliographic records in machine-readable form and to facilitate reformatting for a wide variety of purpose, one of which was production of a record for the creation of library catalogues. This national effort made it possible to download from the database on magnetic tapes supplied by the

Library Congress or the British Library, to obtain the records for the stock of a specific library. Thus, a number of computers readable catalogues proliferated in USA, UK and other European countries.

Advantages

The machine-readable catalogue performs, all the functions of a library catalogue with greater efficiency and speed than any other form. We shall spell out these advantages: A computer readable catalogue:

- can be kept updated with speed and efficiency;
- can search for any bibliographical element, such as author, subject, publisher, price;
- is user friendly and new access points and search capabilities can be added as and when necessary;
- can easily be multiplied;
- union catalogue amongst several libraries by electronic communication is possible;
- easy to construct provided the cataloguer has expertise in the technique and the user feels very easy and has instant access to the pool of information and is well versed with necessary instruction;
- facilitates extensive search facility and possesses all the best qualities of both DDC and CC;
- interchange of catalogue records has led to greater consistency, uniformity and standardisation in catalogue records; and
- identifies the availability of the required item on the shelf or on loan if the computer system is linked to a computerised system for issue and return of the reading materials.

Using a computer, one can store and sort catalogue entries automatically. The resultant computer output can be utilised. For example, the computer 'printout' itself can be used as 'master' and reproduced by offset printing process to obtain any number of copies. Alternatively, the output on magnetic tape may be used as the catalogue to access entries directly and one can search and know from it the availability of any document in a library.

Disadvantages

1) Users should be trained to the computerised systems to exploit fully the capability of a machine-readable catalogue.
2) Both the users and library need to be trained to use the online catalogue
3) Interruption in power supply, breakdown of computer system, lack of proper knowledge of the users to operate the system poses problems for its best use.

As a result of rapid development in computer and communication technologies, a number of libraries in the world are switching over to computer readable catalogues and in the process, several on-line catalogue networks have developed and are available for public access in libraries.

Self Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answer given at the end of this Unit.

4) Discuss the advantages of microform catalogue.

....

....

....

....

....

5) Discuss the advantages of computerised catalogue.

....

....

....

....

....

6.4.9 CD-ROM Catalogue

It is a MARC-based compilation of bibliographic records distributed on CD-ROMs supported with software. CD-ROMs are optical discs and with the help of laser beams it can be written and recorded. It is an offline format like the microform that provides excellent search facilities. The introduction of offline, computer produced book and Computer Output Microform (COM) catalogue was a milestone in the history of library catalogue, in terms of production and maintenance. It reveals many of the problems in card catalogue production and effective maintenance. Besides, many cumbersome tasks that are fundamental to card catalogue were eliminated with the use of computers and machine-readable records.

Advantages

a) It is used because of its fine quality of portability and transportability.

b) It provides the best possible combination of large data-storage capacity, less production and high speed computer based search.

c) Union Catalogues can be prepared for sale purpose with the production of CD-ROMS

Disadvantages

1) Since it is an offline medium, it faces all the problems of conventional forms of offline catalogues.

2) In comparison with online retrieval, CD-ROM is slow.

6.4.10 Comparative Study of Physical form of Catalogues

Forms of catalogue	Easy to use	Easy to scan multiple copies	Easy to keep up-to-date	Easy to produce	Bulky	Easy to guide	Other factors
Printed	Yes	No	Yes	Yes	No	Yes	
Guard book	Yes	Yes, with some effort	Yes	Feasible but important	Yes	Yes	Time consuming to compile. Withdrawals are a problem
Card	Database	Yes	No	No	Yes	Fairly easy	One user can monopolise a complete section.
Sheaf	Yes	Yes	No, unless multiple entries to page which makes updating difficult.	Not usual but it is possible.	Yes	No	Binding mechanism can be awkward. Sheaf slips are filmsy.
Micro form	Database	Yes	Yes	Yes	No	Yes	Requires a machine to read.
Machine readable	Yes, but some training may be necessary.	Yes	No, but other search facilities compensate.	A number of access terminals may be provided.	A terminal as small as a portable typewriter.	Not applicable.	Power and equipment failure are problems to access

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

5) List the best features of the physically forms of library catalogue.

....

....

....

....

6.5 SUMMARY

In this Unit, the different types of physical forms of library catalogues have been enumerated and described. These include conventional forms like Bound register. Printed book form, Sheaf or loose leaf form and Card form, modern forms like Microform and Machine-Readable forms. The advantage and limitations associated with each of these different forms are explained.

With the advent of computers, the entire cataloguing process as well as the physical production of catalogues has undergone many changes, making library catalogue a versatile tool in a library. Modern machine-readable catalogues perform the functions of a library catalogue with more efficiency and speed. Whatever physical form a library catalogue may take, it should have the following essential features: ease of use, updating capability, browsing facility, easy production of multiple copies and occupies less space. A comparative statement of the different features of the physical forms of library catalogue is also given.

6.6 ANSWERS TO SELF CHECK EXERCISES

1) The physical forms of library catalogue are:

   Bound register or ledger form

   Printed book form

   Sheaf or loose-leaf form

   Card form

   Microform and

   CD-ROM catalogue

2) The two reasons why a card catalogue is more popular than a printed book from are:

   i) The card catalogue can always be kept updated which is the most essential requirement of a catalogue.

   ii) Addition or withdrawal of entries to the catalogue does not disturb or dislocate the existing arrangement of entries.

3) The advantages of a microform catalogue are:

   Microform catalogues are compact and occupy less space in libraries They are portable and immediate access to the catalogue is possible. Multiple copies of this catalogue can be prepared easily and inexpensively. These catalogues are very easy to use and maintain.

4) The advantage of computerised catalogue are as follows :

   Using a computer, the entries can be stored and sorted out automatically, but also the resultant computer output can be utilised in a number of ways. For instance, the computer printout itself can be used as a master copy that can be reproduced by offset printing process. Alternatively the output on magnetic tape may be processed into a microform. Eventually, the computer itself can be used as a master copy that can be reproduced by offset printing process. Alternatively the output on magnetic tape may be processed into a microform. Eventually, the computer itself may be used as the catalogue to access the entries directly.

   It can search any document in a library.

5) The best features of the physical form of library catalogue are:

   i) Ease of use, ii) Easy to update, iii) Permit browsing, iv) Easy production of multiple copies, v) Occupy very little space, and vi) Serve as a good guide providing multiple access points.

6.7 KEYWORDS

Bibliographic Record Format	:	The layout of presentation of bibliographical data of a document in a machine-readable form or in machine printout.
Cooperative Cataloguing	:	The sharing by a number of libraries the cost and labour of cataloguing to avoid the duplication of effort common to each.
Centralised Cataloguing	:	The cataloguing of documents by a central agency or library and the distribution there form of printed entries on catalogue card, or in machine-readable form, e.g. Library of Congress printed card service.
Database	:	Information stored on computer files and accessible via a remote terminals and telecommunication link.
Downloading	:	To capture data online from a remote host computer and transfer it to the store of an in house stand-alone system, e.g. a microcomputer, for processing.
Master	:	The plate or then stencil for duplication or for off-set printing, from which a multiple copies are made.
Microform	:	A generic term indicating any form of document in a reduced form whether on flat or on roll film or on microfiche.

6.8 REFERENCES AND FURTHER READING

Hunter, Eric J. *Computerised Cataloguing*. London: Clive Bingley, 1987. Print.

Tedd, Lucy A. *An Introduction to Computer-based Library Systems*. 2nd ed. New York: John Wiley, 1984.132. Print.

Rowley, Jennifer E. *Computers for Libraries*. 3rd ed. London: Library Association, 1993.245-246. Print.

Rowley, Jennifer E. *Organising Knowledge*. London: Gower, 1987. Print.

UNIT 7 FORMATS AND STANDARDS

Structure

7.0 OBJECTIVES

The objectives of this Unit are to introduce you to the formats and standards of bibliographic description of documents.

After reading this Unit you will be able to:

- explain the concept of a bibliographic record format;
- describe a machine-readable record format;
- discuss the development in and structure of ISBD format; and
- describe the structure of MARC, USMARC, UK MARC, ISO – 2709, CCF, UNIMARC and Z39.50.

7.1 INTRODUCTION

The advantage of computers is that they can rapidly and efficiently manipulate and retrieve any information/data, which is stored in machine-readable form. Their use in creation and development of bibliographic databases has raised the hope of developing a universal bibliographic system through the cooperation of several national and international organisations. Data generation and exchange take place at different levels, i.e. international, national, regional and local. Librarians have felt the need for and have arrived at standards to describe different kinds of bibliographic materials in catalogues and other bibliographic lists. The standardisation helps in deciding what elements to consider while describing a bibliographic material. It also extends to standardising their form and order of presentation including the punctuation and capitalisation. AACR2 is an example of such a standard. But it is not enough to prepare online catalogues where we need to code and structure bibliographic elements to utilise the capabilities of a computer. We have devised formats for this purpose and MARC is an example of such a format. MARC is a machine readable cataloguing format devised by Library of Congress. Many other such formats followed later some of which we will discuss in this Unit. These formats are also standards but standards in this Unit refers to bibliographic standards.

Machine-Readable Cataloguing (MARC), International Standard Bibliographic Description (ISBD) and AACR emphasised on preparation of uniform standardised bibliographic records. The standardisation of bibliographic description is aimed at compatibility and promotion of procedures and practices with a strategy for development towards universal bibliographic control which is a precursor to universal access to publications. Application of computers in the development of bibliographic databases has raised the hope of developing universal bibliographic system through the participation of several national and international organisations.

There is a need for exchange formats. i.e. formats that are designed specifically for the transfer of machine-readable bibliographic data between systems. This Unit discusses the formats available at national and international levels which can be adopted by agencies/ organisations or individuals as per their requirements.

7.2 FORMAT

In the UNISIST Reference Manual 'a (machine readable) bibliographic record is defined as a collection of information which pertains to a single document and which is stored in machine readable form as a self contained and unique logical structure'.

A machine-readable bibliographic record is arranged according to a particular format. Format conveys the notion of a formalised framework or structure, which will hold records of varying content according to certain set of rules or conventions controlling the representation of data. These rules may be unique to a system, or shared with other systems.

7.2.1 Bibliographic Record Formats

Bibliography record formats are used to describe the arrangement or structure of computer readable record of bibliographic items. The essential components of a bibliographic record format are:

- A description of document itself in relation to author, title, publisher, etc.
- Choice of elements to use as approach points for retrieval of the record.
- A unique record identifier of the document.
- Descriptors representing the subject matter of the document selected from standard list of subject headings or thesaurus.

7.2.2 Types of Formats

Formats can be of two types:

- Internal / Local format
- Exchange/ Communication/ Interchange format

Internal / Local formats

Internal formats are so called because they are internal / local to a software system. They can be changed specifically as per the needs of the local system and do not have to conform to any external standards. For the internal format (the identifiers used and their records structure), system designers can adopt any conventions that they wish. Rules for input are more restricted, being constrained by existing cataloguing codes and practices and need to be compatible with other systems or with a standard exchange format. Their main aim is economy and efficiency in processing, while the ability to convert, if necessary, from the local format to an exchange format is an extra advantage. This conversion can be done by a computer program. An agency may, therefore, adopt its own internal processing format, but at the same time retain system compatibility with a national or international exchange format. Internal formats can be structured in infinite varieties or ways which depend on the software used.

Exchange / Interchange Communication Formats

Exchange formats are also known as interchange/communication formats. Exchange formats are used for exchange of records between systems. Systems should be sufficiently flexible to cope with the needs of many different software systems. Ideally, they should facilitate the exchange of data which are to be used in wide range of different bibliographic applications, from the production of traditional catalogue cards to records in databases which are used for online access. A major problem for designer of exchange format is the lack of international agreement on standards for constructing bibliographic records.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

1) What do you understand by "format'? What are the two types of formats?

....

....

....

....

7.3 EXCHANGE FORMATS: STRUCTURE AND CONTENT

Bibliographic data formats adopted for exchange of data consist of three basic components. They are:

1) **A defined physical structure:** Rules for the arrangement (on a computer storage medium) of data to be exchanged. This may be linked to a container or carrier into

which data many be placed. The carrier remains constant although the data change from record to record.

2) **Content designator:** Codes to identify the different data elements in the record e.g., author, title, scale of map, starting data of journal, etc.

3) **Rules:** Content of the records governed by rules for the formulation of the different data elements, very closely tied up with content designator. The data elements separately identified by the codes in the exchange format have to be defined, not only in terms of content but also in form, if the records are to be suitable for use by another agency.

7.3.1 Structure

Records in automated systems may be of fixed length or variable length. Fixed length records in a file must contain the same field and any given field must always be of the same length from on record to another. If a field is repeatable, it must be repeated the same number of times in each record. Variable length records may be variable because they do not contain the same number of fields (a given field may not occur in every record) or because a given field is not always of the same length. Bibliographic records do not fit naturally into the fixed length pattern.

An example of a database whose records can be treated at fixed length is a mailing list for officials within an organisation. In each record there can be fields for each of name, job title, province, state, postal code. Few of any of these data elements will always be of the same length, but most of them will have a similar length. Each data element is allocated a number of characters and any space not used will be filled with blank space. The system designer has to allocate to each data element a suitable number of characters which will not be too small to accommodate the majority of records, but not too large also, to waste space by having records full of spaces. Most records in a mailing list will include all the fields mentioned above and there will not be too many fields in a mailing list database which are completely filled with spaces. In bibliographic records, the situation is different.

- Many fields occur only from time to time, for example cover title, ISSN, edition. The edition, appears in bibliographic records only when the document contains specific edition statement or when it is known to be other than a first edition.
- Many fields occur for a variable number of times in a record. For example: A document may have one author or it may have any number of joint authors. The same document may be a member of more than one series. A record may be assigned more than one subject descriptor.
- A data element may vary in length between records. This is true in most systems which do not code their data but record them in ordinary language.

Content Designator

Content designators are used to denote the identity of each data field (instead of identifying data fields by position) within a record. It may be grouped at the beginning of the record with 'pointers' to the data they identify. The storage of the computer file invariably uses the method of indexing the location of the files at the beginning of the file. The computer calculates from the given starting character position of each file the likely position of that file on the tape/disk, moves the same very quickly to a position just before the start of that file at which point the tape / disk moves more slowly in order to locate the data. This index is known as a directory.

7.4 BIBLIOGRAPHIC STANDARDS

The basic need of bibliographic standard is to have bibliographical control which facilitates comprehension and transfer of bibliographic data between bibliographical agencies at national and international levels. It is highly necessary to develop and design a suitable and feasible standard record format uniformly acceptable to all agencies involved in exchange of information.

7.4.1 International Standard Bibliographic Description (ISBD)

Introduction

The developments that took place in library and information science from the early 1960s revolutionised the concept, scope and purpose of bibliographic description. The exponential growth of literature resulted in the development of different tools such as bibliographies, union catalogues, indexing and abstracting services to ensure effective bibliographic control. The centralised and cooperative cataloguing efforts further intensified the need to have a standard bibliographic format that will promote resource sharing among libraries. The efficient information retrieval system must have a format of bibliographic description that suits its requirements. The micrographic, computers and telecommunication technologies brought a sea change in the library environment. The changes in the pattern of library and information services with the introduction of new technologies rendered the traditional manual oriented bibliographic description standards and practices inadequate and thus the need for newer bibliographic standards.

Origin

The **International Standard Bibliographic Description (ISBD)** dates back to 1969, when the IFLA Committee on Cataloguing (subsequently renamed the Standing Committee of the IFLA Section on Cataloguing, now known as the Standing Committee of the IFLA Cataloguing Section) sponsored an International Meeting of Cataloguing Experts. This meeting produced a resolution that proposed creation of standards to regularise the form and content of bibliographic descriptions. As a result, the Committee on Cataloguing put into motion work that ultimately would provide the means for a considerable increase in the sharing and exchange of bibliographic data. This work resulted in the concept of the International Standard Bibliographic Description (ISBD), which has now endured for about 44 years. The individual formats to which the ISBD concept has been applied are now used by bibliographic agencies, national and international cataloguing codes and cataloguers in a wide variety of libraries throughout the world, because of their potential for promoting record sharing.

The first of the ISBDs was the *International Standard Bibliographic Description for Monographic Publications (ISBD(M))* that appeared in 1971. By 1973, this text had been adopted by a number of national bibliographies and with translations of the original English text into several other languages, had been taken into account by a number of cataloguing committees in redrafting national rules for description. Comments from users of the ISBD(M) led to the decision to produce a revised text that was published in 1974 as the "First standard edition."

In 1975, the Joint Steering Committee for Revision of the Anglo-American Cataloguing Rules proposed to the IFLA Committee on Cataloguing that a general international standard bibliographic description suitable for most common types of library resources should be developed. The ISBD (G) was published in 1977. The ISBD (M) was then revised to bring it into line with the ISBD (G), and the "First standard edition revised" was published in 1978.

Keeping in view the peculiarities of different forms of materials such as serials, cartographic materials, non-book materials, printed music, antiquarian, audio visual materials, computer files, etc. ISBD(M) followed by publication of series of specialised ISBDs, viz., ISBD (S), ISBD (CM), ISBD (NBM) ISBD (PM), ISBD (A), ISBD (AVM) and ISBD (CF) [S-Serials; CM- Cartographic Material; NBM- Non- Book Material; PM- Printed Music; A- Antiquarian; AVM- Audio- visual Material; CF- Computer files]. An integrated general format for all types of documents, called General International Standard Bibliographic Description ISBD (G) was also brought out to bring harmony among various ISBDs and to remove incompatibilities among them.

Subsequently, ISBD Review Committee was formed by the Standing Committee of the IFLA Section on Cataloguing met in 1981 to make plans for reviewing and revising the ISBDs covering monographic publications, serials, cartographic materials, and non-book materials. There were three major objectives set out for this project to:

1) harmonise provisions among the ISBDs, achieving increased consistency;

2) improve examples; and

3) make the provisions more applicable to cataloguers working with materials published in non-roman scripts. In addition, two narrower objectives motivated this particular revision effort: (a) to review the use of the equals sign; and (b) to consider proposals regarding the ISBD(NBM) emanating from specialist groups such as the International Association of Music Librarians (most prominent of which was to remove "machine-readable data files" as a format from this standard). By the end of the 1980s, this project had been completed.

The consolidated edition of the ISBD was published in 2007. IFLA's ISBD Review Group is responsible for maintaining the ISBD.

Purpose

The ISBDs seek to serve the following primary purpose: First, and of greatest importance, they are intended to make it possible to interchange records from different sources. As subsidiary purposes, the ISBDs, secondly, have assisted in the interpretation of records across language barriers, since bibliographic items in each record can be easily identified through specialised punctuation and its place for the record so that records produced for users of one language can be interpreted by users of other languages. Thirdly, they have facilitated the conversion of bibliographic records to electronic form. Fourthly the ISBD was to provide a standard form of bibliographic description that could be used to exchange records at international level. This would support IFLA's program of Universal Bibliographic Control.

Structure of an ISBD Record

The ISBD prescribes eight areas of description. Each area, except area 7, is composed of multiple elements. For example, area 1 includes the title proper, general material designation GMD, other title information, parallel title, and statements of responsibility. Elements and areas that don't apply to a particular resource are omitted from the description. Standardised punctuation (colons, semicolons, slashes, dashes, commas, and periods) is used to identify and separate the elements and areas. The order of elements and standardised punctuation make it easier to interpret bibliographic records when one does not understand the language of the description. These eight areas are:

1) title and statement of responsibility area;

2) edition area;

3) material or type of resource specific area (for example, the scale of a or the numbering of a periodical);

4) publication, production, distribution, etc., area;

5) physical description area (for example: number of pages in a book or number of CDs issued as a unit);

6) series area;

7) notes area; and

8) resource identifier (e.g. ISBN, ISSN) and terms of availability area.

Heading

Title Proper = Parallel Title : Sub-title / Statement of authorship; Other authors. – Edition Statement; Edition author statement. – Material or type of resource specific area.- Place of publication : Publisher, Year of publication.

Preliminary pages, Textual pages ; Plates : Illustration;

Size.– (Series: Sub-series; Series number).

Notes.

ISBN

The ISBD (G) prescribed preceding area and punctuation element for elements

Note each area other then the first, is preceded by a point, space, dash, space (. -) (Appendix - 1)

Recent Developments

ISBD and FRBR (Functional Requirements for Bibliographic Records)

In the early 1990s, the IFLA Section on Cataloguing with the cooperation of the Section on Classification and Indexing set up the IFLA Study Group on the Functional Requirements for Bibliographic Records (FRBR). One immediate consequence of this development was the decision to suspend most revision work on the ISBDs while the FRBR Study Group pursued its charge to "recommend a basic level of functionality and basic data requirements for records created by national bibliographic agencies." In 1998, the FRBR Study Group published its Final Report after its recommendations that were approved by the IFLA Section on Cataloguing Standing Committee.

The objective of this "second general review project" was to ensure conformity between the provisions of the ISBDs and FRBR's data requirements for the "basic level national bibliographic record." It also involved a mapping between ISBD elements and FRBR attributes and relationships, developed by Tom Delsey in 2004 and the publication of ISBD (ER) electronic resources and later, the ISBD for 'Continuing Resources'.

The ISBD Review Group has taken initiation to discuss and examine the following:

- to clarify the purpose of area 6 and its relation with area 1 in ISBD(CR) and ISSN: identification or transcription;
- to verify the compatibility of sources of information recommended or prescribed in all ISBDs for area 6 and for area 1 in ISBD(CR) and ISSN, and

- to propose a common phrasing for area 6 in all ISBDs.

7.4.2 ISO - 2709

Introduction

Bibliographical information plays an important role information retrieval for the research community particularly in the field of science and technology. But during the bibliographical information exchange certain problems arise and more when the information interchange is on magnetic tape or CD-ROM. Different international organisations such as UNESCO/PGI, UNISIST, ICSU-AB, IFLA, ISO have taken many steps towards the standardisation of bibliographic exchange formats. The process of standardisation follows a set of codes given by International Standard Organisation (ISO).

Purpose

The major purpose of standardisation is to:

- permit the exchange of bibliographic records between groups of libraries and abstracting and indexing services;
- permit a bibliographic agency to manipulate bibliographic records received from both libraries and abstracting and indexing services;
- serve as the basis of a format for an agency's own bibliographic database by providing a list of useful data elements;and assist the development of individual systems.

Basic Structure

The ISO-2709 describes a generalised structure specially designed for communication between the data processing systems and not intended for use as a processing format within systems. However, this standard was designed specifically for recording and processing data on magnetic tapes and its structures can be used for other data carriers. (see Appendix –2)

- The ISO-2709 format, also known as MARC, is a standard bibliographic information view. The records in the ISO-2709 format have a fixed structure and consist of: **record marker** of 24 characters;
- **data directory**, consisting of 3 numerical marks for each data field, field size and initial symbol position, relating to the first data field; and
- **variable-length data fields** separated from each other by a field divider.

An ISO-2709 record has **three** sections:

- **Record label** : The first 24 characters of the record. This is the only portion of the record that is fixed in length. The record label includes the record length and the base address of the data contained in the record. It also has data elements that indicate how many characters are used for indicators and subfield identifiers. (See Variable Fields, below)
- **Directory** : The directory provides the entry positions to the fields in the record, along with the field tags. A directory entry has four parts and cannot exceed nine characters in length:
  - Field tag (3 characters)

- Length of the field
- Starting character position of the field
- (Optional) Implementation-defined part

- **Data fields** : A string containing all field and subfield data in the record.

Note that although tags are often displayed as labels on bibliographic fields and each bibliographic field has an associated tag, the tags are stored in the directory not in the bibliographic field.

Fields

There are **three** kinds of fields in the ISO-2709 record:

- Record identifier field – identifying the record and assigned by the organisation that creates the record. The record identifier field has tag 001.
- Reserved fields – Reserved fields supply data which may be required for the processing of the record. Reserved fields always have a tag in the range 002-009 and 00A to ZZZ.
- Bibliographic fields – these are in the range 010-999 and 0AA to ZZZ. The bibliographic fields contain data and a field separator. They can also have these optional sub-parts:

Indicator (0-9 characters, as coded in the Leader) – Indicators generally provide further information about the contents of the field, the relationship between the field and other fields in the record, or about action required in certain data manipulation processes (including display labels).

Identifier (0-9 characters) – This identifies data within the bibliographic field. Where used, identifiers are composed of a delimiter (1 char, ? of ISO 646) and an identifying code (1-9 character, as defined in the leader), plus a variable length string containing the data.

Record separator – A single character (of ISO 646).

Example

MARC 21 library cataloguing record was coded in ISO-2709 format. MARC 21 is an instance of ISO-2709 that has the following characteristics:

- tags are in the range 002-999 only;
- there is a two-character indicator on each field, and each character is a separately defined data element; and
- the identifier within data fields (called "subfield code" in MARC 21) is a single ASCII character preceded by of ISO 646.

Principle and Codings

The standard ISO-2709 (standard AFNOR NF Z 47300, December 1987) makes it possible to present any structured bibliographic record a large variety of formats, in particular the MARC or UNIMARC or CCF formats.

A recording comprises in ISO-2709 the following parts:

- **The guide**, continuation of 24 numbered natures from 0 to 20
- **The directory**, which comprises a variable succession of numerical characters
- **Bibliographical data** themselves.

The ISO-2709 standard doesn't limit the content of particular data fields, so this standard offers different bibliographic data view formats, such as USMARC, UNIMARC, RUSMARC, MARC21, UKMARC, DANMARC, etc.

Advantages of ISO-2709

Standards are documented agreements containing technical specifications or other precise criteria to be used consistently as rules, guidelines, or definitions of characteristics, to ensure that materials, products, processes and services are fit for their purpose. Hence ISO-2709 (Format for Bibliographic Information Interchange on Magnetic Tape) has many advantages. Some of the important advantages are given below:

- It provides a small number of mandatory data elements, which are recognised by all sectors of the information community as essential in order to identify an item.
- It gives mandatory data elements that are sufficiently flexible to accommodate varying descriptive practices.
- It also provides a number of optional elements, which may be useful to describe an item according to practices of the agency, which creates the record.
- It provides a mechanism for linking records and segments of records without imposing on the originating agency any uniform practice regarding the treatment of related groups of records or data elements.

7.5 STANDARDS FOR MACHINE-READABLE RECORD FORMAT

7.5.1 MARC and MARC 21

What is a MARC Record?

A MARC record is a Machine-Readable Cataloging record. It is a format standard for the storage and exchange of bibliographic records and related information in machine.

Machine-readable: "Machine-readable" means that one particular type of machine, a computer, can read and interpret the data in the cataloguing record.

Cataloguing record: "Cataloguing record" means a bibliographic record, or the information traditionally shown on a catalogue card. The record includes (not necessarily in this order): 1) a description of the item, 2) main entry and added entries, 3) subject headings, and 4) the classification or call number.

1) ***Description:*** Librarians follow the rules in *Anglo-American Cataloguing Rules*, 2nd ed., 2002 revision to compose the bibliographic description of a library item. It includes the title, statement of responsibility, edition, material specific details, publication information, physical description, series, notes, standard numbers and keywords.

2) ***Main entry and added entries:*** AACR 2R also contains rules for determining "access points" to the record (usually referred to as the "main entry" and "other added entries"), and the form these access points should take. Access points are the retrieval points in the library catalogue where patrons should be able to look up the item.

In other words, the rules in AACR 2R are used to answer questions such as: For this book, should there be entries in the catalogue for more than one author or more than one title? Should the title of the series be noted? How should the author's name be written? Is this a "title main entry" item (no author)?

3) ***Subject headings (subject added entries):*** The librarian uses the *Sears List of Subject Headings* (SLSH), the *Library of Congress Subject Headings* (LCSH), or some other list of standard subject headings to select the subjects under which the item will be listed. Use of an approved list is important for uniformity and consistency, to ensure that all items on a particular subject are found under the same heading and therefore in the same place in the catalogue.

4) ***Classification/Call number:*** The librarian uses a Dewey Decimal Classification or Library of Congress Classification schemes to select the class/call number for an item. The purpose of the class/call number is to place items on the same subject together on the same shelf in the library. Most items are sub-arranged alphabetically by author. The second part of the call number usually represents the author's name, facilitating this sub-arrangement.

Importance

MARC format is most necessary to any type of library for:

- Prevention of duplication of work;
- Better sharing of bibliographic resources; and
- Bibliographic control at international level.

MARC : Terminology

To describe a MARC record, we need to first understand the terminology associated with that.

Signpost: The information from a catalogue card is not simply typed into a computer to produce an automated catalogue. The computer needs a means of interpreting the information on a catalogue record. The MARC record contains a guide to its data, which is called as Signpost, before each piece of bibliographic information.

The following example gives us a clear idea about Signpost.

Signpost	Data
Main entry, personal name with a single Surname The name	 Brenner, Richard J
Title and statement of responsibility Title proper Statement of responsibility	 Make the Team Richard J Brenner
Edition area Edition statement	 1st ed.
Publication, distribution, etc. area Place of publication Name of publisher	 Boston Brown
Physical description area Pagination Illustration	 127 p. ill.

Record with Textual Signpost

Fields

Each bibliographic record is divided into fields. Means there is a field for the author, title, etc. These fields are again subdivided into one or more "subfields". Each field is represented by a "tag".

Indicator

Indicators are one-digit numbers. Each tag is followed by two character positions. In some fields first or second position is used, in some fields both are used and in some fields (like 020, 300) neither is used. When an indicator position is not used, that indicator is referred to as "undefined" and it is marked by character position "#". Each indicator value is a number from 0 to 9.

For example:

Tag Indicator Title Field

245 14 a The Handbook of Library and Information Science.

In the above example "1" is the first indicator and "4" is the second indicator. The value of first indicator in the field i.e., "1" indicates that there should be separate title entry in the catalogue, means title added entry should be prepared and title should be added to the tracing in the card catalogue. A first indicator value of "0" would mean that a title main entry is involved and no additional tracing for the title would be required.

The second indicator i.e., "4" displays the number of characters at the beginning of the field (including spaces) to be disregarded by the computer in the sorting and filling process. For the title *The Handbook of Library and Information Science,* the second indicator is set to "4" so that the first four characters ("T", "h", "e", and the space) will be skipped and the title will be filed under "handbook".

Variable Fields

The data content of a record is divided into variable fields. The MARC 21 formats distinguish two types of variable fields : variable control fields and variable data fields. "Variable control fields" consist of data and a field terminator. The 00X fields (001 to 009) in the MARC 21 formats are variable control fields. "Variable data fields" contain the textual information that describes the bibliographic item being catalogued. All fields except 00X are variable data fields.

Subfields, Delimiters and Subfield Codes

Within a field, there may be several related pieces of data. Each type of data within the field is called a subfield. Each subfield is preceded by a subfield code. Subfield codes are always lower case letters which is preceded by a delimiter. A delimiter is used to separate subfields.

For example :

In the field publication, distribution, etc. (Tag 260), the most commonly used subfields are :

a) Place of publication, distribution

b) Name of publisher, distributor

c) Date of publication, distribution

260 ## $a New Delhi :
$b Allied Publishers,
$c 2001.

Here "$" is delimiter.

MARC : Structure

The format of MARC structure includes :

Leader Record Directory Variable Field

Leader

The leader is the first 24 characters of the records. Each position has an assigned meaning, but much of information in the leader is for computer use. Leader provides information about the ensuing record such as the total length of the record, type of the record code and the bibliographic level.

Record Directory

Immediately following the leader, the directory begins. Directory tells what variable fields are in the record and where they are placed. There is a 12 characters record directory for each variable field. The record directory helps in the retrieval of select fields from the record. The directory ends with a field terminator character.

The MARC format, both by its structure and the content designators has caused a revolution in the design of bibliographic database, worldwide. The existence of many formats based at least on the principles behind MARC format proves the impact that MARC had. The creators of various databases such as INIS, AGRIS and INSPEC have used MARC structure as the basis for the creation of their own communication format. Why one standard? One could devise her/his own method of organising the bibliographic information, but s/he would be isolating her/his library, limiting its options, and creating much more work for himself. Using the MARC standard or format prevents duplication of work and allows libraries to better share bibliographic resources. Choosing to use MARC enables libraries to acquire cataloguing data that is predictable and reliable. If a library were to develop a "home-grown" system that did not use MARC records, it would not be taking advantage of an industry-wide standard whose primary purpose is to foster communication of information.

Using the MARC standard or format also enables libraries to make use of commercially available library automation systems to manage library operations. Many systems are available for libraries of all sizes and are designed to work with the MARC format. Systems are maintained and improved by the vendor so that libraries can benefit from the latest advances in computer technology. The MARC standard or format also allows libraries to replace one system with another with the assurance that their data will still be compatible.

7.5.2 USMARC

Introduction

USMARC formats are standards for the representation and communication of bibliographic and related information in machine-readable form. A USMARC record involves three elements: the record structure, the content designation and the data content of the record. The structure of USMARC records is an implementation of national and international standards, e.g., Bibliographic Information Interchange (ANSI Z39.2) and Format for Bibliographic Information Interchange on Magnetic Tape (ISO-2709). Content designation, the codes and conventions established to identify explicitly and characterise further the data elements within a record and to support the manipulation of those data, is defined in the USMARC formats.

The content of most data elements is defined by standards outside the formats, e.g., Anglo-American Cataloguing Rules, Library of Congress Subject Headings, National Library of Medicine Classification. The content of other data elements, e.g., coded data is defined in the USMARC formats. A USMARC format is a set of codes and content designators defined for encoding a particular type of machine-readable record. USMARC formats are defined for the following types of data: bibliographic, holdings and authority.

USMARC format for bibliographic data contains format specifications for encoding data elements needed to describe, retrieve, and control various forms of bibliographic material. The USMARC Format for Bibliographic Data is an integrated format defined for the identification and description of different forms of bibliographic material. USMARC specifications are defined for books, archival and manuscripts control, computer files, maps, music, visual materials and serials. With the full integration of the previously discrete bibliographic formats, consistent definition and usage are maintained for different forms of material. USMARC Format for Holdings Data contains format specifications for encoding data elements pertinent to holdings and location data for all forms of material. USMARC Format for Authority Data contains format specifications for encoding data elements that identify or control the content and content designation of those portions of a bibliographic record that may be subject to authority control. USMARC formats are maintained by the Library of Congress in consultation with various user communities.

USMARC Format: Utility

USMARC formats are communication formats, primarily designed to provide specifications for the exchange of bibliographic and related information between systems. They are widely used in a variety of exchange and processing environments. As communication formats, they do not mandate internal storage or display formats to be used by individual systems. USMARC formats, particularly the bibliographic and authority formats, were developed to enable the Library of Congress to communicate its catalogue records to other institutions. The formats have had a close relationship to the needs and practices of United States libraries. USMARC formats were designed to facilitate the exchange of bibliographic and related information on magnetic tape within the United States. An attempt has been made to preserve compatiblity with other national and international formats, e.g., CANMARC and UNIMARC. Lack of international agreement on cataloguing codes and practices has made complete compatibility impossible. National agencies in the United States and Canada are given special emphasis and consideration in the formats because they serve as sources of authoritative cataloguing and as agencies responsible for certain data elements. USMARC format provides content designation only for data that are applicable to all copies of the bibliographic entity described. USMARC format provides limited content designation for the encoding of this information and for identifying the holding institution, e.g., subfield $5 in the 700-740 added entry fields in the bibliographic format. Information that does not apply to all copies of a title and is not of interest to other institutions, is coded in local fields.

Structural Features

USMARC formats are an implementation of the Bibliographic Information Interchange (ANSI Z39.2). The formats also incorporate other relevant ANSI standards, e.g., Magnetic Tape Labels and File Structure for Information Interchange (ANSI X3.27). All information in a USMARC record is stored in character form. USMARC communications records are coded in Extended ASCII,as defined in USMARC Specifications for Record Structure, Character Sets, Tapes. The length of each variable field can be determined either from the length-of-field portion of the directory entry or from the occurrence of the field terminator character.

Content Designation

The goal of content designation is to identify and characterise the data elements that comprise a USMARC record with sufficient precision to support manipulation of the data for a variety of functions. USMARC content designation is designed to support functions that include:

Display the formatting of data for display on a CRT, for printing on 3x5 cards or in book catalogues, for production of COM catalogues, or for other visual presentation of the data. Information retrieval, identification, categorisation and retrieval of any identifiable data element in a record. Some fields serve multiple functions. For example, field 245 (Title Statement) serves both as the bibliographic transcription of the title and the statement of responsibility and as an access point for the title. USMARC formats provide for display constants. A display constant is a term, phrase, and/or spacing or punctuation convention that may be system generated under prescribed circumstances to make a visual presentation of data in a record more meaningful to a user.

Organisation of the Record

A USMARC record consists of three main sections: the leader, the directory, and the variable fields which is shown below:

Leader	Directory	Variable Control Fields	Variable Data Fields

Fig. 7.1: Sections of a USMARC Record

Leader: The leader consists of data elements that contain coded values and are identified by relative character position. Data elements in the leader define parameters for processing the record. The leader is fixed in length (24 characters) and occurs at the beginning of each USMARC record and provides general information about the record.

This part of the record consists of data elements to process the record. It is a fixed field consisting of a total 24 characters with positions counted form 0 to 23.

Character position

0-4 Logical record length

0-5 Record status

0-6 Type of record

0-7 Bibliographic level

8-9 Blank

10-11 Indicator count & subfield code count

12-17 Base address

17-19 Implementation defined positions

20-22 Entry map

23 Undefined & set to ‘o’

They are used to indicate nine types of data such as length, status, type and bibliographical level of the record and blank characters, subfield code, base address, etc. Length indicates the total number of characters used in the records, status tells whether it is new, changed or deleted record, type tells whether it is printed or not, and bibliographical level states whether it is an analytic, monograph, serial or a collection and so on.

Directory: The directory contains the tag, starting location, and length of each field within the record. Directory entries for variable control fields appear first, in ascending tag order. Entries for variable data fields follow, arranged in ascending order according to the first character of the tag. The order of the fields in the record does not necessarily correspond to the order of directory entries. Duplicate tags are distinguished only by location of the respective fields within the record. The length of the directory entry is defined in the entry map elements in Leader/20-23. In the USMARC formats, the length of a directory entry is 12 characters. The directory ends with a field terminator character.

Variable fields: The data content of a record is divided into variable fields. USMARC formats distinguish two types of variable fields: variable control fields and variable data fields. Control and data fields are distinguished only by structure. The term fixed fields is occasionally used in USMARC documentation, referring either to control fields generally or to specific coded-data fields, e.g., 007 (Physical Description Fixed Field) od 008 (Fixed-Length Data Elements) which is enumerated below:

001 Control number

002 Subrecord map of directory

003 Subrecord relationship

004 Related record directory

005 Date and time of latest transaction

006 Fixed length data elements

007 Physical description fixed field

008 Fixed length data elements: Coded information useful for retrieval and manipulation of data.

009 For local use

Data fields in USMARC follow the variable control field. Their definition depends on the bibliography format blocks, authority format blocks and holding format blocks. However, here the bibliographic format block is mentioned. Bibliographic format blocks 0XX = Control information, numbers, and codes.

XX = Main entry

2XX = Titles and title paragraph (title, edition, imprint)

3XX = Physical description, etc.

4XX = Series statements

5XX = Notes

6XX = Subject access fields

7XX = Added entries other than subject

or series; linking fields

8XX = Series added entries, etc.

9XX = Reserved for local implementation

Variable Control Fields

The 00X fields in the USMARC formats are variable control fields.

Variable control fields consist of data and a field terminator. They contain neither indicators nor subfield codes.

Variable control fields contain either a single data element or a series of fixed-length data elements identified by relative character position.

Variable Data Fields

All fields except 00X are variable data fields.

Four levels of content designation are provided for variable data fields in ANSI Z39.2:

a) a three-character tag, stored in the directory entry;

b) indicators stored at the beginning of each variable data field, the number of indicators being reflected in Leader/10 (Indicator count);

c) subfield codes preceding each data element, the length of the code being reflected in Leader/11 (Subfield code count); and

d) a field terminator following the last data element in the field.

Indicators

Hence it is clear that variable data fields are made up of single as well as groups of data elements. Each variable data field consists of indicator, subfield codes, data elements and field terminator. Further a tag is assigned to each variable data field and that tag is stored in the directory. For example 100 is the tag for main entry-personal name. Indicator is a code supplying additional information about the field and it is located at the beginning of the field.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

2) Enumerate the different types of USMARC Communication Formats and discuss their structure.

....

....

....

....

7.5.3 UKMARC

Introduction

UKMARC format was developed primarily to facilitate the production of the printed British National Bibliography (BNB) and thus closely reflects the cataloguing practice of the British Library in its interpretation of AACR2 and other standards.

A distinction can however be drawn between the UKMARC format as a national format and its use by the national library in preparing records for the national bibliography, and UKMARC Manual seeks to make such a distinction.

MARC is an acronym for Machine Readable Cataloguing. The MARC format was primarily developed as a convenient way of storing and exchanging bibliographic records. It has since been extended to include non-bibliographic forms of library material, such as maps and music scores. There is also a MARC format for the storage and exchange of authority records.

The original MARC format was developed by the Library of Congress in the mid-1960s. A pilot project, known as MARC I, was set up to investigate the feasibility of producing catalogue data in machine-readable form. Similar work was being done in the United Kingdom by the Council of the British National Bibliography Ltd. which had set up the BNB MARC Project to examine whether the production of the printed British National Bibliography (BNB) could be simplified by using machine-readable bibliographic records.

Genesis of the UKMARC format

In 1969 the *British National Bibliography* introduced a MARC tape service for current British books and a machine-readable version of the weekly printed BNB was published. UKMARC primarily reflected the requirements of BNB but was gradually being developed to cater for use by the wider information community. By 1975 when the first edition of the UK MARC Manual was published, UKMARC had become a national communications format and its use in the production of records for BNB represented just one particular application of the format.

Since the early 1970s an extended family of more than twenty national MARC formats grew up. Attempts were made to preserve compatibility between national formats, but differing national requirements made complete compatibility impossible and differences in data content meant that editing was required before records could be exchanged. The solution to the problem of incompatibility was to create an international MARC format. UNIMARC was given which facilitates the exchange of records created in any MARC format.

UKMARC and AACR2

The publication of the second edition in 1978 was a landmark in standardising procedures for the description of books and other items and for the construction of headings and references and in addressing the needs of machine processing.

The British Library decided to adopt AACR2 with effect from 1981 and the changes this made necessary to UKMARC were incorporated into the second edition of the UKMARC Manual in 1980. Throughout the 1980s there was a growing international commitment to networks and shared cataloguing based on AACR2 which by then had been adopted by a number of non-English speaking countries. Although the basic concepts of AACR2 did not change, three sets of rule revisions were required to take account of international usage, new library materials, and rapid technological change. This process was consolidated with the publication of the 1988 revision of AACR2. In 1994 a set of amendments, known as Amendments 1993, was published comprising clarifications and slight changes to existing rules which the British Library applied to BNBMARC records.

Structure of the UKMARC format

UKMARC format provides a record structure for the description of bibliographic and other items and identifies and describes the data elements within that record. UKMARC format has been defined for books, serials, cartographic items, music and audio-visual materials, but originally UKMARC was used for the exchange of bibliographic records on magnetic tape and the way in which records are organised reflects this.

Because computers can only store and manipulate numbers, the data content within a UKMARC record, i.e. a traditional catalogue record as defined by AACR2, consists of strings of characters arranged in fields. UKMARC format uses the Extended ASCII (American Standard Code for Information Interchange) character set to determine the numeric value to be used for each character of data. The data content of these fields, or the subfields within them, may be indicated by either a tag or by its position in the record. The end of each field is indicated by a character usually represented as a # (a special character with no ASCII equivalent).

UKMARC format uses tags consisting of three numeric characters, e.g. 100, and, wherever possible, parallel meanings have been preserved between tags, e.g. between fields 100 and 700. Tags 950-999 and all tags ending in 9 are reserved for local use. UKMARC specifies no structure or meaning for local fields. Theoretically, all fields except 001 may be repeated, but the nature of the data often precludes repetition, e.g. a bibliographic record may contain only one main title.

There are two types of fields: fixed length fields, in which coded characters in particular character positions give information about the item, e.g. whether the item is a monograph or a serial; and variable-length fields, in which the sense of the data is not dependent on a particular character position, e.g. title.

Fields are arranged in functional blocks, sometimes referred to as areas. These blocks organise the data according to its function in a traditional catalogue record:

Blocks Definition

001-009

Control fields

010-099

Coded and other information

100-244

Main entry access points

245-299

Title and title paragraph

300-399

Physical description

400-499

Series statements

500-599

Notes

600-699

Subject access points

700-799

Added entry access points

800-899

Series access points

900-945

References

946-999 *Local fields*

Fields 001-009 are termed control fields. Unlike bibliographic fields, control fields do not contain either indicators or subfield codes. Control fields contain either a single data element or a series of fixed length data elements identified by their relative character position.

The data content of a field is further defined by the use of the following elements:

Indicators

Each variable length bibliographic data field in UKMARC records is introduced by two indicators, each in the form of a single character in the range 0-9, A-Z (giving a possible total of 35 values). These digits 'indicate' the relationship between the field in which they occur and other fields in the record, and about how the data in the field should be manipulated for catalogue production. For example, the 100 tag, followed by indicators .10, i.e. 100.10, shows that the person is entered under the last element of the name.

Subfields

Subfield codes distinguish data elements within a field that require (or might require) separate manipulation. Subfield codes in the UKMARC format consist of an alphabetic lower case character following a $. Subfield codes are defined independently for each field. Each subfield code has its own punctuation and typographic value, and therefore no punctuation needs to be input at subfield boundaries. The small amount of additional punctuation used, particularly within notes fields, follows AACR2.

Levels

Levels are 1-digit numbers input following the indicators and separated from them by a colon, e.g. 700.10:1, used to show that a work included within a publication has been catalogued as a bibliographic entity in its own right. If a separate entry is required for each of the three levels in that collection. Levels in this sense are not used in BNBMARC records (added entries, i.e. fields 700-745, are used instead)

MARC 21

In 2000, the British Library after consultation with the UK library and information community unanimously decided for a move from UKMARC, the national cataloguing format maintained by the British Library since 1975, to the MARC 21 bibliographic format developed by the Library of Congress and the National Library of Canada after merging their respective national formats. Because of the that MARC 21 has become the format favoured by other national libraries and by online bibliographic utilities and their end-users; is the format supported by the majority of library systems; and offers participation in an international bibliographic community following common standards, and the advantage of copy cataloguing at much reduced cost and with no need to maintain conversion programs. The British Library undertook to support libraries and other organisations, including overseas users of UKMARC, in preparing their implementation of MARC 21.

There are five types of MARC 21 formats. These are:

Format for Bibliographic Data : It provides details for creation and design of bibliographic records. It describes the records and provides access to them. It defines the different data elements in terms of codes, tags and indicators for processing the records.

Format for Authority Data : Authority format provides codes and code values for accepted and used headings for authority control. It includes personal and corporate names as well as subjects and their subdivisions.

Format for Holdings Data : The format includes codes, tags and indicators that enable to provide holdings information of single part and multi part items as well as serial items. It enables to interface the automated system with the union catalogue or inter library loan system.

Format for Classification Data : It includes codes and indicators for classification data a in LC scheme and DDC. It provides details of classification schedules helping in assigning and validating class numbers.

Format for Community Information Data : It is suitable for coding community non-bibliographic information. One may be interested in creating database for experts in an area for which there are details in this format. Similarly for organisations, events, programmes or services the format provides details.

The first there formats and particularly are most commonly used in cataloguing in libraries.

UKMARC VS USMARC

There is some difference between USMARC and UKMARC formats in respect of terminology and other aspects. Adoption of different texts of AACR caused some differences between the two formats. The second edition of AACR had its impact on the second edition of UKMARC manual which appeared in 1980 primarily in terms of coordinated treatment for materials of all kinds. In contrast USMARC format never confined to AACR only but reflected various cataloguing codes applied in American cataloguing practices. Differences of terminology between two formats such as "Leader" in USMARC versus "Record label" in UK MARC format of Bibliographic level" in USMARC versus "Class of record" in British format. Besides, there were also differences between these formats in respect of calculation of fixed field position, fixed length data elements and control numbers used. For example, the record control number used in USMARC is the LC card number whereas in UK MARC it is ISBN.

How does MARC 21 differ from UKMARC

The main differences are that of scope and development. MARC 21 is a set of five communication formats for representing and exchanging data in machine-readable form, of which the bibliographic format is the most widely used. UKMARC is a national format, developed solely for bibliographic purpose. MARC 21 bibliographic format has been developed as a standard format for conveying information in a systematic way about print and a wide variety of other types of materials. In this respect, it can be said to be better suited to the diverse nature of library collections today.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

3) Differentiate between USMARC and UKMARC directory.

....

....

....

....

7.5.4 UNIMARC

Introduction

UNIMARC was the brain-child of IFLA. It was conceived of as a tool for an international network. Although the record structure which became ISO-2709 was accepted early on; during the very first cooperative project between L.C. and B.N.B.

Development of UNIMARC

Despite cooperation there emerged several versions, e.g. UKMARC, INTERMARC and USMARC, whose paths diverged owing to different national cataloguing practices and requirements. Since the early 1970s an extended family of more than 20 MARC formats has grown up. Differences in data content means that editing is required before records can be exchanged.

One solution to the problem of incompatibility was to create an international MARC format (UNIMARC) which would accept records created in any MARC format. So records in one MARC format could be converted into UNIMARC and then be converted into another MARC format. The intention was that each national agency would need to write only two programs – one to convert into UNIMARC and one to convert from UNIMARC - instead of one program for each other MARC format, e.g. INTERMARC to UKMARC, USMARC to UKMARC, etc.

So in 1977 the International Federation of Library Associations and Institution (IFLA) published *UNIMARC: Universal MARC format*, stating that "The primary purpose of UNIMARC is to facilitate the international exchange of data in machine-readable form between national bibliographic agencies". This was followed by a second edition in 1980 and a *UNIMARC Handbook* in 1983. All focussed primarily on the cataloguing of monographs and serials and took advantage of international progress towards the standardisation of bibliographic information reflected in the International Standard Bibliographic Descriptions (ISBDs).

In the mid-1980s it was seen necessary to expand UNIMARC to cover documents other than monographs and serials. So a new description of the format – the *UNIMARC Manual* – was produced in 1987. By this time UNIMARC had been adopted by several bibliographic agencies as their in-house format. So the statement of purpose was amended to include "UNIMARC may also be used as a model for the development of new machine-readable bibliographic formats".

UNIMARC Manual was issued in 1987. This edition of the format removed the provisional status form and revised field that were specific to sound recording, visual projections, video recording, motion pictures, graphics, printed music, and microforms, in addition to the remaining provisional cartographic fields. Fields for electronic resources were added as provisional.

The 1994 edition supersedes the previous ones and contains additions and changes agreed upon by the permanent UNIMARC Committee from June 1990. The changes are described in Appendix O. The new loose-leaf formats was intended to be updated with change pages as needed.

UNIMARC is maintained by IFLA and UBCIM office.

Purpose and Scope of UNIMARC

i) The primary purpose of UNIMARC is to facilitate the international exchange of bibliographic data in machine-readable form between national bibliographic agencies.

ii) UNIMARC may also be used as a model for the development of new machine-readable bibliographic formats.

iii) UNIMARC is intended to be a carrier format for exchange purpose. It does not stipulate the form, content, or record structure of the data within individual systems.

iv) UNIMARC does provide recommendations on the form and content of data when it is to be exchanged. Records are usually structured in exchange tape format as the last stage in any conversion process, after form, content, and content designators have been converted to the UNIMARC standard.

v) Those organisations intending to use UNIMARC for data interchange will find it useful to coordinate their internal format content designators and field and subfield definitions with those in UNIMARC to reduce the complexity of data conversion when the records are converted into the UNIMARC exchange tape structure.

vi) It facilitates the description, retrieval and control of bibliographic items. This is achieved by providing a structure for recording bibliographic information which is input by reference to international standards.

Scope

The scope of UNIMARC is to specify the content designators (tags, indicators and subfield codes) to be assigned to bibliographic records in machine-readable form and to specify the logical and physical format of the records. It covers monographs, serials, cartographic materials, music, sound recording, graphics projected and video materials, rare books and electronic resources.

Features Relating To UNIMARCas an Exchange Format

i) An interesting feature of the format is the inclusion of fields in blocks defined by type of data element. Up to the development of UNIMARC, the major national MARC formats had ordered the different fields in a way that reflected the order of the field on a traditional catalogue card.

ii) UNIMARC avoided this towards one particular end product of a machine-readable bibliographic record and put all name access points in one block instead of supplying different fields for author as main entry from author as added entry.

iii) The most novel feature of UNIMARC is its treatment or links between one bibliographic item and another.

iv) Bibliographic items have relationships with each other. They may have previous editions, they may, as in the case of serials, have related, earlier titles. Moreover, they may be in the same journal or series as each other. In special cases, some bibliographic items are translations of others.

v) Another kind of relation is the sharing of common subject or authorship and it has a number of different ways of showing these linking relationships.

The UNIMARC format

The UNIMARC format, like any other version of MARC, involves three elements of the bibliographic record:

- Record structure
- Content designation
- Data content

Record structure

The record structure is designed to control the representation of data by storing it in the form of strings of characters known as *fields*.

All data in the record must be stored using one or more character sets. Since computers can store and manipulate only numbers, each symbol, alphabetical character etc. is assigned a number following the rules of a particular character set. For example, one character set assigns the number '75' to 'K'. UNIMARC allows the use of certain character sets, approved by the International Organisation for Standardisation (ISO).

The record structure established by UNIMARC is an implementation of the relevant standard: *Format for Bibliographic Information Interchange on Magnetic Tapes (ISO 2709-1981)*. This structure utilises record labels and directories. As few users need concern themselves with such items, the description below covers the way a cataloguer sees the record.

Content designation

Certain conventions are followed in order to identify the data elements within records. Such elements which include author, title and subject access are further characterised where necessary. This supports the manipulation of the data for a variety of purposes:

- To provide multiple access points for searching,
- To allow the typography and layout to be varied,
- To permit certain elements of the record to be omitted where this is required.

In addition, UNIMARC records may be formatted for visual display on a VDU, for output on CD-ROM or fiche and for printing out as hard copy.

In general, UNIMARC provides content designation only for data which is applicable to all copies of a work. However, information which applies only to some copies (or even a single copy) of a work may be of interest beyond the holding institution. In such cases UNIMARC assigns specific fields for such details. These fields are also available for cases where the information is for in-house purposes only.

Data content

The content is the data which is stored in the fields within the record. Data can be coded data or bibliographic data.

- Coded data is used to represent such items as control numbers, publication type, and main language of text. There is also provision for the characteristics of special types of items such as printed music.
- Bibliographic data is defined by reference to the International Standard Bibliographic Description for that type of material. In addition, each record can carry a class number and subject heading.

The fields, which are identified by three-character numeric tags, are arranged in functional blocks. These blocks organise the data according to its function in a traditional catalogue record. In the table below, fields 0 to 1 hold the coded data while fields 2 to 8 contain the bibliographic data:

Block	**Example**
0— Identification block	010 International Standard Book Number
1— Coded information block	101 Language of the work
2— Descriptive information block	205 Edition statement
3— Notes block	336 Type of computer file note
4— Linking entry block	452 Edition in a different medium
5— Related title block	516 Spine title
6— Subject analysis block	676 Dewey Decimal Classification
7— Intellectual responsibility block	700 Personal name - primary intellectual responsibility
8— International use block	801 Originating source
9— Reserved for local use	

In addition to the 9 –block any other tag containing a 9 is available for local implementation.

The fields defined by UNIMARC provide for different kinds and levels of information. This can be shown by looking at a typical record in the UNIMARC format.

Putting UNIMARC to work

Bibliographic records in the UNIMARC format are designed for use in automated library systems. Depending on the versatility of the system a range of related functions can be supported by manipulating the data. Two such functions are information retrieval and displaying citations.

Information Retrieval

In the UNIMARC format each data element is identified for the purpose of information retrieval. Using computer software, it is possible to search on most of the MARC fields and subfields in the record. For example:

- Keywords (i.e. significant words)
- Subject headings
- Author
- Name, topical name, geographical name as subject
- Title and series title
- Standard numbers (ISBN, ISSN etc.) and numbers assigned by agencies (a national bibliographic agency, a government printing office etc.)
- Classification numbers
- Publisher
- Publication date and type
- Acronyms formed from name and title words
- Coded items. For example, FICTION would select the above record because field 105 character position 11 codes it as fiction.

While each record in the UNIMARC format is a discrete entity, a catalogue consisting of many such records becomes a database enhanced with the capacity to respond to highly specific or comprehensive search strategies. The range of search options will, of course, depend on the kind of software employed.

Displaying citations

UNIMARC offers a choice of formats for displaying records. Naturally, readers will not want to consult the full MARC record simply because the format is intended not for human perusal but for processing by computer.

7.5.5 Common Communication Format (CCF)

Historical Backdrop

The rapid growth of international bibliographic exchange formats including ISO's well known format for bibliographic information interchange on magnetic tape (ISO- 2709) and lack of compatibility amongst them poses problems both to libraries and other information services to exchange records with one another. Hence, necessity arose for each of these organisations to agree upon a common standard format for easy exchange purposes. As such to study the desirability and feasibility of establishing optimum compatibility between existing bibliographic exchange formats led to the convening of the International Symposium on Bibliographic Exchange Formats with the initiation by the Unesco General Information Programme in April 1978 in Taormina, Sicily. The Symposium was organised by UNISIST International Centre for Bibliographic Description in collaboration with International Council of Scientific Unions Abstracting Board, IFLA and ISO.

Genesis of CCF

As a direct result of the Symposium, an ad hoc Group was constituted for developing the Common Communication Format (CCF). After prolonged deliberations by the experts, the Group decided :

1) That the structure of the new format would be in accordance with the ISO – 2709.

2) That the core record would consist of a small number of mandatory data elements essential to bibliographic description, identified in a standard manner.

3) That the core record would be larger in number by adding optional data elements, identified in a standard manner.

4) That a standard technique would be developed for accommodating levels, relationships, and links between bibliographic entities.

Besides, it was also resolved that CCF should provide a meaningful link between the major exchange formats basing on ISBD.In the course of deliberations, the Group duly considered and compared all the data elements in the Reference Manual, UNIMARC, Guidelines for ISDS, MEKOF-2 ASIDIC/EUSIDIC/ICSU – AB/NFAIS Interchange specifications and the USSR- US Common Communication format. Commonly used data elements in those six formats were identified and formed the basic guide and core of CCF on the above six standard formats. The Group identified a small number of data elements which were used virtually by all information-handling communities, including both the libraries and the abstracting and indexing organisations. These commonly used data elements form the core of the CCF. Attempt was also made to device a technique to establish relationship between bibliographic records and between elements within bibliographic records. The concept of the record segment was developed and later refined and finally a method for designating relationship between records, segments and fields was duly approved by the Group. Consequently the first edition of CCF was published in 1984 and subsequently the second edition was brought out in 1988, The publication entitled," Implementation notes for users of CCF was brought as the guide for the benefit of CCF users. Subesequently UNESCO/ PGI brought 3rd edition of CCF in 2 volumes. There volume one CCF (B): Bibliographic Information and volume 2 CCF (F): Factual. Various bibliographic agencies in Asia, North and South America and Europe recommended changes which were incorporated into this new edition.

Rationale

Within an information system, the records which form the database will usually exist in a number of separate but highly compatible formats. At least there will be:

- A format in which the records will be input to the system,
- A format best suited to long – term storage,
- A format to facilitate retrieval, and

- A format in which records will be displayed.

Need of CCF

The need of CCF is of paramount importance if two or more organisations wish to exchange records with one another. It cannot be achieved unless each of these organisations agree upon a common standard format for exchange purposes.

If there is a single national standard exchange format, information interchange with that country is possible so also will be greatly facilitated both technically and economically. But on the other hand, if each nation's standard format is different from all others, then it will be more problematic and complex to have international information interchange among national bibliographic agencies because of the number of computer programs that must be written to accommodate the translation of records form one format to another.

At present many national standard exchange formats exist. However, a number of these formats are very similar to one another, others differ significantly. Very rarely if two national formats are likely to be identical then their records can be handled by the same computer programs.

Scope and Uses of CCF

The CCF is designed to provide a standard format for three major purposes :

1) To permit the exchange of bibliographic records between groups of libraries and the abstracting and indexing services.

2) To permit a bibliographic agency to use a single set of computer programs to manipulate bibliographic records received from both the libraries and the abstracting and indexing services.

3) To serve as the basis of a format for an agency's own bibliographic database, by providing a list of useful data elements.

Uses

These uses have been accommodated in the following ways:

1) By specifying a small number of mandatory data elements which are recognized by all sectors of the information community as essential in order to identifying an item.

2) By providing mandatory data elements that are sufficiently flexible to accommodate varying descriptive practices. A section entitled "USE" for each field and subfield indicates whether-the use of that data element is mandatory or optional.

3) By providing a number of optional elements which may be useful to describe an item according to the practices of the agency which creates the record.

4) By providing a mechanism for linking records and segments of records without imposing on the originating agency any uniform practice regarding the treatment of related groups of records or data elements.

Structure

The record structure of the Common Communication Format comprises a specific implementation of the international standard ISO- 2709, Each CCF record consists of four major parts:

a) Record Label

b) Directory

c) Datafields

d) Record Separator

Record Label

Each CCF record begins with a fixed-length label of 24 characters to provide parameters to process record.

Directory

The directory is a table containing a variable number of fourteen-character entries ; the table is terminated by a field separator character. Each directory entry corresponds to an occurrence of a datafield in the record, and is divided into five parts :

1) Tag
2) Length of datafield
3) Starting character position
4) Segment identifier
5) Occurrence identifier

1) **Tag :** A three character code identifying the datafield which corresponds to the directory entry.

2) **Length of Datafield** : A four digit number showing how many characters are occupied by the datafield, including indicators and datafield separators but excluding the record separator code if the datafield is the last field in the record.

3) **Starting character position :** A five digit number giving the position of the first character of the datafield relative to the base address of data, i.e. the first character of the first of the datafields.

4) **Segment identifier :** A single character which designates the datafield as being a member of particular segment.

5) **Occurrence identifier:** A single character (chosen from 0-9 and A-Z) which differentiates multiple occurrences of datafields that carry the same tag within the same record segment.

A single directory entry is organised as follows:

TAG 3characters	LENGTH OF DATAFIELD POSITION 5 characters	STARTING CHARACTER POSITION 5 characters	SEGMENT IDENTIFIER 1 character	OCCURRENCE IDENTIFIER 1 character

Datafields

A datafield consists of:

1) Indicators,

2) One or more subfield each of which is preceded by a subfield identifier, and

3) A data field separator

(1) ***Indicators :*** Two bytes reserved for use as defined for each data filed. These may supply further information about the contents of the data field or about the action required in certain data manipulation processes.

(2) ***Sub fields* :** A subfield consists of sub field identifier followed by a data string which is terminated by either another sub field identifier or a field separator.

(3) ***Data field separator* :**The data field separator (Character 1/14 of ISO 646) constitutes the final character of every data field.

A data field which has a single subfield will be organised as follows:

INDICATORS 2 characters	SUBFIELD IDENTIFIER 2 characters	SUBFIELD Variable	FIELD SEPARATOR 1 character

A data field which has two subfields will be organised as follows:

INDICATORS 2 characters	FIRST SUBFIELD IDENTIFIER 2 characters	FIRST SUBFIELD VARIABLE	SECOND SUBFIELD IDENTIFIER 2characters	FIELD SEPARATOR 1 character

Record

The record separator (Character 1/13 of ISO 646) is the final character of the record. It follows the field separator of the field datafield of the record.

Features of CCF

The unique features of CCF attract many institutions, information centers, and other organisations of various countries are mentioned as under:

1) It can be used to produce catalogue cards as all the necessary data elements are incorporated.

2) It is used friendly and rather catalogue friendly for no cataloguing rules are imposed.

3) It is a flexible and very popular format among UN organisations and international bodies. Many developing countries are adopting it for the creation of bibliographic records in machine readable form. Bibliographic agencies following AACR2 can be conveniently converted to CCF format.

4) It provides basic data elements and has facility for optional elements and private fields thus enabling an agency to incorporate new standard elements considered important.

5) It facilitates a library and bibliographic agency to use a single use of computer program for the exchange of data.

6) Some of the mandatory data elements are flexible and can accomodate varying descriptive practices.

7) Its simple set of data elements that can be used at any bibliographic level and are dissociated from cataloguing codes.

8) It is the logically defined record structure which uses the fourth element of the ISO 2709 directory to denote the bibliographic level and filed occurrence.

9) CCF is specifically designed for retrieval and output within an institution. It neither includes its own cataloguing rules nor recommends any particular cataloguing code oriented towards any specific type of output format.

10) It identifies and defines relationship between the data elements and their respective content designators independent of category of items or different types of materials. It outlines a specific system of content designators for bibliographic records pertaining to all forms of documents.

7.6 INDIAN STANDARDS

Standardisation of record format has not received due attention in Indian libraries. At national level, Indian standards institution now renamed as Bureau of Indian Standards (BIS) had evolved a standard for bibliographical references in 1963 for use in non-computerised systems. However it could not keep track of the development in the media and forms of documents in the following years. A revision of the standard appeared in 1979 wherein ISBD was suggested as a substitute format to be adopted by agencies willing to do so. It was a format incomplete and inadequate in many respects. The sixteenth Indian standards convention of ISI held in Bhopal in October 1975 discussed the issue of standardisation of bibliographic information in the contest of machine-readable records. Continued efforts taken by ISI in this direction resulted in design and publication of a standard IS: 11370-1985 titled 'Guide for data element and record format for computer Based bibliographic Description of different kinds of Documents' in July 1986. Structure of this format confirms to ISO 2709-1981.

Structure of IS: 11370

The general structure of the bibliographic record format is given below:

Leader

Directory

Data fields

Records separator

Among the data elements, descriptive block is based on ISBD. Subject in analysis block includes POPSI, PRECIS, keywords and synopsis or abstract. The Bureau of Indian Standards is yet to take steps to revise it. During the late 1980's NISSAT organised a tripartite meeting (CALIBNET, DELNET, INFLIBNET) to sort out the difference in choice of format – Common Communication Format (CCF) vis-à-vis UNIMARC. Finally it was decided that database producer can use either of the formats. NISSAT also constituted a group to draft the INDIMARC guidelines based on the framework prescribed by the CCF.

7.7 SUMMARY

Current exchange formats may be affected by rapidly changing technology for data storage, transfer, retrieval, etc. For example the development of Open System

Interconnection procedures for data transfer, growth of CD-ROM, Multimedia, Hypertext CD-ROMs, DVD technology, web server based databases, etc., as medium for data storage and distribution. But this is the right time for us to start thinking and working in this direction to solve the problems, which we are going to face in near future.

7.8 ANSWERS TO SELF CHECK EXERCISES

1) Format conveys the notion of a formalised framework which will hold records of varying content according to certain set of rules of conventions controlling the representation of the data. These rules may be unique to a system, or shared with other system., Formats can be of two types:

   - Internal/Local format
   - Exchange/Communication/Interchange format

2) There are three USMARC communication format:

   - USMARC format for bibliographic data (UFBD)
   - USMARCformat for authority data (UFAD)
   - USMARC format for holdings and location (UFHL)

All the three USMARC formats are implementations of ANSI Z39.2, American National Standard for bibliographic information interchange on magnetic tape which conforms with the ISO 2709.

The physical structure of all USMARC formats are similar to the structure of the UK MARC record, although there are some differences in terminology and definition of fields:

Leader	Directory	Variable Control fields	Variable Data fields

1) Nature and function of directory are identical to UK MARC with the following exceptions:

   - USMARC records have less restriction on length of fields, with maximum 9999 characters:
   - USMARC contains no subrecord directory entries as analytical records are not handled via the directory:
   - Character positions are counted from zero.

7.9 KEY WORDS

Data : Each record contains data which refer to s a separate entity, object of unit recognised by the system, e.g., books, journals, employee, customer, etc.

Field : Field is a part of record which contains data referring to one characteristics of the entity or unit represented by the record. Fields may be fixed in length or variable in length. The number and nature of field is generally determined at the time record structure is designed and also taking care of applications.

Logical Record : A record for a bibliographic or other item presented in machine readable form may be described as a logical record. It is not necessarily stored in the computer system in the same form, it may be broken up and the parts stored in the different places on a computer disk or it may be grouped with other records in one large physical record.

Physical Record : Physical record refers to the physical arrangement on the storage medium.

Record : A record is a group of related data elements treated as a unit. A record may be divided into identifiable fields and subfields which speeds up the process of search and retrieval by computer.

Subfield : Subfields are part of a field which cannot stand completely in isolation from the data in the field, but require individual treatment.

Tag : A tag consists of one or more characters or digits which uniquely denotes a data element or a whole field in a record.

Variable Field : A field which can be extended to any length as per requirement.

7.10 REFERENCES AND FURTHER READING

Byrne, Deborah. *MARC Manual: Understanding and Using MARC Records.* Englewood: Libraries Unlimited, 1991. Print.

Gredley, Ellen and Hopekinson, Allan. *Exchange Bibliographic Data: MARC and Other Intermediary Format.* London : Library Association, 1990. Print.

Indian Standards Institution. *Guide of Data Elements and Record Format for Computer-based Bibliographical Databases for Bibliographic Description of Different Kinds of Documents (IS: 11370-1985).* New Delhi: Indian Standard Institution, 1986. Print.

Prasad, A.R.D. “Brief Introduction to Z39.50 Protocol”. *National Convention on Library and Information Networking*, 6-9 Nov. 2001. Print.

CCF/B: The Common Communication Format for Bibliographic Information. Ed Peter Simmons and Alan Hopkinson. Paris:Unesco, 1992. Print.

UNIT 8 CATALOGUING OF NON-BOOK MATERIAL

Structure

8.0 OBJECTIVES

This Unit aims to give an overview of the bibliographic description of the various types of non-book material and the rules followed for their cataloguing according to AACR-2R.

After reading this unit, you will be able to:

- identify and describe the bibliographic information about each item of non-book material as prescribed in AACR-2R;

- discuss the relevant rules for effective cataloguing of non-book material;
- identify and mention access points for different items to be catalogued; and
- catalogue electronic resources according AACR 2002, RDA and FRBR.

8.1 INTRODUCTION

Library is a centre for the communication of ideas and knowledge. It is also a repository of records of human civilisation serving as an information clearing house to the community. The librarian must organise and order the flow of information so as to maximise the use of the resources of a library. In the present era of scientific and technological advancement, the sources of knowledge are not confined to conventional print media only. The impact of ICT has directly caused the spread of non-print materials. These documents are responsible for the communication of recent information and new knowledge. As a result, the libraries world over have acquired large collection of such materials. These materials need to be collected, organised and disseminated at the right time. The information looses its dynamism if not communicated timely. The nature of non-print material demands a separate treatment for organisation that forms the subject of the present Unit.

Evolution of Non-Book Materials (NBM)

The most primitive media for recording knowledge were stones, rocks, clay tablets, parchment, vellum, papyrus and palm leaf. The book in paper medium is the outcome of the printing press of John Gutenberg. The advent of microforms or non-book material goes back to 1953 with the invention of microphotography. Although these have been available for a considerable time, it is only in the past six decades that we have seen their active use and availability in the library market.

The educational use of audio-visual materials began during the 1950s and libraries started acquiring the materials in the late 1960s. During this time the governments and national professional organisations became interested in the educational possibilities of the materials as opposed to their entertainment value they had been labelled with previously. Thus, the history of the non-book materials has passed from the stone age to the electronic age. Libraries have been providing information in many different media. They circulate books, microforms, art prints, periodicals, disc records, audio and video tapes, games and simulations, motion pictures, films, slides, film strips, models, realia and so on.

The world is now witnessing an exponential growth of information and there is a need to provide quick access to information. The exorbitant cost of printing and publishing and the need to reduce time lag in conventional printing necessitated the shift in focus from traditional print media to non-print media. Over the last two decades, there has been tremendous growth in the development of new technologies in photographic, micrographic, computers, fiber-optics and telecommunications technologies that affect the preparation, organisation, storage and retrieval of information.

8.2 NON-BOOK MATERIAL

NBMs are those materials which do not come within the definition of a book, periodical or pamphlet and which require special holding e.g. audio-visual materials, microforms or computer files, electronic resources. It is generally understood to be any resource material, which is not a printed book but contributes to the learning process. The NBM require special treatment in terms of their bibliographic description in order to exploit information from those formats. Webster's Third New International Dictionary defines "Non-book being something other than a book; being a manuscript, microfilm, map on other library holding that is not a book." Harrold's Librarian's Glossary describes NBM as "those library materials which come with in the definition of special holding, i.e., audio-visual materials, vertical file materials, microforms or computer software."

8.2.1 Types of Non-Book Material

There are various types of NBM which are mentioned as under:

Cartographic Materials

- Ariel Chart
- Leaf
- Ariel Remote Sensing image
- Atlas
- Celestial Globe
- Chart
- Globes
- Map
- Plan
- Relief Model
- Remote Sensing Image
- Space Sensing Image
- Topographic Drawings

Motion Pictures and Video Recordings

- Film Catridge
- Cassette
- Video Cassette
- Video Disc
- Video Reel

Graphic Materials

- Art Original
- Art Print
- Art Reproduction
- Chart
- Film Strip
- Kit
- Photograph
- Picture
- Post Card
- Slide.
- Technical Drawing
- Transparency

Manuscripts

- Item (for collection of manscripts.)
- Box

Music

- Score
- Condensed
  - Minature Score
- Chorus Score

Sound Recordings

- Sound Cartridge
- Sound Cassette
- Sound Disc
- Sound Track Film Reel (Cassette)

Computer Files

- Data Files
- Program File
- Object Program

Three-Dimensional Artefacts and Realia:

- Art Original
- Realia
- Game
- Diorama
- Model

Microform

- Aperture Card
- Microfiche
- Micro Film Cartridge Cassette Reel

8.2.2 Utilities of Non-Book Material

Following are some of the utilities of NBM:

- NBM as storage media provide potential alternative access to information, which has enormous information storage capacity with low cost.
- Data damage is drastically reduced with NBM and information can be retrieved and transferred speedily and accurately.
- Their durability quality provides the benefit of repetitive use of information without deterioration or loss.
- Presently there is a trend towards compressing information carriers so that they occupy less space and make storage and distribution easy. Books are "space eaters" where as NBM are "space savers".
- They have the quality of security, accessibility, portability, reliability, economy, easy retrieval and easy updating.
- Rare books can be stored in microform. The whole set of Encyclopedia Britannica; Chemical Abstracts and Biological Abstracts can be stored in a CD-ROM.
- Monotonous topics in history produced in coloured picture with enchanting voice as videotapes helps to remember and grasp the topic easily as media resources are powerful forms of communication. It has been ascertained that a person can remember 10% of information on reading, 20% of what is heard, 30% of what is viewed, but 50% of information are remembered if it is heard, seen and discussed which can only be possible with the audio-visual materials.

8.2.3 Constraints of Using Non-Book Material

Following are some constraints in using NBM:

- In a paper oriented society majority of the people still are not acquainted with microform or the screen display of information and this, they are not quite reconciled to this change.
- Special methods of handling materials are troublesome to operate. Hence, special awareness program to users are very much essential.
- Reading NBM sometimes gives strain to the eyes.
- Acquiring of equipments for operation of NBM, its maintenance and storage facilities amounts to high cost.
- Like books, NBM are not issued for use at home.
- The cost of microfilming may not be justified if the record has short retention schedule with little or no reference. It requires special storage devices for the NBM.
- Special types of equipment are required to retrieve the information from different types of materials, e.g.., microfilm reader or reader/printer is required to retrieve information from microform.
- As and when necessary, information on microform can't be manually changed or updated as on paper.
- High obsolescence of equipment due to technology adds on to cost.

Self Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

1) State the important reasons for advent of non-book materials.

....

....

....

....

2) State the different types of NBM.

....

....

....

....

8.3 PROBLEMS OF CATALOGUING NON-BOOK MATERIAL

New media materials due to their inherent peculiar format, very frequently pose serious problems in their arrangement, cataloguing and retrieval of information contained in them. Hence, the cataloguer may encounter the following possible problems.

- Information may be difficult to obtain from the documents to be catalogued than with conventional documents having title page.
- Information collected from one source in the NBM may differ with that obtained from another source of documents.
- It may be harder to reach the cataloguing decisions with reference to choice of access point i.e. determination of heading or in other words to decide the person who is chiefly responsible for the intellectual content of the document, which is less experienced in case of books and serials.
- Information about physical description of different types of NBM definitely creates problems for cataloguer rather than that for conventional documents.
- Not possible to obtain information through the naked eyes as it requires special equipments.

Cataloguing of NBM in comparison to books has a number of special problems. Eric J. Hunter has pointed out some specific problems regarding frequent variation in its physical forms and difficulties in getting the author equivalent. John Horner at the same time in his book **Special Cataloguing** has discussed a number of possible problems along with the two above problems as stated by Hunter. The problems stated by Horner are:

a) Machinery may be needed to use the documents.

b) The materials may be more fragile, rare and expensive than normal book-form materials.

c) Special subject knowledge and that of the relevant rules in catalogue code may be needed to catalogue the materials thoroughly and quickly.

d) Special knowledge and experience of the physical form may be needed.

e) Special cataloguing tools that is, codes and thesauri may be needed.

f) Hence, it may be necessary to compile one's own aids because of the variety of the type of material.

Despite these bottlenecks, the entire spectrum of the new media cannot be kept out of the libraries of today and tomorrow. With the publication of AACR-2, Amendments and AACR-2002, the cataloguing of NBM has become easy, clear and standardised. Hence, through proper cataloguing their use in the libraries gets facilitated.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

3) Mention the major problems of cataloguing of NBM.

....

....

....

....

8.4 CATALOGUING OF NON-BOOK MATERIAL

AACR-2 specifies sources of information to be used in describing a publication; in the case of printed monograph, for example, such sources include the title page, the verso of the title page, etc. Of these, the source of bibliographic data to be given first preferences as the source from which a bibliographic description is prepared is called the chief source of information. The rules identify a chief source of information for each type of material.

Chief Sources of Information

Type of Materials	**Sources**
Cartographic materials	Cartographic item itself Container or case, the cradle and stand of globe, etc.
Manuscripts	Title page and Colophon
Music	Title Page
Sound recordings	
Disc	Label
Tape (open reel-to-reel)	Reel and label
Tape Cassette	Cassette label
Tape Cartridge	Catridge and label
Sound recording on film	Container and label
Motion picture and video	Film itself and its container

Recordings	(if integral part of item)
Graphic Materials	Item itself including any labels and the container
Computer files	Internal user label Information issued by publisher, creator, etc.
Three – dimensional artefacts and Realia	Object itself with any accompanying textual material and container
Microforms	Title frame

8.4.1 Punctuation Marks

One of the significant features of the ISBD is a set of prescribed punctuation. The prescribed punctuation mark precedes each element in the description and signifies the nature of that element. The prescribed punctuation marks are used as a device of recognition for both machine and human manipulation of bibliographic records.

Specific and detailed rules with regard to prescribed punctuation are given in each chapter in AACR-2. The details of punctuation marks of ISBD (G) has been provided in Unit 7. Besides some punctuation marks necessary for NBM have been given below:

Parentheses ( )

Parentheses are used to:

1) enclose physical details of accompanying material.

2) enclose the number of logical records after the designation for a data file; the number of statements and the name of programming language the designation for a program file; the number of logical records, or statements in each file after the designation for a multipart file; or the name, number, etc. mentioned after the designation for an object program.

3) enclose the number of frames of microforms a filmstrip and the speed of a film or recording.

Plus Sign +

1) precedes a statement of accompanying material.

2) is used to indicate the Northern Hemisphere when giving the declination of the center of a celestial chart.

Square Brackets [ ]

(1) enclose information taken from outside the prescribed source or sources.

(2) enclose the general material designation.

8.4.2 Areas of Description

AACR-2 prescribes detailed rules for each area of description. The general rules are presented in AACR2R Chapter 1 and rules for specific types of materials are given in chapters 2 to 12. Adequate examples are included to illustrate the rules. The major elements in bibliographic description are discussed below.

In presenting data in the bibliographic description, information taken from the chief source of information is preferred. If the information required is not available or is insufficient from the chief source, other sources are used. Rules for each area are enumerated in Chapters 2 to 12. Information taken from outside the prescribed sources is enclosed in brackets.

8.4.3 Structure of Description

The bibliographic description of NBM follows the same norms as for books and other materials. The main structure of the bibliographic entry comprises the heading, the description and the subject description. The structure of the bibliographic description according to AACR 2R is given below.

First Level (1.OD1)

The level was designed for minor-item and for entries in catalogues with a policy of minimum description. The bibliographic elements to be included are set forth in the following schematic illustration.

Title proper/First statement of responsibility, if different from main entry heading in form of number or if there is no main entry heading. Edition statement. Material (or type of publication) specific details first publisher, etc., date of publication, etc. Extent of item – Note(s) –Standard number.

It would primarily be sufficient to identify item in a small library collection

Shastri Ravindra	
1946	Hamare Deshbashiyon[manuscript].-**[s.l:s.n]**, 10 leaves; 24cm. Holograph, signed poem in Hindi. Two leaves are stained by water. 1. Hindi Poetry. I. Title.

Second Level (I.OD2)

This level was designed for the standard range of item found in the library and for entries in catalogues with a policy of standard description. The following elements are included:

Title proper [General material designation] = Parallel title: other title information/First statement of responsibility, Each subsequent statement of responsibility. –Edition statement/First statement of responsibility relating to the edition,- material (or type of publication) specific details,-first place of publication, etc.: First publisher, etc., Date of publication, etc.. – extent of item: other physical details; Dimensions. –(Title proper of series/Statement of responsibility relating to series, ISSN of series; Numbering within the series. Title of sub-series, ISSN of sub-series; Numbering within sub-series) – Note(s). –Standard number. This level might appropriately be used in medium sized library.

Shastri Ravindra	
1946	Hamare Deshbashiyon[manuscript].-**[s.l:s.n]**, 10 leaves; 24cm. Holograph, signed poem in Hindi. Two leaves are stained by water. 1. Hindi Poetry. I. Title. ○

Third Level (1.0D3)

It includes all the rules applicable to the item being catalogued. This level represents full description and is recommended for items which, in the context of the catalogue are considered to be important and rare. All elements set forth in the rules which are applicable to the item being described are included, it is appropriate to large libraries and research collections.

8.4.4 GMD (General Material Designation)

GMD prescribes the broad class of publication to which the item belongs, for example, sound recording, music, map, film- scripts, microform, motion picture, machine-readable data file will be replaced by computer file etc.

This is an optional addition in AACR-2. The use of GMD is very much essential in case of non-book materials and it may not be preferred by the libraries for book form materials. For the purpose of using GMD the cataloguer may use the following terms immediately following the title proper enclosed in square brackets:

Cartographic Materials	Map
	Globe
Computer File	Computer File
Graphic Materials	Art Original
	Art Reproduction
	Chart
	Filmstrip
	Flash Card
	Kit
	Picture
	Slide
	Technical Drawing
	Transparency

Manuscript————————————	Manuscript
Microform————————————	Microform
Motion Picture————————————	Motion Picture
Music————————————	Music
Sound Recording————————————	Sound Recording
Art original, Diorama, Game Model, Realia	Realia
Video Recording————————————	Video Recording

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

4) Mention the chief source of information for motion picture and computer file of cataloguing of NBM.

....

....

....

....

8.5 BIBILOGRAPHIC DESCRIPTION OF NON-BOOK MATERIALS (AACR-2 REV.ED.)

Various types of NBMs along with relevant rules of AACR 2, 1988 Rev have been discussed with examples for your better understanding about their bibliographic description.

8.5.1 Cartographic Materials

Rule 3.01A Scope

The rules in this chapter cover the description of cartographic materials of all kinds. Cartographic materials include all materials that represent the whole or part of the earth or any celestial body. These include two and three – dimensional maps and plans (including maps of imaginary places); aeronautical, navigational and celestial charts ; globes ; block diagrams ; sections ; aerial photographs with a cartographic purpose; birds eye-views (map views), etc.

Rule 3.0 B2 Chief Source of Information

The chief source of information (in order of preference) is:

a) The cartographic item itself; when an item is in a number of physical parts, treat all the parts (including a title sheet) as the cartographic item itself.

b) Container (portfolio, cover, envelope, etc.) or case, the cradle and stand of a globe, etc.

Rule 3.1B: Title proper, 3.1C: GMD, 3.1D: Parallel title and 3.1E including 3.4 follows the rules given in rule 1 and applicable to cartographic materials.

Rule 3.3B1 Give the scale of a cartographic item (except as noted below) as a representative fraction expressed as a ratio (1). Precede the ratio by scale. Give the scale even if it is already recorded as part of the title proper or other title information.

Scale ca 1: 45, 00000

If a scale statement found in the chief source of information or accompanying material is not expressed as a representative fraction, give it as a representative fraction in square bracket.

Scale [1:221,440]

Rule 3.3B3 If the scale within an item varies and the outside values are known give scales connected by a hyphen.

Scale 1: 11,000-1:20,000

If the values are not known, mention that scale varies

Rule 3.3C *Statement of Projection*

Give the statement of projection if found in the item, e.g. conic equidistant projection

Rule 3.5B2 If there is more than one map, etc. or one or more sheets, specify the number of maps and number of sheets. e.g. 6 maps on 1 sheet.

Rule 3.5D5 Give the physical description of maps and globes as follows:

1 globe: col., plastic, on metal stand; 22 cm. in diam. no box 12 × 10 × 10 cm.

1 map: col.; 200 × 350 cm. folded to 20 × 15 cm. in plastic case 25 × 20 cm.

There are various types of Notes given in chapter/rule 3.7B for benefit of the cataloguer

Cataloguing of a Map: **For Example:**

Tourist map of West Bengal. Department of Tourism. Editor. A.K. Ghosh .-Scale 1:250,000 – 1:3500,000.- Kolkata: Government Printing Press. 2007 . There are 15 coloured maps on sheet 20" x 12" folded to 25 x 12 cm. Place names are in Bengali and English in red coloured ink. Also shown important tourist places in West Bengal.

	West Bengal.	Department of Tourism.
	1:250,000- guide.	Tourist map of West Bengal [map]/edited by A.K. Ghosh.-scale 1:3500,00.-Kolkata: Government Printing Press,2007. 15 map:col.,on sheet 20"x12" folded to 25x12cm. Useful guide for tourists. Places names in Bengali and English in red colour ink. Shows important tourist places of West Bengal along with a separate 1.West Bengal-Description and Travel.I.Ghosh,A.K.II. Title. O

8.5.2 Manuscripts (Including Manuscript Collections)

Rule 4.0a. Scope

Rule 4.0A10

The rules in this chapter cover the description of manuscripts (including type-script) materials of all kinds, including manuscript books, dissertations, letters, speeches, etc., legal papers (including printed forms completed in manuscript), and collections of such manuscripts for reproductions of manuscripts published in multiple copies, see chapter 2 or 11 as appropriate for manuscript cartographic items, see also chapter 3 for manuscript music and chapter 5 of AACR2R.

Rule 4.0B1 Chief Source of Information

The chief source of information for manuscripts is the manuscript itself. Within manuscripts, use (in this order of preference) information from the : Title page, Colophon Caption , Heading etc., Content of the manuscript and reference sources. For collection of manuscripts, treat the whole collection as the chief source.

Rule 4.2B. Edition statement:

Transcribe a statement relating to a version of a manuscript that is different from other versions.

Rule 4.4B. Date of manuscript

Rule 4.5B1. Give the number of leaves or pages

e.g. VII, 18 leaves; leaves 43-43; [3] , 122 p.

Rule 4.5C1. (Physical Details):

Name the material on which the item being described is written if it is other than paper e.g. [2] leaves : vellum; [1] leaf : parchment.

Rule 4.5 D for dimensions

Rule 4.6. Series area :

This area is not used for manuscripts

Rule 4.7B Notes area

Edition, place of publication, name of publisher, series, ISBN and materials are not used as these are irrelevant to manuscript. If it is handwritten, the word is used as holograph (s). Ms will be used for manuscript and Mss for collections of manuscripts.

Hamare Deshbashiyon. Manuscript. Hindi Poems.

10 leaves. 2 leaves stained by water. There is a signed holograph on the manuscript. Author- Ravindra Shastri

Example

Hamare Deshbashiyon. Manuscript. Hindi Poems.

10 leaves. 2 leaves stained by water. There is a signed holograph on the manuscript. Author- Ravindra Shastri

Shastri, Ravindra

Hamare Deshbashiyon [manuscript]/ by Ravindra Shastri. – 1946

10 leaves; 24cm.

Holograph, signed

Poem

In Hindi

Two leaves are stained by water

1. Hindi Poetry I Title

	Shastri Ravindra	
	1946	Hamare Deshbashiyon[manuscript]/by Ravindra Shastri.-[s.l:s.n], 10 leaves; 24cm. Holograph, signed poem in Hindi. Two leaves are stained by water. 1. Hindi Poetry. I. Title.

8.5.3 Music

Rule 5.0a Scope

The rules in chapter 5, AACR 2 (Music) cover the description of published music. They do not cover the other unpublished music in detail, though the use of an additional term in the physical description (5.58) and the use of specific provisions of chapter 4 (Manuscripts) will furnish a sufficiently detailed description for the general library catalogue. For the description of recorded music, see chapter 6. For microform reproduction of music, see chapter 11.

Rule 5.0B 1 Chief Source of Information

If the title page consists of a list of titles including the title of the item being catalogued, use as the chief source of information whichever of these: title page, the cover, or the caption furnishes the fullest information. In all other cases, use the title page or title page substitute (see 2.OB1) as the chief source of information. If information is not available from the chief source, take it from the following sources (in this order of preference): Caption, Cover, Colophon, Other Preliminaries and other sources

Rule 5.5 B 1 – 5.5 B 3: Extent of item

For example: 1 score + 1 piano conductor or part (Bp.) + 16 parts.

Rule 5.5 C 1 : Illustration

Rule 5.5 D 1 : Dimensions

Rule 5. E 1 : Accompanying material

Rule 5.7 B 1-5.7 B 20, : Note area

For Example:

Home sweet home: A musical play David Smith. Lyrics - John Austin. Book by Derek Edward. Publisher, Random House. New York. 1976

One vocal score. 10 inches.

	Smith, David	
	John Austin;	Home sweet home: a musical play[music]/by David Smith; Lyrics by book by Derek Edward.-New York: Random House,1976. 1vocal score(51p.); 10in.I.Austin,John. II.Edward,Derek.III. Title. O

8.5.4 Sound Recordings

Rule 6.01 Scope

In Chapter 6 of AACR2R, the rules cover the description of sound recording in all media, i.e. discs, tapes (open reel to reel, catridges, cassettes) piano rolls and sound recordings on film. The use of appropriate specification in the physical description (*Rule 6.5*) and special notes will furnish a sufficiently detailed description for such item.

Rule 6.0 B1 Chief Source of Information

The chief source of information for each major type of sound recording is as follows:

Type	**Chief Source of Information**
Disc	Dis and label
Tape (open reel-to- reel)	Reel and label
Tape cassette	Cassette and label
Tape cartridge	Cartridge and label
Roll	Label
Sound recording on film	Container and label

Treat accompanying textual material or container as the chief source of information if it furnishes a collective title. If information is not available from the chief source, take it from the following sources (in this order of preference).

Accompanying textual material

Container (sleeve, box, etc.) and other sources.

Rule 6.1E Other Title Information :

If any sub-title i.e. the other title information is to be transcribed after the title putting a colon (*Rule 6.1E*)

Hello Dolly ! [GMD] : Original motion picture sound track

Rule 6.1F   Statement of Responsibility :

It relates to the heading chosen for the type of sound recording which is to be incorporated by giving a diagonal slash after the Title/Sub-title (*Rule 6.1F*). e.g. Melville [GMD]/ written and narrated by Thomas Heath.It is to be remembered that, the cataloguer is required to add a word or short phrase to the statement of responsibility if the relationship between the title and the person(s) or body named in the statement is not clear, e.g. Born to run [GMD]/ [written and performed by] Ken Russel.

Rule 6. 7B 21   Items without a Collective Title :

In case a sound recording lacks a collective title, either it is to be described as a unit or separate description for each separate titled part is to be made as mentioned in the Notes section (6.*7B 21*).

Place of Publication, Distribution etc. **:** e.g. London : RCA Victor : Distributed by Middle Earth Co.

It is more appropriate to prefer a trade or brand name rather than the name of the publisher if both appear on the label. Prefer label information rather than information appearing in accompanying material or container.

If the statement of function of publisher, distributor is found, the information is to be added as mentioned below.

New York : Sunflower; [London] : Virgin Records [distributor] Date of Publication, distribution etc.

The date of a published sound recording is to be provided after the name of publisher after giving a comma.

Rule 6. 7B7   But if the date of recording appears on a published sound recording, it is to be given in a note (Sec *6. 7B7*)

[New York] Music Guild, 1971

Note : Recorded in 1965.

Sometimes the name of the publisher is unknown, whereas the place and name of the manufacturer are found in the item, then that place and name are to be cited.

[s.l. : s.n.], 1986 (London : High Fidelity Sound Studies)

Rule 6.5B   Extent of Item

The number of physical units of sound recording be noted as

1) Sound discs

2) Sound cassettes

The playing time of a sound recording is to be transcribed as in many cases, playing time is included on the item as part of the information on the label, container or accompanying material. e.g. 2 sound cassettes (35 min. each). 1 sound disc (45 min.)

Rule 6.5C   Other Physical Details

The type of recording i.e. the way in which the sound is encoded on the item. e.g. 1 sound cassette (50 min.) : digital

In case of a sound cassette, the playing speed can be given.

e.g. 1 sound cassette (40 min.) : analog, 1 5/16ips.

Rule 6.5D5 Dimension for Sound Cassette

Give the dimensions of a cassette if other than the standard dimension (e.g. the standard dimensions of an analog cassette are 3 7/8 × 2 1/2 in.). Give the width of a tape if other than the standard width (e.g. the e.g. 1 sound cassette (85 min.): analog, mono, 71/4 × 31/2 in; ¼ in. tape.

Rule 6.6B Series Statement

The series statement is to be recorded in parentheses after the last element of the physical description preceded by a full stop, space, dash, space.

e.g. (Audio – cassette library for professional librarians; 48).

Rule 6.7B Notes

There are 19 types of notes given and if the cataloguer decides that a particular information is of importance for the users' viewpoint, then it can be mentioned in the order as mentioned below by starting a new paragraph for each note or precede each note by a full stop, dash, space.

Example

Farewell My Friend. Sound Recordings. Arranged and composed by Roger Taylor.Stereo. Also issued in cassette

Songs: Roger Taylor with orchestra and in part

With background vocals

Taylor, Roger

Farewell my friend [sound recordings] / [composed and Arranged by] / Roger Taylor.- London: Royalty Records, 1991.

1 sound disc: 33 1/3 rpm, stereo: 12 in.

Title from container

Also issued in cassette tape

Songs: Roger Taylor with orchestra and in part

With background vocals

Fare well my	Friend [sound recording]/[composed and arranged by Roger Taylor].-
	London: Royalty Records,1991. 1 sound disc: 33 1/3 rpm,stereo: 12in. Title from container. Also issued in cassette tape. Songs: Roger Taylor with orchestra and part with background vocals. O

8.5.5 Motion Pictures and Video Recordings

In AACR 2, Chapter 7 prescribes rules for the descriptive cataloguing of all types of media involving a sequence of images projected in rapid succession so that they can create the illusion of movement. These include video recordings with all of their manifestations. This chapter also includes motion pictures which come in many packages such as film cassette and film reels, etc.

Rule 7.0B *Sources of Information*

The chief source of information for motion pictures and video recordings as it is for other library items, is the work itself. As you know, when a book is catalogued, the title page is the chief source of information. Likewise for motion pictures and video recordings the chief source of information is the item itself, its container (and container label) if the container is an integral part of the piece. However, if the information is not available from the chief source, then the following sources can be consulted :

a) Accompanying textual material (e.g. scripts. shot lists, publicity material).

b) Container (if not an integral part of the piece.) and other sources.

Rule 7.1B – 7.1E1 (For title proper, GMD, Paralle titles and other title information).

The rules for description of information are identical as for sound recordings.

Rule 7.1E2

If the item is a trailer containing extracts from a larger film, add [trailer] as other title information.

Home sweet home [GMD] : [trailer]

Rule 7.1F *Statement of Responsibility:*

This area covers those persons or bodies credited in the chief source of information with participation in the production of a film (e.g. as producer, director, animator) that are considered to be of major importance, while all other statements of responsibility are to be recorded in Notes.

Classroom control [GMD]/ University of London

Audiovisual centre ; produced and directed by W.C. Collins.

Rule 7.1G	*Items without a Collective Title :* Follow the rules for description as mentioned under 6.1 G.

Rule 7.2B	*Edition Statement*: Same rules as for sound recording (6.2B).

Rule 7.4. Publication, Distribution, etc. : Exactly identical with sound recording.

Rule 7.5 Extent of Item

The physical units of a motion picture or video recording can be given the number of parts in Arabic numerals as mentioned under : (a) 1 videocassette (b) 2 video discs (c) 3 film reels.

Rule 7.5B2	*Playing Time:*

It is to be recorded as indicated in the item. If the playing time is mentioned as 'about 10 minutes', then it will be written as (ca. 10 min.). (a) 2 film cassettes (25 min. each) (b) 1 Video reel (4 min., 20 sec.)

Rule 7.5C	*Other Physical Details:*

It recommends to give sound characteristics, colour etc in extent of item area.

1 video cassette (20 min.) : sd., col. with b & w sequences.

Rules 7.5D and 7.5E:

for incorporating the dimension and accompanying material.

Rule 7.7 Note Area

It facilitates scope for recording 18 types of notes for essential information about the item which are not recorded previously in the entry.

Rule 7.7B1	*Nature or Form:*

Make notes on the nature or form of a motion picture or video recording.

Documentary, T.V. Play

Rule 7.7B6	*Statement of Responsibility*:

List featured players, performers, narrators, and /or presenters.

Presenter: Chris Lewis

Cast: Jack Robinson

Credits : The persons who have contributed to the artistic/or technical production of a motion picture or video recording whose name(s) are not named in the state of responsibility.

Credits : Screen play, A. Ronaldo; music, Robin Smith and followed by name for camera, editor, etc.

Rule 7.7B10	*Physical Description*:

It consists of sound characteristics, length of the film, technicolour, video recording system (laser optical CAV, VHS Hi-Fi) and three-dimensional film.

Rule 7.7B11 Accompanying Material:

The information like 'cast list and credits on box'; 'with shot list' be provided in this note area.

Rule 7.7B14 Audience :

Make a brief note of the intended audience.

Intended audience : Elementary grades.

Rule 7.7B16 Other Formats :

Give the details of other formats in which the content of the item has been issued.

Issued also as cassette (VHS or Sony U-Matic).

Rule 7.7B17 Record a brief objective summary of the contents of an item.

Rule 7.7B18 Contents:

Give the titles of individual works contained in, or the parts of a motion picture or video recording. Add to each title, if any statement of responsibility area and the duration if indicated.

Example

Nutritional Snacks and Fast Food. A film. Director: Johnny Lever. National Film Board of Canada.

International Film Bureau. 1989. One film reel. 52 minutes. Teacher's guide also attached. For primary and intermediate students. Produced to show that snacks fulfill nutritional needs of students and cannot be equated with junk food.

Nutritional	Snacks and fast foods [motion picture]/National Film Boardof Canada;
Children	Director,Jojny Lever.- Toronto: International Film Bureau,1989. 1film reel (52min.);sd,col.;16mm.+1 teacher guide. Title from data sheet. Technicolour. Intended audience: Primary and Intermediate Grade. Summary: Shows that snacks fulfil all the nutritional need of the and can not be equated with junk food .

8.5.6 Graphic Materials

Rule 8.0A. Scope

It covers the description of graphic materials of all kinds whether opaque (e.g. two dimensional art originals and reproductions, charts, photographs, technical drawings) or intended to be projected or viewed (e.g. filmstrips, radiographs, slides) and collections of such graphic materials.

Rule 8.0B1. Chief Source of Information:

It is the item itself including any labels, etc. that are permanently affixed to the item or a container that is an integral part of the item. If the item being described consists of two or more separate physical parts (slide set, etc), treat a container that is the unifying element as the chief source of information if it furnishes a collective title and the items themselves and their labels do not. If the information is not available form the chief source, take it form the following source (in order of preference) :

Container (box, frame, etc.)

Accompanying textural materials (manuals, leaflets, etc.)

Other sources

Rule 8.4: Publication, Distribution etc. Area.

Rule 8.4 F2

Record the date of creation of an art original, unpublished photograph, or other unpublished graphic item.

Portrait of Charles Dickens – 1964

Garden flowers [GMD]/Geoff Arnold. — [1898 ?]

Fair Resemund. — [1910-1973]

(Unpublished photographs)

Rule 8.4. G : Place of manufacture, name of the manufactures, date of manufacture.

Rule 8.4 G 1 : If the name of publisher is unknown, give the name of place and manufacturer as instructed in 1.4 G, if they are found in the item and have not been recorded in a statement of responsibility.

Fig. : (s.l. : s.n. , 1966 ?) (London: Allen press)

Rule 8.5 B 1 : Record the number of physical units of a graphic item.

4 wall charts 1 filmstrip catridge

200 slides 6 stereograph reels

Rule 8.5 B 2 : Add to the designation for a filmslip, filmstrip, etc. like

1 filmstrip (26 fr.)

1 flip chart (6 sheets)

Rule 8.5 C : Other physical details.

Rule 8.5C 1: (a) Art originals. Give the medium (Chalk, Oil , Pastel etc.) and the base

(board, canvas, fabric, etc.)

1 art original : Oil on canvas.

(b) Art prints:Give the process in general terms (engraving, lithograph, etc.

(c) Charts and flip charts e.g. 1 flip chart (8 sheets) : double sided.

(d) Filmstrips and filmslips. Give an indication of the sound if the sound integral.

1 filmstrip (30 fr., 2 title fr.) : sd., col.

(e) Flash cards: 12 flash cards : col.

(f) Photographs. If the photograph is a transparency not designed for projection or negative print, indicate this. Give an indication of the colour.

1 photo : negative, b & w

(h) Slides. Give the indication of the sound if the sound is integral. Add the name of the system e. g. (3 M talking slide) after the indication of the sound.

10 slides : sd. (3 M Talking Slide), col.

(j) Technical Drawing. Give the method of reproduction if any (blue print, photocopy etc.)

1 technical drawing : blue print

Rule 8.5 D : Dimensions.

Give for all graphic materials except film strips, filmslips and stereograph the height and the width in centimeters to the next whole centimeter up.

1 Technical drawing : blue print, 87 X 87 cm.

28 photographs : b & w; 13 x 8 cm.

Rule 8.5 D5 : Slides.

Do not give the dimensions if they are 5 X 5 cm (2 x 2 in.)

1 slide : col.

1 slide : col. ; 7 x 7 cm

Rule 8.5 E : Accompanying material.

Rule 8.5 E 1 : 30 slides : col. + 1 sound disc

(30 min. analog 33 1/3 rpm. mono. ; 12 in.)

Rule 8.7 B 1 to 8.7 B 19 : Refer for various kinds of notes.

Example

Department of Navy. United States. Modern Warfare and Navy. Coloured film. 2 reels, 40 frames. 35 mm. Accompanying booklet.

Describes the strategic role of missiles in defence and attack by a country.

	Modern war	fare and missiles[film strip]/produced by U.S.Department of Navy.-
	attack by a	New York:[s.n],1983. 2reel (40fr.):sd.col.;35 mm.+1 Boklet. Summary: Describes the strategic role of missiles in defence and country. 1. U.S Department of Navy.

8.5.7 Computer Files

Rule 9.0A   *SCOPE*

Rule 9.0A1

The rules in this chapter cover the description of files that are encoded for manipulation by computer. These files comprise data and programs. Computer files may be stored on, or contained in carriers, available for direct access or by remote access.

The rules in this chapter do not cover electronic devices such as calculators etc. ; see chapter 10 for such materials. Programs residing in the permanent memory of a computer (ROM) or firmwire are considered to be part of the device and should be described in conjunction with the device (e.g. the programming language of a particular computer, such as : Applesoft in ROM).

Rule 9.0B1.   *Chief Source of Information.*

The chief source of information for computer files is the title screen (s). If there is no title screen take the information form other formally presented internal evidence (e.g. main menus and program statements).

If the information required is not available from the internal sources or the sources existed above, take it from the following sources (in this order of preference) : Physical carrier or its labels

Other published description of the file and other sources

Rule 9.3 File Characteristics Area

Rule 9.3B1 Designation:

When the information is readily available indicate the type of file. Use one of the following terms:

Computer data, Computer program (s), Computer data and program (s)

Optionally, if GMD are used (see 1.1 C1), omit computer from the file designation.

Rule 9.3B2. Number of records, statements, etc.

If a file designation is given and if the information is readily available, give the number of approximate number of files that make up the content (use file or files preceded by an Arabic numeral) and /or these other details :

a) Data. Give the number or approximate number of records and/or bytes.

   Computer data (1 file: 600 records, 2400 bytes)

b) Programs. Give the number or approximate number of statement and/or bytes.

   Computer program (1 file : 200 statements)

c) Multipart files. Give the number or approximate number of records and/or bytes, or statements and/or bytes, in each part.

Computer data (2 files : 800, 1250 record) and programs (3 files : 7260, 3490, 5076, bytes)

Rule 9.5 Physical Description Area

Rule 9.5 B1

Record the number of physical units of the carrier by giving the number of them in Arabic numerals and one of the following terms as appropriate.

Computer catridge, Computer disc, 3 computer cassettes, 2 computer laser optional card

Rule 9.5 C1

If the file is encoded to produce sound, give sd. If the file is encoded to display in two or more colours, give col.

1 computer disc : sd., col., single sided, single density, softsectored.

Rule 9.5 D 1 Give the dimensions of the physical carrier

1 computer disk : col. ; 5 ¼ in.

a) Cartridge.

   1 computer chip cartridge ; 3 ½ in.

Rule 9.5E1.

Give the details of accompanying material as instructed in 1.5E. 1 computer disk; 5 ¼ in. +1 user manual and addendum. (accompanying materials title : user manual and addendum)

Rule 9.7 Note area:

The cataloguer is free to provide any one of the notes.

Example

Wizard master. computer programme. Conceived and designed by Edward Miller. 4 files.

Santa Clara, CA. Activision. 1999.

1 basic game. Along with a manual. S y s t e m requirements: Atari 2600, left joystick container .

Miller,Edward	
Miller.- ([63p.]:col.ill. Left joystick	Wizard master[computer file]/conceived and designed by Edward computer programme (4 files).-Santra Clara, CA, Activision,1999. 1 computer chip cartridge: col.;3 1/4in.+1 base game manual ;18cm.)+1 updated game manual system requirements: Atari 2600;container. Title from cartridge lebal. I. Title. O

8.5.8 Three Dimensional Artifacts and Realia

Rule 10.0A *Scope*

Rule 10.0A1

The rules in this chapter cover the description of three-dimensional objects of all kinds (other than those covered in previous chapter), including models, dioramas, games (including puzzles and simulations) , braille cassettes, sculptures and other three dimensional artworks, exhibits machines and clothing. They also cover the description of nature all/occurring objects, including microscope specimens (or representations of them) and other specimens mounted for viewing. For the description of three-dimensional cartographic materials (e.g. relief models, globes), see chap. 3 of AACR2R.

Rule 10.0B1 Chief Source of Information:

The chief source of information for the materials covered in this chapter is the object itself together with any accompanying textual material and container issued by the publisher or manufacturer of the item. Prefer information found on the object itself (including any permanently affixed labels) to information found in the accompanying textual material or on a container.

Rule 10.4 . Publication, Distribution, etc. area.

Rule 10.4 C. Place of Publication, Distribution etc.

Rule 10.4C2.

Do not record a place of publication, distribution, etc., for a naturally occurring object (other than on mounted for viewing or packaged for presentation) or for an artifact not intended primarily for communication. Do not record s.n. in such a case.

Rule 10.5Physical Description Area

Rule 10.5 B Extent of Item (including specific material designation)

Rule 10.5B1 Record the number of physical units of a three dimensional artifact or object by giving the number of parts in arabic numerals and one of the terms listed below, as appropriate.

Art original, Art reproduction, Braille cassette, Diorama, Exhibit, Game, Microscope slide, Mock – up, Model, if non of these terms is appropriate, give the specific name of the item or the names of the parts of the item as concisely as possible.

1 Clockwork toy train, 2. Jigsaw puzzles, 3. quilts

Rule 10.5C. Other Physical Details

Rule 10.5C1. Material, When appropriate, give the material (s) of which made. If the material (s) cannot be named concisely, either omit the statement or give it in a note. Give the material of which a microscope slide is made if it is other than glass. Such as 2 models (various pieces) : polystyrene

1 diorama (various pieces) : Polystyrene

1 statue : marble

1 quilt : cotton

Rule 10.5D Dimensions

Rule 10.5D2 If the object is in a container, name the container and give it's dimensions either after the dimensions of the object or as the only dimensions.

1 model 10 pieces : col. ; 16 x 32 x 3 cm. in case 17 x 34 x 6 cm.

1 Jigsaw puzzle : wood, col. ; in box 25 x 32 x 5 cm.

Rule 10.5D3 If, in a multipart item, the objects and/or their largest or larger size, separate by a hyphen.

3 sculptures : marble; 150 – 210 cm. high.

Rule 10.7Note area

Rule 10.7B1 Nature of the item.

Give the nature of the item unless it is apparent from the rest of the description.

Study of a figure on motion

Section of fetal pig mandible

Rule 10.7B10. Physical Description.

Make notes on important physical details that are not included in the physical description area, especially if these affect the use of the item. If the physical description includes various pieces and a description of the pieces is considered to be useful, give such a description.

Four times actual size – The parts of the ear are painted to show anatomical structure.

Includes headdress, beaded shirt, trousers, and moccasins pattern : Pennsy lvania wild goose.

Contains 1 small stage, 5 foreground transparencies, 2 backgrounds, 5 story sheets, and 1 easel.

Example:

Art. Hand Weaving. Weaver: Ramesh Meher.

Patterns by Suresh Meher.

Coloured sampler. Cotton.

Meher,Ramesh	
	Hand weaving [art original].-[s.l:s.n].-1968. 1sampler: cotton, ccol.;125x30cm. Woven by Ramesh Meher. Pattern: Suresh Meher. I.Meher,Suresh. II. Title.

8.5.9 Microforms

Rule 11.0A Scope

It includes microfilms, microfiches, microopaques and aperture cards. Microforms may be reproductions of existing textual or graphic material or they may be original publications.

Rule 11.0B1 : Chief Source of Information

The chief source of information for microfilms is the title frame (i.e. a frame, usually at the beginning of the item, bearing the full title and, normally, publication details of the item.) The chief source of information for aperture cards is, in the case of a set of cards, the title card, or, in the case of a set of cards microfiche, and microopaques is the title frame. If there is no such information or if the information is insufficient, treat the eye-readable data printed at the top of the fische or opaque as chief source of information.

If information is not available from the chief source, take it from the following sources

(in order of preference) :

The rest of the item (including a container that is an integral part of the item)

Container, Accompanying eye-readable material, Any other source

Rule 11.3 :

Special data for cartographic materials, music and serials

Rule 11.3A **:**Cartographic materials

Rule 11.3A1 **:**Give the mathematical data of a cartographic item in microform as instructed in 3.3.

Rule 11.3B **:** Music.

Rule 11.3B1 **:** Give the physical presentation of music in microform as instructed in 5.3.

Rule 11.3C **:**Serials

Rule 11.3C1 **:** Record the numeric and /or alphabetic chronological or other designation of a serial microform or a serial reproduced in microform as instructed in 12.3.

Rule 11.5B1 **:**Record the number of physical units of a microform as : 20 aperture cards ; 3 microfilms reels, 4 microfiches, etc.

Rule 11.5C1 **:**If a microform is negative, then indicate as : 1 microfilm : negative.

Rule 11.5C2 **:**If it consists of, illustrations, then indicate as 3 microfiche : ill. 2 microfilm reels : col. ill.

Rule 11.5D2 **:**Aperture card. Give the height and the width of an aperture card mount in cms.

e.g. 10 aperture cards 9 X 19 cms. 16 aperture cards. 6 x 16 cm.

Rule 11.5D3 **:**Microfiches. Give the height X width of a microfiche in cm.

e.g. 2 microfiches ; 10 x 15 cm – 14 x 17 cm

Rule 11.5D4 **:**Give the width of a microfilm in millimeters.

1 microfilm reel, 16 mm.

1 microfilm catridge, 35 mm.

Rule 11.5E **:**Accompanying material

1 microfilm reel, 16 mm + 1 pamphlet (3o p. : ill., 20cm.)

Rule 11.7B **:**Notes

Rule 11.7B1 :Nature, scope, or artistic or other form of an item,

Rule 11.7B2 :Language

French, with English translations

Rule 11.7B3 **:**Source of title proper. Make notes on the source of the title proper if it is other than the chief source of information.

Title from container.

Rule 11.7B14 **:**Audience. Make a brief note of the intended audience.

For high school students.

Rule 11.7B17 **:**Summary

Rule 11.7B18 **:**Contents.

Example:

Sixth Conference on Alternative to Fuel. Held at Mumbai. 2007. Proceedings.

Editor: Ray and K. Rao. ONGC.

4 microfiches : negative

Conference	on Alternative to Fuel***(6th:2007:Mumbai)***
2007/edited	Conference on alternative to fuel: no.6[micro form]: proceedings,by s. Ray and K.Rao.- Mumbai: ONGC, 2007. 4 microfiches: negative. I.Ray,S. II.Rao,K.III.ONGC.

8.5.10 Electronic Resources

Scope

The rules in chapter-9 of AACR2R cover the description of electronic resources. Electronic resources consist of data (information representation numbers, text, graphics, images, maps, moving images, music, sounds, etc.) programs (instruction, etc., that process the data for use), or combinations of data and programs. (Rule 9.OA1)

Prescribed Sources of Information:

Information issued by the publisher, creator, etc., container

The title and statement of responsibility, GMD, parallel titles, other title information,Edition are same as other NBM.

But if an electronic resource lacks an edition statement but is known to contain significant changes from other editions (e.g., changes in the data involving content, standardized coding, etc.; changes in the programming including changes in the program statements, programming language, and programming routines and operation; the addition of sound of graphics; improvement of graphics), supply a suitable brief statement in the language and script of the title proper and enclose it in square brackets. (Rule 9.2B3)

[Windows 95 ed.]

Types of Resource: The term for the electronic resource to be catalogued be indicated. e.g. Electronic data and program (s)

Extent of Resource: The number of records be given

Electronic data (1 file: 500 records, 180,000 bytes)

Electronic program (1950 statements)

Electronic data (2 files: 800, 1250 records) and programs (3 files: 7260, 3490, 5070 bytes)

Publication, Distribution, etc. Area

The bibliographic description of the above areas are exactly same as other materials. However do not record the place of publication, distribution, publisher, distributor for an unpublished electronic resource, so do not record s.l. and s.n. in such cases.

Extent of Item: The number of physical units of the carrier be given

e.g. 1 computer disk

2 Photo CDs

Other Physical Details

e.g. 1 computer disk: sd., col., single sided,. Single density, soft sectored

1 computer optical disk: col. ; 4 ¾ in.

1 computer tape cassette ; 3 ½ x 2 ½ in.

For accompanying material:

1 computer disk; 3 ½ in. + 1 demonstration disk.

Rule 9.7B Note Area

There are 22 types of notes are provided for the benefit of the cataloguer. Most of the notes are just same as other non-book materials. However some of the notes are different are provided below:

*Rule 9.7B1***.** Nature and scope, system requirement, and mode of access

Rule 9.7B8 Type and extent of resources.

e.g. File size: 520,300,280,400,320 records

File size unknown

*Rule 9.7B19 Numbers***.**

Give important numbers associated with the item other than ISBN or ISSM. e.g. APX-10050

*Rule 9.7B22. Item described***.**

For remote access resources, always give the date on which the resource was viewed for description.

e.g. Description based on contents viewed June. 16, 2008.

8.6 CHANGES IN AACR 2R AND AMENDMENTS 2002

There are significant changes in the 2002 revision which are discussed below:

Chapter3 (Catrographic Materials)

The changes to chapter 3 are of three major types:

- Additional rules or additions to existing rules for the description of cartographic materials in electronic form;
- Miscellaneous changes to existing rules to bring them into line with current practice:
- Editorial changes.

The first category is the most substantial, involving changes to the mathematical and other material specific details area {formerly: Mathematical data area). Three new rules were added: rule 3.3E (Type and extent of resources); rule 3.3F (Digital graphic representation); and, rule 3.3G (Numbering related to serials). The overall goal is to enable more accurate description of cartographic materials that are electronic resources and/or continuing resources. Rule 3.3D (Statement of coordinates and equinox) has been changed to allow the recording of coordinates in decimal degrees as well as in degrees, minutes, and seconds. Additional examples have been added to rule 3.7B8 (Mathematical and specific details) to reflect the addition of rule 3.3F and the changes to rule 3.3D changes in the second category include: the changing of "of map section" to "section," and "relief model" as other physical details in rule 3.5C1.Editorial changes have been made as necessary to match new terminology and to relief changes made in other chapters.

Chapter 12 (Continuing Resources) and other Related Rules

The scope of chapter12, now called "Continuing Resources" instead of serials, expanded to include resources that have either not been covered in the rules or not adequately covered. Chapter 12 now encompasses:

- Successively issued resources (i.e., serials);
- Ongoing integrating resources (e.g., updating loose-leafs, updating Web sites);
- Some categories of finite resources i.e., reprints of serials, resources with character of serials but whose duration is limited, and finite integrating resources).

General Material Designation for Cartographic Materials

The general material designations "globe" and "map" in list 2 in rule 1.1C1 have been replaced with the single general material designation "cartographic material."

Chapter 9 (Electronic Resources) and other Related Rules

The revision to chapter 9, now called "Electronic Resources," fall into two categories:

- Changes to align to the International Standard Bibliographic Description for Electronic Resources (ISBD(ER));
- Changes to accommodate the particular nature of electronic resources.

Changes falling in the first category include: the clarification of the scope of chapter 9 and the provision of a distinction between direct access and remote access electronic resources; the addition of an instruction at new rule 9.4B2 to consider all remote access electronic resources as published; and, changing the name of the file, characteristics area (9.3) to "Type and extent resource area."

Changes in the second category include: changing of the chief source of information from the title screen to the resource and the removal of the preference given to internal sources; the addition of an option at rule 9.5B1 to allow for the use of conventional terminology to describe a physical carrier e.g., "1 CD-ROM" instead of "1 computer optical disc"; and, the addition of rule 9.7B22 (item described) to instruct the cataloguer to always give the date viewed when describing remote access electronic resources. In addition, more current examples of electronic resources have been included and the glossary has been updated with new and revised definitions.

The general material designation in list 1 and list 2 of rule 1.1C1 has been updated from "computer file" to "electronic resource."

"Work" in Music

Uniform Titles

Changes have been made in chapter 25 to clarify the use of the term "work" in the rules for music uniform titles, including moving the definition in the glossary for "musical work" to a footnote to rule 25.25A

8.7 RESOURCES DESCRIPTION AND ACCESS (RDA)

What is RDA?

RDA stands for "Resources Description and Access", the new standard that will be the successor to AACR2R.

Introduction

RDA has been developed as a new standard resource description and access design for the digital world.

It has two parts (A and B) instead of the three parts (Parts I, II, and III) originally proposed. Part A covers description and access elements and part B covers authority control for the form of access point.

It provides:

- A flexible framework for describing all resources – analog and digital
- Data that is readily adaptable to new and emerging database structures
- Data that is compatible with existing records in online library catalogues.

Objectives and Principles for the Design of RDA

The objectives and principles set out in this section are those that govern the overall design of RDA as a standard for resource description and access. They address matters of scope, formulation, currency, etc.

Comprehensiveness, Consistency, Clarity, Rationality, Currency, Compatibility, Adaptability, Ease and efficiency of use, Format, Generalisation, Specificity, Non-redundancy, Terminology, Reference structure,

The Key Features of RDA

The first key feature is that RDA will be designed as an online product for use in a web environment. This will allow different views of the rules to be presented, for example to present a concise version of the rules, or rules of particular interest to, say, those cataloguing music. The second key feature is that the structure will be aligned more directly with the FRBR and FRAR models. This more flexible framework will help address the challenges of describing digital resources. The data that is produced should also be more readily adaptable to newly emerging, more efficient, database structure.

The third key feature is that instructions for recording data will be presented independently of guidelines for data display. This will provide more flexibility, enabling the records to be used in a variety of online environments with different structure of syntax for data storage or display.

The final key feature is that RDA will contain clear general instructions, written in plain English. The instructions will be supplemented by detailed rules or by references to other standards as needed, and they will be backed by guidance on the principles behind the rules. This will enable the code to be used more readily beyond the library world. Together these changes will pave the way for improved catalogue design and a grater user focus.

Purpose of RDA

RDA – Resource Description and Access will be a new standard for resource description and access, designed for the digital world.

Built on foundations establishment by the Anglo – American Cataloguing Rules (AACR), RDA will provide a comprehensive set of guidelines and instructions on resource description and access covering all types of content and media.

RDA will enable users of library catalogues and other systems of information organisation to find, identify, select, and obtain resources appropriate to their information needs.

Need of a New Standard

AACR2 was first published in 1978. Although it has been updated many times through the revision process that was established by the JSC, it is largely designed for an environment dominated by the card catalog. The International Conference on the Principles and Future Development of AACR that was held in Toronto in 1997 identified substantive problems with AACR2. Although the updates issued in the years following that conference addressed some of these problems, it became clear that a fundamental rethinking of the code was required to respond fully to the challenges and opportunities of the digital world.

FRBR and its Relationship with RDA

The acronym "FRBR" stands for Functional Requirements for Bibliographic Records. FRBR was developed by an IFLA Study Group (1992-1997), and IFLA continues to monitor the application of FRBR and promotes its use.

FRBR includes a conceptual model of entities and relationship and attributes; identifies specific user tasks that bibliographic records are intended to fulfill: find, identify, select, obtain; and recommends a set of elements for inclusion in national bibliographic records.

FRBR provides the conceptual foundation for RDA. RDA includes the FRBR terminology when appropriate (for example, use of the names of bibliographic entities: "work" "expression", "manifestation", and "item"), will use the FRBR attributes as the basis for specific data elements to be included in bibliographic descriptions, will address FRBR relationship, and will use the FRBR user tasks (find, identify, select, obtain) as the basis for defining a set of mandatory data elements. RDA will also use FRAD (Functional Requirements for Authority Data) as the basis for instructions on authority control.

Each section contains general guidelines and a chapter for each entity. Each chapter is associated with an FRBR user task. The chapters on recording attributes and relationships for concepts, objects, and events are placeholders for completeness in mapping to FRBR and FRAD and may be developed further in future releases of RDA. The instruction on recording attributes and relationship for places will not initially go beyond the scope of AACR2 – chapter 23

Outline of the RDA Structure

There are 10 sections which focus firstly on recording attributes for FRBR entities, and then on recording relationship between these entities.

Recording attributes (Section 1- 4)

Recording relationships (Section5-10)

RDA does not include instructions on how to create or format subject headings, it does refer to them with regard to their relationship to Group 1 FRBR entities. One of a number of key elements in RDA is that it establishes a clear line of separation between the recordings of data and presentation of data. The ISBD order of areas, data elements and punctuation will not be required. Information on presenting data RDA data in an ISBD display have been provided separately. The concept of main entry as used in a card catalogue is no longer applicable in online catalogues, and this term will not be used in RDA. That records created by using RDA would be compatible with AACR2 records.

AACR 2 and MARC 21 are two different standards designed for two different purposes. AACR 2 is largely a content and display standard while MARC 21 is largely an encoding standard. RDA is being developed only as a content standard rather than as an encoding standard.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

4) State the physical description of microform.

....

....

....

....

8.8 SUMMARY

In this Unit, attempt has been made to describe the cataloguing rules item-wise for each of the non-book materials as laid down in AACR-2. Also many appropriate and relevant examples are provided along with each area of bibliographic description for easy understanding. Besides, the physical description of each type of non-documentary resource which differ from one another are also mentioned exhaustively after explaining rules vividly and clearly. The 'Entry Format' providing all the necessary items of bibliographic descriptions provided at the end of description of various types of non-book materials are discussed from 2.3 – 2.10. Even in many cases, entries are prepared citing the information about a non-book material following the rules for cataloguing (including the choice of access point and areas of description) for quick and easy understanding. There is also a discuss in about electronic resources, FRBR and RDA.

8.9 ANSWERS TO SELF CHECK EXERCISES

1) 1) Bringing direct solution to the space problem in the libraries felt world over, it saves 70 to 90% less space than the books.

2) Transporting these documents easily from one library to other, having the quality of portability.

3) Saving the documents from its decay as felt in case of conventional documents of hard copies on paper.

4) Requiring special storing devices, which makes it possible to store the collection of a large library in a limited space.

5) Audio-visual documents require the help of machine process to read which increase the reading interest of the users.

2) **Types of NBM**

There are various types of non-book materials which are mentioned as under:

Cartographic Materials:

- Ariel Chart.
- Ariel Remote Sensing image.
- Atlas.
- Celestial Globe.
- Chart.
- Globes
- Map.
- Plan.
- Relief Model.
- Remote Sensing Image.
- Space Sensing Image.
- Topographic Drawings.

Manuscripts:

- Leaf.
- Item (For Collection of Mans.)
- Box.

Music:

- Score.
- Condensed
- Minature Score.
  - Chorus Score.

Sound Recordings:

- Sound Catridge
- Sound Cassette.
- Sound Disc
- Sound Track Film Reel. (Cassette)

Motion Pictures and Video Recordings:

- Film Catridge
- Cassette.
- Video Cassette
- Video Disc
- Video Reel

Computer Files:

- Data Files.
- Program File.
- Object Program

Three-Dimensional Artefacts and Realia:

- Art Original
- Realia.
- Game
- Diorama
- Model

Graphic Materials:

- Art Original.
- Art Print.
- Art Reproduction.
- Chart.
- Film Strip.
- Kit.
- Photograph
- Picture.
- Post Card.
- Slide.
- Technical Drawing
- Transparency

Microform:

- Aperture Card.
- Microfiche
- Micro Film Catridge Cassette Reel

3) ➢ Information collected from one source in the NBM may differ with that obtained from another source of the documents.

➢ It may be harder to reach the cataloguing decisions with reference to choice of access point i.e. determination of heading or in other wards to decides the person who is chiefly responsible for the intellectual content of the document, which is less experienced in case of books and serials.

➢ Information about physical description of different types of NBM definitely creates problem for cataloguer rather than that of conventional documents.

➢ Very difficult to obtain information through the naked eyes as it requires special equipments.

4) The chief source of information for motion pictures and video recordings as it is for other library items, is the work itself. As you know, when a book is catalogued, the title page is the chief source of information. Likewise the item itself, its container (and container label) if the container is an integral part of the piece. However, if the information is not available from the chief source, then the following sources can be consulted :

c) Accompanying textual material (e.g. scripts. shot lists, publicity material).

d) Container (if not an integral part of the piece.) and Other sources.

The chief source of information for computer files is the title screen (s). If there is no title screen take the information form other formally presented internal evidence (e.g. main menus, program statements).

5)

11.5B1 : Record the number of physical units of a microform as : 20 aperture cards ; 3 microfilms reels, 4 microfiches etc.

11.5C1 : If a microform is negative, then indicate as : 1 microfilm : negative.

11.5C2 : If it consists of, illustrations, then indicate as 3 microfiche : ill. 2 microfilm reels : col. ill.

11.5D2 : Aperture card. Give the height and the width of an aperture card mount in cms.

e.g. 10 aperture cards 9 X 19 cms. 16 aperture cards. 6 x 16 cm.

11.5D3 : Microfiches. Give the height X width of a microfiche in cm.

e.g. 2 microfiches ; 10 x 15 cm – 14 x 17 cm

11.5D4 : Give the width of a microfilm in millimeters.

1 microfilm reel, 16 mm.

1 microfilm catridge, 35 mm.

11.5E : Accompanying material

1 microfilm reel, 16 mm + 1 pamphlet (3o p. : ill., 20cm.)

8.10 KEYWORDS

Access Point	:	A name, term, etc under which a bibliographic record may be searched and identified.
Accompanying Materials	:	Material issued with and intended to be used with, the item being catalogued.
Artifact	:	Any object made or modified by one or more persons.
Art Original	:	An original two – or three – dimensional work of art created by the artist (e.g. a painting, drawing or sculpture).
Audio Tape	:	A generic term designating a sound recording on magnetic tape.
Computer File	:	A file (data and / or programs) encoded for manipulation by computer.
Digital	:	Refers to the use of discrete signals for representing data in the form of numbers or characters. Most forms of digital representation in data processing have been done on the basis of binary numbers.
Film Strip	:	A length of film containing a succession of images intended for projection one at a time, with or without recorded sound.
Microform	:	A generic term for any medium, transparent or opaque, bearing micro-images.
Realia	:	An artifact or a naturally occurring entity, as opposed to a replica.
Score	:	A series of staves on which all the different instrumental and/or vocal parts of a musical work are written.
Sound Recording	:	A recording on which sound vibrations have been registered by mechanical or electrical means so that the sound may be reproduced.
Title Frame	:	A frame containing printed or written material not part of the subject content of the item.
Video Recording	:	A recording on which visual images, usually in motion and accompanied by sound, have been registered; designed for playback by means of a television set.

8.11 REFERENCES AND FURTHER READING

Aichele, Jean et.al. *A Manual of AACR2: Examples for Motion pictures and Video Recordings.* Ed Edward Swanson and Marilyn H. McClaskey. Minnesota:Soldier Creek Press, 1981. Print.

Anglo-American Cataloguing Rules. 2nd rev. ed. Chicago: ALA, 1988. Print.

Beacom, Matthew. "Crossing a Digital Divide: AACR 2 and Unaddressed Problems of Networked Resources." Paper presented at the *Library of Congress Bicentennial Conference on Bibliographic Control for the New Millennium*, November 15-17, 2000. Print.

Daily, J.E. *Organizing Non-print Materials*. New York, Marcel Dekker, 1986. Print.

Fleischer, Eugene Band Godman, Helen. *Cataloguing Non-print Media : A Manual based on the Anglo-American Cataloguing Rules II.* New York: Neal-Schuman, 1980. Print.

Frost, Carolyn O. *Cataloguing Non-book Materials: Problems in Theory and Practice*. Littleton Colo: Libraries Unlimited, 1983. Print.

Gorman, Michael. "What Is the Future of Cataloguers and Cataloguing?" Paper presented at the *63rd IFLA General Conference*, August 31-September 5, 1997. Print.

Horner, John Leonard. *Special Cataloguing with Particular Reference to Music, Maps, Serials and the Multimedia Computerized Catalogue.* London: Clive Bingley, 1973. Print.

Hsieh-Yee, Ingrid. Chapter 7: "Internet Resources" (p. 203-253). *Organizing Audiovisual and Electronic Resources for Access: A Cataloging Guide*. Littleton Colo:Libraries Unlimited, 2000. Print.

Huthwaite, Ann. "AACR2 and Its Place in the Digital World: Near-term Revisions and Long-term Direction." Paper presented at the *Library of Congress Bicentennial Conference on Bibliographic Control for the New Millennium,* November 15-17, 2000. Print.

Maxwell, Margaret F. *Handbook of AACR2* : *Explaining and Illustrating Anglo – American Cataloguing Rules*. 2nd ed. Chicago, American Library Association, 1980. Print.

Olson, Nancy B. *Cataloguing of Audio - Visual Materials: A Manual Based on AACR2.* Mankoto, Minn, Minnesota: Scholarly Press, 1981. Print.

Rout, R K and B K.Choudhury. *Illustrative Examples in Cataloguing of Non-book Materials (According to AACR2, 1988 R)*. New Delhi: Reliance Publishing House, 1996. Print.

Swanson, Edward. *A Manual of AACR2 Examples for Manuscripts*. Lake Crystal; Minnesotta : Soldier Greek Press, 1981. Print.

Teague, S. John. Microform. *Video and Electronic Media Librarianship*. London, Butterworths, 1985. Print.

CC: DA/Chair/2004-05/03.January 10, 2005 page 1.

AACR3: Resource Description and Access. Web. 21 September 2012.< http:// www.libraries.psu.edu/tas/jca/ccda/docs/chair22.pdf>.

Websites

http://www.collectioncanada.gc.ca/jsc/rdafaq.html http://ds.internic.net/z39.50/z39.50html

http://www.ifla.org/V1/3/nd1/isbdlist.htm

http://www.ifla.org/VII/S13/frbr.htm

http://www.unimarc.info/bibliographic/2.3/en/chapter1

http://www.ddb.de/news/pdf/code comp 2003 europe 2 pdf

http://www.loc.gov./marc/umb/um01206.html

Lynch, C.A. (1994, Aug-23). Using the Z39.50 inf. Retrieval protocol in the internet environment (Draft RFC). Available via anonymous FTP at <ds. Internet.net

http://lcweb.loc.gov/marc/umb/

http://www.csdl.tamu.edu/DL95/papers/levy/levy.html

http://lcweb.loc.gov/catdir/bibcontrol/huthwaite_paper.html

http://www.ifla.org.sg/VII/s13/wgfrbr/faq.htm

http://www.ifla.org/VII/s13/pubs/isbd.htm

http://lcweb.loc.gov/catdir/bibcontrol/huthwaite_paper.html

http://lcweb.loc.gov/catdir/bibcontrol/beacom_paper.html

UNIT 9 BASICS OF SUBJECT INDEXING

Structure

9.0 OBJECTIVES

After reading this Unit, you will be able to:

- discuss the origin and development of subject indexing;
- explain the concept and purposes of subject indexing;
- discuss the principles of indexing;
- explain the different stages of intellectual operations in the subject indexing process;
- analyse and discuss the various problems associated with subject indexing in order to arrive at appropriate solutions;
- explain the reasons for evaluation of an indexing system; and
- discuss the major criteria for evaluation of the indexing systems.

9.1 INTRODUCTION

All library work is a matter of storage and retrieval of information, and cataloguing and indexing are specially performed to achieve that. Subject cataloguing is intended to embrace only that activity which provides a verbal subject approach to materials added to library collections. Subject indexing is used in information retrieval especially to create index records to retrieve documents on a particular subject. Descriptive cataloguing makes it possible to retrieve the materials in a library by title, author, etc. — in short all the searchable elements of a cataloguing record except the subjects.

Until the second half of the nineteenth century, descriptive cataloguing was the basic library cataloguing practice that was found necessary. Libraries were much smaller than they are today, and scholarly librarians then were able, with the aid of printed bibliographies, to be familiar with everything available on a given subject and guide the users to it. With the rapid growth of knowledge in many fields during the nineteenth century and the resulting increase in the volume of books and periodicals, it became desirable to do a preliminary subject analysis of such works and then represent them in the catalogue or in printed indexes in such a way that they could be retrievable by subject. Subject cataloguing deals with what a book or other library item is about, and its purpose is to list, under one uniform word or phrase all the materials on a given topic that a library has in its collection. A subject heading is a uniform word or phrase used in the library catalogue to express a topic. The use of authorised words or phrases only, with cross-references from unauthorised synonyms, is the essence of bibliographic control in subject cataloguing. In the literature of LIS, the phrases *subject cataloguing* and *subject indexing* are used more or less interchangeably. In this context it is to be pointed out here that it was Charles Ammi Cutter who first gave a generalised set of rules for subject indexing in his Rules for a Dictionary Catalogue (RDC) published in 1876. But he never used the term 'indexing'; rather he used the term 'cataloguing'. In this course material, the phrase *subject indexing* includes *subject cataloguing* also. The literature differentiates the two as subject cataloguing is intended to embrace only that cataloguing activity which provides a verbal subject approach to library collections, especially macro documents (i.e. books). It refers to the determination and assignment of suitable entries for use in the subject component of a library's catalogue. The primary purpose of the subject catalogue is to show which books on a specific subject are possessed by the library. Subject indexing refers to that indexing activity which provides a verbal subject approach to micro documents (e.g., journal articles, research reports, patent literature, etc.). Subject indexing provides a subject entry for every topic associated with the content of a micro document.

The representation of documents and the knowledge expressed by them is one of the central and unique areas of study within Library and Information Science (LIS) and is commonly referred to as indexing. Subject approach to information has been a long and extensive concern of librarianship and is assumed to be the major approach (access method) of users for a very long period. Indexing has traditionally been one of the most important research topics in information science. Indexes facilitate retrieval of information in both traditional manual systems and newer computerised systems. Without proper indexing and indexes, search and retrieval are virtually impossible.

9.2 SUBJECT INDEXING: ORIGIN AND DEVELOPMENT

The origin and development of subject indexing is intimately related with the historical development of libraries through ancient and medieval periods to modern days. The libraries of the ancient world used to arrange documents under some subjects. The catalogue, which worked as an index to this store, was predominantly a systematic subject listing according to a scheme of subject headings. The arrangement more or less conformed to the arrangement of documents in the store.

The specific usage of the term *index* goes back to ancient Rome. There, when used in relation to literary works, the term *index* was used for the little slip attached to papyrus scrolls on which the title of the work (and sometimes also the name of the author) was written so that each scroll on the shelves could be easily identified without having to pull them out for inspection. From this developed the usage of *index* for the title of books. In the first century A.D., the meaning of the word was extended from "title" to a table of contents or a list of chapters (sometimes with a brief abstract of their contents) and hence to a bibliographical list or catalogue. Only the invention of printing around 1450 made it possible to produce identical copies of books in large numbers, so that soon afterwards the first indexes began to be compiled, especially those to books of reference. By the end of 13th century alphabetisation by names of authors under systematic subject arrangement was well known. The index to the store or the shelf-list used to be supplemented with an author index to satisfy the author approach of the users of store. Index entries were not always alphabetised by considering every letter in a word from beginning to end. Most early indexes were arranged only by the first letter of the first word, the rest being left in no particular order at all. Gradually, alphabetisation advanced to an arrangement by the first syllable, that is, the first two or three letters, the rest of an entry still being left unordered.

The 15th century saw the entry of the university libraries which brought about a qualitative change. Efforts were made to rank subjects and on devising indexing or cataloguing methods for better utilisation of documents. Towards the end of 15th century the practice of supplementing systematic listing with an alphabetical subject index was introduced. Only very few indexes compiled in the 16th and early 17th centuries had fully alphabetised entries, but by the 18th century full alphabetisation became the rule. Alphabetical indexing gained a new momentum as intellectual debates among the scholars required ready reference to scholarly works with the rise of universities. The pressmarks, which were mainly used for storage of documents, started being used in catalogues as a retrieval tool. But the pressmarks could not ensure a flexible hierarchical order of subjects and hence it was discarded in favour of notation. In the 19th century, subject access to books was provided by means of classification. Books were arranged by subject and their surrogates were correspondingly arranged in a classified catalogue. Only in the late 19th century, alphabetical subject indexing became widespread and more systematic. Classification system was primitive in nature. It could not go deep enough to the extent of individualising subjects of documents. The separate existence of the classed catalogues and indexes stirred up imagination for compilation of a catalogue which was very much akin to a dictionary in form. Thus were born the forerunners of our dictionary and classified catalogues.

Preparation of back-of-the-book index, historically, may be regarded as the father of all indexing techniques. Indexing techniques actually originated from this index. It was of two types: Specific index, which shows broad topic on the form of one-idea-one-entry, i.e. specific context of a specific idea; and Relative index, which shows various aspects of an idea and its relationship with other ideas. Specific index cannot show this, it only shows broad topic on the form of one-idea-one-entry, i.e. specific context of a specific idea.

The dictionary catalogue brought some relief into the sharp conflict between subjects of documents and the practice of naming them. Charles Ammie Cutter, who was both a classificationist as well as a theoretician of the library catalogue, observed that the name of the subject assigned to a document did not indicate its specific subject. Rather it indicated the class to which the subject of the document belonged to. For example assigning the subject 'plant' to a document discussing the plant 'cactus'. The practice was deficient in helping a user who came for information on a specific subject. The root of the conflict remained deep in the classification system also as the classification was not coextensive with the subjects of the documents. Hence, whatever was left out in classification became conspicuous by their absence while giving class names to individual entries as subject heading. Cutter, who was an advocate of dictionary catalogue wanted to solve the conflict at the cataloguing level. The year 1876 is particularly important for the library profession for the publication of two outstanding books: (1) *A Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library,* by Melvil Dewey; and (2) *Rules for a Dictionary Catalogue,* by Charles Ammie Cutter. The first sought to solve the problems by organising the document store and simultaneously providing an alphabetical subject index for easy access to it while the second, expressing doubts about the efficacy of class headings to be used as specific subject heading, decided to go through a different way by some specific method for naming of the subjects. While Dewey offers a ready made list of names (class names in this case) Cutter suggested some methods for building them up in order to name them more specifically. Cutter's rules for specific subject headings for use in a dictionary catalogue seemed to have appealed the library professional. Subsequently there was a demand for some 'standard list of subject headings' which could be used in carrying out the specifications in Cutter's rules. This paved the way for the publication of a list of subject headings by the American Library Association (ALA), to be used in a dictionary catalogue. The list was later revised and published in two more editions which ultimately established a pattern for subsequent subject heading lists like subject headings used in the dictionary catalogue of Library of Congress and Sears List of Subject Headings.

Use of the above standard lists of subject headings raised important questions relating to the use of terminology (whether common or popular terms or scientific and technical terms were to be used), and sequencing of terms in the subject heading (what should be the sequence of terms in case of compound subject headings). But Cutter as well as compilers of several standard lists of subject headings failed to provide satisfactory answers to the above noted questions.

The first quest for logical approach towards solving the above noted problems is evident in J. Kaiser's *Systematic Indexing* (1911). Kaiser was the first person who gave the idea of categorising the terms under two fundamental categories: concretes and processes. He recommended the citation order of these categories into index string. Kaiser suggested that many composite subjects could be analysed into a combination of concepts indicating a 'concrete' object and a 'process'. In such cases the concrete should be given precedence over 'process' in the order of citation of index terms in a compound subject heading. Kaiser failed to analyse deeply the various types of intricacies involved in naming of subjects. Nevertheless, his work remains unique till date as he is the first person to suggest certain logical processes for naming subjects in terms of fundamental categories and a citation order of index strings.

Dr. S. R. Ranganathan was the first to analyse the universe of subjects in depth and suggesting a complete theory of naming subjects using a subject indexing language. He realised the fallacy of trying to symbolise the extremely flexible and dynamic multidimensional universe of subjects into a linear, rigid notational model. Just as ready-made class numbers cannot be given according to his scheme of classification to all subjects of the past, present and future, so also subject headings cannot be made available ready-made. He therefore enunciated certain rules on the basis of which subject names could be framed. Ranganthan developed a mechanical procedure for doing it and called it the chain procedure. The basic contention of chain procedure is that a multidimensional universe of subjects cannot be fitted into a rigid one-dimensional model and hence, a chain of terms is required to name a subject where the term indicating the specific subject is stated in a particular context. Chain procedure demonstrated that it is not necessary to depend on the flair of some authorities for supply of names of subjects. One can very well build up one's own authority file and use subject names consistently. The names used will be uniform for all libraries following the same scheme of classification. The chain, which is a string of terms, gets organised or arranged following the classification scheme used. Qualities of the classification scheme therefore very much determine the qualities of the subject headings drawn according to chain procedure.

J. E. L. Farradane devised a scheme of pre-coordinate indexing system known as Relational Indexing in the early 1950s. The basic proposition of Farradane's Relational Indexing was to identify the relationship between concepts by following the learning process through which we develop our power of discrimination in time and space. Farradane's Relational Indexing has been the subject of scholarly research, but was never implemented. Still we can say that Farradane's contribution to the area of subject indexing was: analysis of relationship among each pair of terms, use of relational operators, and representation of relationship among terms by relational operators leading to the creation of 'Analets'. 'Analet' refers to a pair of terms linked by any of the relational operators as developed by Farradane. Each relational operator is denoted by a slash and a special symbol having a unique memory. For example,

Chemistry / Θ Encyclopaedia

The contribution of E. J. Coates in subject indexing was not original in nature. Coates merely synthesized the ideas of Cutter, Kaiser, Ranganathan and Farradane. Coates applied his idea on British Technology Index (now Current Technology Index) of which he was editor from its inception in 1963 until his retirement in 1976.

Preserved Context Index System (PRECIS), developed by Derek Austin and applied to BNB in 1971 as an alternative to the chain procedure for deriving subject index entries, sought to rectify the problem of co-extensiveness by generating entries with a lead term and the full context of the document. Depending heavily on computer to generate mechanically all index entries from input strings, PRECIS developed its own code for preparation of input strings by the human indexer and its subsequent processing by computer. Its emphasis has been on generating printed index for BNB. Though PRECIS was fairly successful in its original mission it does not have the simplicity of chain procedure and considerable skill is required to use it effectively.

Postulate-based Permuted Subject Indexing (POPSI) sought to overcome the shortcomings of chain procedure from entirely different perspective. It recommended postulates and principles for analysing the subjects into elementary categories and their subsequent ordering. The postulates are not rigid and hence give flexibility to indexers. As it is essentially distilled out of chain procedure it has managed to retain most of the helpful features of chain procedure such as simplicity. Over the years, Bhattacharya, Neelameghan, Devadasan, Gopinath and others have given a sound theoretical foundation to POPSI in terms of 'General Theory of Subject Indexing Languages' (GT-SIL). The GT-SIL seeks to analyse the deep structure of Subject Indexing Languages in terms of semantic structure, elementary structure and syntactic structure of subject propositions. In essence, GT-SIL is a logical abstraction of the structures of outstanding subject indexing languages such as those of Cutter, Dewey, Kaiser and Ranganathan.

It is evident from the above discussion that the research on development and use of various subject indexing systems was devoted to techniques of constructing pre-coordinate subject headings. A greater part of the pre-coordinate subject indexing system was devoted to syntactical rules of indexing. Rigidity of significance order may not meet the approaches of all users of the index file, though this problem is solved by rotating terms or multiple entry system. It is also evident that even the acceptance of multiple entry system covers only a fraction of the possible number of total permutations. Thus, a large portion of probable approaches or access points are left uncovered. This gap widens rapidly with every increase in the number of terms in a subject heading due to the demand for more specific subject headings. The index file may fail to provide a particular combination which the user is looking for. It may also provide a combination which proves too broad for a particular search. The above considerations and difficulties stemming from the pre-coordination of terms led to the development of post-coordinate indexing or simply coordinate indexing systems like Uniterm, Optical Coincidence Card / Peek-a-boo, Edge-Notched Card, etc during 1960s.

Computers began to be used to aid information retrieval in the 1950s. The Central Intelligence Agency (CIA) of USA is said to be the first organisation to use the machine-produced keywords from Title Index since 1952. H P Luhn and his associates produced and distributed copies of machine produced permuted title indexes in the International Conference of Scientific Information held at Washington in 1958, which he named as Keyword-In-Context (KWIC) index and reported the method of generation of KWIC index in a paper. American Chemical Society established the value of KWIC after its adoption in 1961 for its publication 'Chemical Titles'. A number of varieties of keyword index are evident in the literature. They differ only in terms of their formats but indexing techniques remain more or less the same.

The publication of Science Citation Index (SCI) by Eugene Garfield of the Institute of Scientific Information (ISI), Philadelphia in 1963 provided a new approach to the bibliographic file organisation. The online version of the SCI, known as SCISEARCH, was published in 1974. ISI also brought out the Social Science Citation Index (SSCI) and Arts and Humanities Citation Index (A&HCI) in 1973 and 1978 respectively. The publication of the citation classics, with the first issue of *Current Contents* in 1977, forms an important and interesting venture of the ISI.

It has already been mentioned above that the traditional subject indexing systems and techniques have taken a new turn with the applications of computers in 1950s. In fact, all attempts at computerised indexing were based on two basic methods: Statistical analysis; and Syntactic and semantic analysis. In the arena of computerised indexing, there has been considerable research on the user-interface design, indexing systems using Artificial Intelligence techniques like Natural Language Processing (NLP), Knowledge Representation Model and Expert System-based subject indexing systems. As a result of the phenomenal growth of content on the web as an indexing problem, we have seen a continued interest in the development of tools and techniques to index the Web resources. Different search tools and technologies were developed in finding the resources on the Web so far to make computers understand the semantics underlying contents of the web resources.

9.3 MEANING AND PURPOSE

The term *index* came from the Latin word *indicare* which means 'to point out, to guide, to direct, to locate'. An index indicates or refers to the location of an object or idea. It is a methodically arranged list of items or concepts along with their addresses. The process of preparing an index is known as *indexing*. According to the British Standards (BS 3700: 1964), an index is "a systematic guide to the text of any reading matter or to the contents of other collected documentary material, comprising a series of entries, with headings arranged in alphabetical or other chosen order and with references to show where each item indexed is located". An index is, thus, a working tool designed to help the user to find his way out of a mass of documented information in a given subject field, or document store. It gives subject access to documents irrespective their physical forms like books, periodical articles, newspapers, audio-visual documents, and computer-readable records including web resources.

It appears from the foregoing discussion that an index indicates or refers to the location of an object/idea/concept. A concept is a unit of thought. The semantic content of a concept can be re-expressed by a combination of other and different concepts, which may vary from one language or culture to another. What the particular body of information is about, in a document constitute its subject. A subject can be defined as any concept or combination of concepts representing a theme in a document. An indexing term is defined as the representation of a concept in the form of either a term derived from natural language or a classification symbol.

A subject is then any concept or combination of concepts which is expressed in the document. The readers' task is to interpret the words and sentences in the document in order to understand the concepts. Whether a reader understands a document depends on how precisely the author expresses the concepts he refers to and whether the reader is aware of the concepts the author expresses. The basic idea is that the concepts exist before the author writes the document and the reader reads the document.

Similarly, the indexer's task is to identify concepts in the document and re-express these in indexing terms. This is done first by establishing the subject content, or in other words the content of concepts in the document. Thereafter the principal concept presented in the subject content is identified, and finally, the concepts are expressed in the indexing language. The indexing is successful when the document and the indexing term express the same concepts.

- **Purpose**

Modern subject indexing practice has its roots in Charles Ammi Cutter's *Rules for a Dictionary Catalog* published in 1876. In chapter 1 Cutter's statement of the basic objectives of a catalogue is: (i) *To enable a person to find a book of which the subject is known, and (ii) To show what the library has on a given subject (and related subjects)*. This implies that the main purpose of subject indexing is to satisfy the subject query of the users by enabling an enquirer to identify documents on a given subject and providing information on the presence of material on allied or related subjects.

The first objective refers to the need to locate individual items, and the second refers to the need to collocate materials on the same subject as well as related subjects. A subject is a set of interrelated component ideas in which each component idea is related directly or indirectly to other component ideas. A subject of a document is amenable for structuring into subject heading. It is a kind of linear structuring of subject surrogates, and some criteria for formatting or modeling it into an accessible procedure. The purpose of subject indexing is to:

a) satisfy the subject approach to information;

b) identify pertinent materials on a given subject or topic;

c) enable the enquirer to find materials on related subjects;

d) link related subjects by a network of references;

e) prescribe a standard methodology to subject cataloguers/indexers for constructing uniform subject headings;

f) bring consistency in the choice and rendering of subject entries, using standard vocabulary and according to the given rules and procedures;

g) be helpful to users in accessing any desired document(s) from the catalogue or index through different means of such approach;

h) decide on the optimum number of subject entries, and thus economize the bulk and cost of indexing; and

i) provide user-oriented approach in naming the subjects through any vocabulary common to a considerable group of users, specialists or laymen.

Self Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

1) What do you mean by 'Subject Index' and 'Subject Indexing'?

....

....

....

....

2) State the purposes of subject indexing.

....

....

....

....

3) Traces the historical development of subject indexing.

....

....

....

....

9.4 CATALOGUING VERSUS INDEXING

There has always been confusion about the distinction between 'subject cataloguing' and 'subject indexing'. Basically, cataloguing is the process of creating bibliographic description of a document as a whole entity, and subject cataloguing and classification assign subject labels which together describe the overall topic of the document. Indexing involves delving into a document for analysing its contents at a much deeper level to provide access to many of the concepts contained within it at greater depth. Although most articles in a periodical issue and many books are listed under only one or two subject headings, a back-of-the-book index contains hundreds of subject terms associated with the content of an individual book.

Subject cataloguing usually refers to the assignment of subject headings to represent the overall contents of whole documents (e.g. books, reports, periodicals, etc.) within the catalogue of a library. *Subject indexing* is a term used more loosely; it may refer to the representation of the subject matter of the parts of whole documents as in the case of a back-of-the-book subject index. Thus, a library may enter a book under the subject heading *'roses'* in its catalogue to indicate its overall subject matter, the detailed contents of the book are only revealed by the back-of-the-book subject index. This distinction between the terms *'subject cataloguing'* and *'subject indexing'*, one referring to complete bibliographic items and other to parts of them, is artificial, misleading, and inconsistent. The process by which the subject matter of documents is represented in databases—printed or electronic form—is almost referred to as *'subject indexing'*, whether overall documents or their parts are discussed. Thus, the subject index might refer to the representation of the content of the complete books or complete technical reports as well as to the parts of documents (e.g. chapters in books, papers within the periodicals or conference proceedings, etc.). On the other hand, libraries may choose to represent parts of books (e.g. chapters or papers) within the catalogue which is usually referred to as *analytical cataloguing.*

The situation is even more confusing when the term *classification* is considered. The term *classification* refers to the process of assigning class numbers, drawn from a given classification scheme, to documents, especially for the purpose of arranging these items on the shelves of the libraries, in catalogues, etc. But the subject catalogue of a library can be either alphabetically based (in an alphabetical subject catalogue or dictionary catalogue) or arranged according to the sequence of a classification scheme (in a classified catalogue). Suppose a librarian picks up a book and decides that it is about 'banking'. He or she might assign the subject heading *Banking* to this document. Alternatively, the Dewey Decimal classification number 332.1 may be assigned to it. Many people would refer to the first operation as *subject cataloguing* and to the second as *classification.* These terminological distinctions are quite meaningless and only serve to create confusion due to failure to understand the distinction between the *conceptual analysis* and *translation* stage in indexing. In short, *subject indexing* is conceptually identical to *subject cataloguing*. Its process involves classification, forming classes of objects on the basis of their subject matter and representing them either in the verbal plane (by using a readymade list of subject headings or a thesaurus) or in the notational plane (by using a scheme of classification). In this Unit, the term *subject indexing* or simply *indexing* is used as a matter of convenience to refer to all activities of subject cataloguing.

9.5 INDEXING PRINCIPLES AND PROCESS

9.5.1 Need and Purpose of Indexing Principles

Before we discuss the principles of indexing, it is important to know why we need to have principles of indexing. We need principles of indexing:

1) To set out the general directions for the consistent application of subject indexing techniques;

2) To serve as a useful guide for developing new indexing techniques and to develop one that already exists;

3) To facilitate the evaluation of indexing systems;

4) To provide theoretical rationale for particular standards or guidelines for designing subject indexing system and its application;

5) To promote understanding of different subject indexing systems by identifying commonalities underlying them and providing a structure for their comparison; and

6) To determine how the subject headings are established and applied.

9.5.2 Indexing Principles

Indexing principles may be stated as:

a) **The user as focus:** The wording and structure of the subject heading should match what the user will seek in the index;

b) **Unity:** A subject index must bring together, under one heading all the documents which deal principally or exclusively with the subject, whatever the terms, applied to it by the authors and whatever the varying terms, applied to it at different times. It must use a term which is unambiguous and does not overlap in meaning with other headings in the index.

c) **Common Usage**: The subject heading chosen must represent common usage or, at any rate, the usage of the class of users for whom the documents on the subject within which the heading falls is intended. Whether a popular term or a scientific one is to be chosen should depend on the approaches of the users.

d) **Specificity:** The heading should be as specific as the topic it is intended to cover. As a corollary, the heading should not be broader than the topic. Rather than using a broader heading, the cataloguer should use two specific headings which will approximately cover it.

9.5.3 Indexing Policy

Indexers must take policy decisions about how many terms should be included in an index entry, how specific the terms should be and how many entries an index should incorporate. Together this gives a depth of indexing. The depth of indexing describes the thoroughness of the indexing process with reference to exhaustivity and specificity. While taking such a policy decision, indexers should strive for a balance between specificity and exhaustivity and should consider the requirement of the users of the index along with the cost and time factors.

Exhaustivity in Indexing

Exhaustivity in indexing is the detail with which the topics or features of a document are analysed and described. In other words, an exhaustive index is one which lists all possible index terms associated with the thought content of a document. In contrast to higher exhaustivity, higher specificity increases precision at the cost of impaired recall. Greater exhaustivity gives a higher recall leading to the retrieval of all the relevant documents along with the retrieval of a large number of irrelevant documents or documents which only deal with the subject in little depth.

Specificity in Indexing

The specificity describes how closely the index terms match the topics they represent in a document. It is the extent to which the indexing system permits us to be precise when specifying the subject of a document we are processing. Higher specificity leads to high precision, whereas lower specificity will lead to low precision, but high recall. Specific indexing provides specific terms for all or most topics and features and results in a larger indexing vocabulary than more generic indexing. Specificity tends to increase with exhaustivity in indexing vocabulary as the more terms we include, the narrower those terms will be. A high level of specificity increases precision.

9.5.4 Indexing Process

The representation of documents and the knowledge expressed by them is one of the central and unique areas of study within library and information science (LIS) and is commonly referred to as indexing. A common demand in the LIS field is for a set of rules or a prescription for *how to* index. When this demand is raised it is usually based on the assumption that it is possible to explain the intellectual operations in the subject indexing process. The indexing process basically consists of two intellectual steps: conceptual analysis and translation.

Conceptual analysis

This step refers to the identification of different component ideas associated with the thought content of the document and establishment of interrelationship between those component ideas. According to Ranganathan, it involves the work in the idea plane which is carried out in two stages, although these tend to overlap in practice:

a) examining the document and establishing its subject content;

b) identifying the principal concepts present in the subject;

- **Examining the document and establishing its subject content**

In the first stage of the conceptual analysis of the thought content of the document, it is examined for the establishment of its subject content. A complete reading of the document often is impracticable, but the indexer should ensure that no useful information has been overlooked. While examining the document, the indexer should give particular attention to a number of places in the document: the title; the abstract, if provided; the list of contents; the introduction, the opening chapters and paragraphs, and the conclusion; illustrations, diagrams, tables and their captions; words or groups of words which are underlined or printed in an unusual typeface.

- **Identifying the principal concepts present in the subject**

In this stage of the indexing process the indexer identifies the principal concepts in the subject. The second stage is laid over the first stage in the sense that the indexer should not go back to the document to look for concepts. Rather, the indexer should look for concepts within the findings of the first step; that is the natural language representations of the subject content. The indexer does not necessarily need to retain, as indexing elements, all the concepts identified during the examination of the document. After examining the document, the indexer needs to follow a logical approach in selecting those concepts that best express its subject. While selecting the principal concepts of the document the indexer should take into consideration the purpose for which the indexing data will be used. Indexing data may be used for the purpose like preparation of subject headings for the subject catalogue, production of printed alphabetical indexes to different types of information products, and computerised storage of indexing data elements for subsequent retrieval of the documents.

Translation

During the first two stages the indexer has established the subject content of the document and identified the principal concepts in the subject. The indexer is hereafter ready to translate the concepts into the indexing language. This step refers to the expression of principal concepts as identified while analysing the thought content of the document into the language of the indexing system. According to Ranganathan, it involves the work in the verbal plane which calls for the familiarity with different components of the given indexing language: controlled vocabulary, syntax and semantics including their working roles for displaying the indexing data in a subject index.

If the concepts that the indexer has identified during the second stage are present in the indexing language the indexer should translate the concept into preferred terms. At this point in the indexing process the indexer should be aware that indexing languages may impose certain constraints in translating the concepts. If the indexer uses a controlled indexing language, this may not permit the exact representation of a concept encountered in a document. The concern is that the concepts that the indexer identified during the second stage of the indexing process might not be present in the indexing language. The indexer is then forced either to choose a term that does not express exactly the same concept or add a new term to the vocabulary to represent the concept. Here, the indexer is required to be familiar with the particular indexing language and the specific rules and mechanisms of the indexing language.

9.5.5 Indexing Language

An indexing language is a set of terms and devices used to establish the relationship between terms for representing the content of the documents as well as queries of the users. It consists of three basic elements: controlled vocabulary, syntax and semantics. Controlled vocabulary has been defined as a limited set of terms showing their relationships and indicating ways in which they may usefully be combined to provide subject index to the documents and to search for these documents, in a particular system. Syntax comprises a grammatical structure or a set of rules that govern the sequence of occurrence of terms/words in representing the content of the document. Semantics refers to the systematic study of how meaning is structured, expressed and understood in the use of an indexing language. More discussion on indexing languages can be seen in the Unit 10 of this Course.

9.5.6 Problems in Indexing

An indexer analyses a text and strives to ascertain meaning. Ideally this analysis anticipates a searcher at some future time, looking for text with the same meaning. But, meaning is not fixed at either end of this process. And even if the meaning is relatively unambiguous or stable, the terms used to represent it are not. Thus, most indexing processes encounter problems at two levels**:**

- Interpreting meaning as intended by the author and as construed by the potential user;
- Choosing the terms to represent that meaning that will enable this communication to be clear and as true as it can be. (Bearing in mind that such fidelity is a relative thing to begin with)

Fidelity in the context of IR denotes the accuracy with which term(s) used to represent the name of the subject represent the meaning. A number of problems and issues associated with indexing are:

a) Subjects of documents are complex—usually multi-worded terms;

b) Users' request for information tend to multidimensional;

c) Choice of terms—among different categories, viz. entities, activities, abstracts, properties and heterogeneous concepts (synonymous to semantic factoring);

d) Choice of word forms—among different forms, viz. noun vs. adjective, singular vs. plural;

e) Homographs—if neglected, will give rise to reduced relevance. Seriousness of the problem will depend on the coverage of the system.

f) Choice of the kind of vocabulary that should be used, and syntactical and other rules necessary for representing complex subjects;

g) Identification of term relationship—semantic vs. syntactic;

h) Decision about the exhaustivity level (i.e. the depth to which indexing should be done);

i) Decision about the specificity level (i.e. The levels of generality and specificity at which concepts should be represented);

j) Ensuring inter indexer consistency (i.e. consistency in indexing between several indexers), and intra-indexer consistency (i.e. consistency in indexing by the same indexer at different times); and

k) Ensuring that indexing is done not merely on the basis of a document's intrinsic subject content but also according to the type of users who may be expected to benefit from it and the types of requests for which the document is likely to be regarded as useful.

9.5.7 Quality in Indexing

The quality of an index is defined in terms of its retrieval effectiveness—the ability to retrieve what is wanted and to avoid what is not. Quality in indexing leads to a better performance in retrieving documents. The governing idea is that indexing should be neutral, objective, and independent of the particular indexer's subjective judgment. An indexing failure on the part of the indexer may take place at the following stages of indexing process:

- Failure in establishing concepts during conceptual analysis of the content of a document;

- Failure to identify a topic that is of potential interest to the target user group;
- Misinterpretation of the content of the document, leading to the selection of inappropriate term(s);
- Failure in translating the result of conceptual analysis into the indexing language;
- Failure to use the most specific term(s) to represent the subject of the document;
- Use of inappropriate term(s) for the subject of a document because of the lack of subject knowledge or due to lack of seriousness on the part of the indexer; and
- Omission of important term(s).

The quality of indexing depends on two factors: (i) qualification and expertise of the indexer; and (ii) quality of the indexing tools. In order to achieve quality in indexing, the indexer should have adequate knowledge of the field covered by the documents s/he is indexing. S/he should understand the term of the documents as well as the rules and procedures of the specific indexing system. Quality control would be achieved more effectively if the indexers have contact with users. An indexer who has contact with the users might better be able to represent the documents in accordance with how the users think. The idea is that the indexer should attempt to determine *the* subject of the document taking into account the users' questions and information needs. This might help the indexer when a document contains multiple concepts. In such a situation, the indexer can select only those concepts to represent the content of a document which are regarded as most relevant by a given community of users. Indexing quality can be tested by analysis of retrieval results, e.g. by calculating recall and precision ratios.

Indexing Consistency

It is assumed that there is a relationship between indexing consistency and the indexing quality. That is to say, an increase in consistency can be expected to cause an improvement in indexing quality. Traditionally, consistency in indexing has long been considered as an acceptable indicator of indexing quality. Consistency in indexing is essential for effective retrieval. Indexing consistency refers to "the extent to which agreement exists on the terms to be used to index some document" (Lancaster, 2003). Consistency is a measure that relates to the work of two or more indexers. It should, remain relatively stable throughout the life of a particular indexing system. Consistency is particularly important if information is to be exchanged between agencies in a documentary network. An important factor in reaching the level of consistency is complete impartiality in the indexes. The goal of the consistency is to promote standard practice in indexing.

It has for long been observed that different indexers tend to assign different index terms to the same document as they differ considerably in their judgment as to which terms reflect the contents of the document most adequately. Essentially, indexing consistency is seen as a measure of the similarity of reaction of different human beings processing the same information. Indexing consistency in a group of indexers is defined as the degree of agreement in the representation of the essential information content of the document by certain sets of indexing terms selected individually and independently by each of the indexers in the group.

In the process of indexing, indexers choose what topics to represent and what to call those topics. The goal is to select and name topics consistently so that all of the material about any given topic will be found together. Ideally, if two indexers use the same thesaurus or classification system to index the same document, they are supposed to assign the same index terms or class numbers. In practice, indexers are not always consistent with each other, because subject indexing is essentially a subjective process. Indexers may miss important points of the document, and add irrelevant terms. This would stem from insufficient knowledge of indexers about the subject. Decades of research on consistency between indexers and by the same indexer at different times has documented medium to high levels of inconsistency.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

4) Explain the steps involved in subject indexing.

....

....

....

....

5) What are the different levels of indexing principles?

....

....

....

....

6) Distinguish between Exhaustivity and Specificity in Indexing.

....

....

....

....

7) Identify the different stages in which indexing failure can take place.

....

....

....

....

9.6 EVALUATION OF INDEXING

An indexing system is a sub-system of an information retrieval system and hence, its performance is directly linked up with the overall performance of the entire information retrieval system.

Evaluation of an indexing system essentially means measuring the performance of the system, success or failure, in terms of its retrieval efficiency (ease of approach, speed and accuracy) to the users, and its internal operating efficiency, cost effectiveness and cost benefit to the managers of the system.

The foundation of the Institute of Information Scientists in the UK in 1958 coincides closely with the beginning of the notion of experimental evaluation of information retrieval systems in general and indexing system in particulars. Although there had been some earlier attempts, we usually mark the start of the tradition as the Cranfield experiments, which ran from 1958 to 1966.

9.6.1 Purpose of Evaluation

- To identify the level of performance of the given indexing system,
- To understand how well the given indexing system fulfills the queries of the users in retrieving the relevant documents,
- To compare the performance of two or more indexing systems against a standard,
- To identify the possible sources of failure of the given indexing system or inefficiency with a view to raising the level of performance at some future date,
- To justify the existence of the given indexing system by analysing its costs and benefits,
- To establish a foundation of further research on the reasons for the relative success of alternative techniques, and
- To improve the means employed for attaining objectives or to redefine goals in view of research findings.

9.6.2 Efficiency and Effectiveness of Indexing System

By effectiveness we mean the level up to which the given indexing system attains its stated objectives. The effectiveness may be a measure of how far an information retrieval system can retrieve relevant information withholding non-relevant information. The effectiveness of an indexing system can be measured by calculation of recall and precision ratios. By efficiency we mean how economically the indexing system is achieving its stated objectives. Efficiency can be measured by such factors, such as, at what minimum cost and effort does the system function effectively. It may be necessary that the cost factors are to be calculated indirectly, such as response time (i.e. that is time taken by the system to retrieve the information), user effort (i.e. the amount of time and effort required by a user to interact with the indexing system and analyse the output retrieved in order to get the required information), the cost involved, and so on.

9.6.3 Evaluation Criteria

It is evident from the history of the experimental evaluation of information retrieval systems that there has been a remarkably coherent development of a set of criteria for the evaluation of indexing systems. These evaluation criteria generate argument, disagreement and heated dispute, but there remains a relatively stable common core, which has, despite its limitations, served us well over the last 50 years. The most important criteria used for evaluating an indexing system are: Recall and Precision.

Recall and Precision

- **Recall**

Recall refers to the index's ability to let relevant documents through the filter. Recall ratio is a ratio of the relevant documents retrieved to the total number of relevant documents potentially available. It measures the completeness of the output. Hence, the recall performance can be expressed quantitatively by means of a ratio called recall ratio as mentioned below:

$$\text{Recall ratio} = \frac{R}{C} \times 100$$

Where, R = Number of relevant documents retrieved against a search

C = Total number of relevant documents available to that particular request in the collection.

- **Precision**

When the system retrieves items that are relevant to a given query it also retrieves some documents that are not relevant. These non-relevant items affect the success of the system because they must be discarded by the user, which results in wastage of a significant amount of time. The term 'precision' refers to the index's ability to hold back documents not relevant to the user. Precision ratio is a ratio of the relevant documents retrieved to the number of documents retrieved. It measures the preciseness of the output, i.e. how precisely an indexing system functions. If recall is the measure of system's ability to let through wanted items, precision is the measure of the system's ability to hold back unwanted items. The formula for calculation of precision ratio is:

$$\text{Precision Ratio} = \frac{R}{L} \times 100$$

Where, R = Total number of relevant documents retrieved against a search

L = Total number of documents retrieved in that search

The search result against a query is to separate the all documents into two parts: (a) One part is the set of relevant documents, and (b) the other part is the set of irrelevant documents. The following matrix can be used as a common frame of reference for evaluation of indexing system with reference to the calculation of recall and precision ratios:

	←User relevance decision→		
Retrieved	a Hits	b Noise	a + b Total retrieved
Not retrieved	c Misses	d Dodged	c + d Total not retrieved
Total	a + c Total relevant	b + d Total not relevant	a + b + c + d Total collection

From the above matrix, recall and precision ratios can be calculated according to the following manner:

- Recall ratio = [a / $(a+c)$] × 100
- Precision ratio = [a / $(a+b)$] × 100

Where, a = Hit (Retrieval of relevant documents by the system. It adds to precision).

b = Noise (Retrieval of irrelevant documents by the system along with the relevant documents against a search).

c = Misses (The system fails to retrieve the relevant documents that should have been retrieved. It adds to the noise).

d = Dodged (The system correctly rejects to retrieve the documents that are not relevant to the given query).

It needs to be pointed out here that 100% recall and 100% precision are not possible in practice because *recall and precision tend to vary inversely in searching*. When we broaden a search to achieve better recall, precision tends to go down. Conversely, when we restrict the scope of a search by searching more stringently in order to improve the precision, recall tends to deteriorate.

Relevance

In human history, relevance has been around forever, or as long as humans tried to communicate and use information effectively. The concept of "relevance" is the fundamental concept of information science in general and information retrieval, in particular. Evaluation of indexing will never be effective until there is an understanding of the percept of relevance. Relevance is one of the important types of measures used in the evaluation of an information retrieval system and is highly debated issue in information retrieval research. There does not seem to be any consensus among the experts on the definition of *relevance*.

The first full recognition of relevance as an underlying notion came in 1955 with a proposal to use "recall" and "relevance" (later, because of confusion, renamed *precision,* sometimes it was called as *pertinence*) as measures of retrieval effectiveness in which relevance was the underlying criterion for these measures. But, the term *pertinence* refers to a relationship between a document and an information need, whereas the term *relevance* refers to a relationship between a document and a request statement (i.e. expressed information need). It refers to the ability of an information retrieval system to retrieve material that satisfies the needs of the user.

We know that the main objective of indexing, forming an essential component of an IR system, is to determine the aboutness of documents for subsequent retrieval of information object relevant to user queries. Relevance denotes how well a retrieved set of documents meets the information need of the user i.e. to what extent the topic of a retrieved set of information objects matches the topic of the query or information need.

In most of the evaluation studies *relevance* was applied to stated requests (i.e. expressed need). But, it has now been well established that the users' requests do not reflect their information needs completely. Therefore, the current view is that the *relevance* is to be judged in relation to both expressed and unexpressed needs rather than restricting only to stated requests. It is dependent on the degree to which a user is able to recognize the exact nature of his/her information need and the degree to which his/her need is accurately expressed in the form of a request (i.e. request statement). Information retrieval systems create relevance—they take a query, match it to information objects in the system by following some algorithms, and provide what they consider relevant. People derive relevance from obtained information or information objects. They relate and interpret the information or information objects to the problem at hand, their cognitive state, and other factors—in other words, people take the retrieved results and derive what may be relevant to them. Relevance is *derived* by inference.

Although "relevance" is extensively used in evaluation of information retrieval, there are considerable problems associated with reaching an agreement on its definition, meaning, evaluation, and application in information retrieval. There are a number of different views on "relevance" and its use for evaluation. This is because there are degrees of relevance. Relevance is a subjective factor depending on the individual. The same questions, posed by two different enquirers, may well require two different answers. It is because of the fact that enquirers seek information from their own corpus of knowledge. Thus it appears that the *relevance* is highly subjective and personal. It is a relation between an individual with an information need and a document.

Other Important Criteria

Perry and Kent are credited for bringing the concept of evaluation into information retrieval systems during 1950s. The evaluation criteria they suggested were:

i) **Resolution factor:** The proportion of total items retrieved over a total number of items in the collection.

ii) **Pertinency factor**: The proportion of relevant items retrieved over a total number of retrieved items. This factor was popularly named as the precision ratio in the subsequent evaluation studies.

iii) **Recall factor:** The proportion of relevant items retrieved over a total number of relevant items in the collection.

iv) **Elimination factor:** The proportion of non-retrieved items (both relevant and non-relevant) over the total items in the collection.

v) **Noise factor:** The proportion of retrieved items those are not relevant. This factor is considered as the complement of the pertinency factor.

vi) **Omission factor**: The proportion of non-relevant items retrieved over the total number of non-retrieved items in the collection.

Perry and Kent suggested the following formulae for the estimation of the above mentioned evaluation criteria:

L / N	= Resolution factor	(N—L) / N	= Elimination factor
R / L	= Pertinency factor	(L—R) / L	= Noise factor
R / C	= Recall factor	(C—R) / C	= Omission factor

Where, N = Total number of documents

L = Number of retrieved documents

C = Number of relevant documents

R = Number of documents that are both retrieved and relevant

C. W. Cleverdon (1966) identified six criteria for the evaluation of an information retrieval system. These are:

i) ***Recall*:** It refers to the ability of the system to present all the relevant items;

ii) ***Precision*:** It refers to the ability of the system to present only those items that are relevant;

iii) ***Time lag*:** It refers to the time elapsing between the submission of a request by the user and his receipt of the search results.

iv) ***User Effort*:** It refers to the intellectual as well as physical effort required from the user in obtaining answers to the search requests. The effort is measured by the amount of time user spends in conducting the search or negotiating his enquiry with the system. Response time may be good, but user effort may be poor.

v) ***From of presentation*** of the search output, which affects the user's ability to make use of the retrieved items, and

vi) ***Coverage of the collection*:** It refers to the extent to which the system includes relevant matter. It is a measure of the completeness of the collection.

9.6.4 Major Retrieval Experiments

- **Cranfield Test**

The first extensive evaluation of retrieval systems was carried out by the ASLIB with financial assistance from the National Science Foundation at the College of Aeronautics, Cranfield, UK, under the supervision of C. W. Cleverdon in 1957. It is popularly known as the Cranfield 1 Project. Its objective was to investigate into the comparative efficiency of four indexing systems: UDC, Faceted Classification, Alphabetical Subject Heading List and Uniterm Indexing system. Cranfield 1 was the most significant study in the decade of 1958-68. Even then it was not free from criticisms. One of the main criticisms against this study was the artificiality without much relation to the real life situation. The drawbacks of Cranfield 1 necessitated the conduct of further tests. The second stage of Cranfield studies, known as Cranfield 2, began in 1963 and completed in 1966.

- **MEDLARS Test**

The largest evaluation of an operating system was performed by Lancaster on the performance of the Medical Literature Analysis and Retrieval System (MEDLARS) of the US National Library of Medicine in the period between August 1966 and July 1967. The study involved the derivation of performance figures and conduct of details of failure analyses for a sample of 3000 real searches conducted in 1966-67. The objective of the MEDLARS test was to evaluate the existing system and to find out how it could be improved. The document collection available on the MEDLARS service at the time of the test consisted of about 700,000 items. Search requests from twenty-one user groups, selected for the study were taken. On receipt of a request, the system operator formulated it in ***Mesh***(Medical Subject Headings) terms using expansion and search logic.

- **TREC Experiment**

Prior to 1991, research in information retrieval was limited to small test collections, and there was little transfer of research ideas into commercial systems. The Text REtrieval Conference (TREC), launched in 1991, introduced the first large test collection of full-text documents in order to enable IR researchers to scale up from small collection of data to larger experiments, along with the idea that providing an open testing event, with common tasks and a standard evaluation scenario, would lead to the acceleration of research on a realistic scale. It was funded by DARPA (Defence Advanced Research Project Agency, Department of Defence, USA) and operated by the NIST (National Institute for Science and Technology, USA). It is an annual competition/collaboration/ get-together between research groups interested in different aspects of information retrieval. Every year, a set of tasks is defined, broadly information retrieval/search tasks. The TREC series of experiments has drawn attention of the LIS professionals all over the world since its inception and has shown that significant research results can be obtained through international efforts and collaboration.

A wide range of information retrieval strategies was tested in different TREC experiments (i.e. from TREC 1 in 1992 to TREC 16 in 2007). Some notable examples are:

- Boolean search;
- Statistical and probabilistic indexing and term weighting strategies;
- Passage or paragraph retrieval;
- Combining the results of more than one search;
- Retrieval based on prior relevance assessments;
- Natural language-based and statistically based phrase indexing;
- Query expansion and query reduction;
- String and concept-based searching;
- Dictionary-based searching;
- Question-answering;
- Content-based multimedia retrieval; and
- Relevance judgements.

Self Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

8) What do you mean by the evaluation of an indexing system? What are its purposes?

....

....

....

....

9) How do you measure the 'efficiency' and 'effectiveness' of an indexing system?

....

....

....

....

10) How recall and precision ratios help in measuring the performance of an indexing system?

....

....

....

....

11) Distinguish between 'Relevance' and 'Pertinence'.

....

....

....

....

12) What is the significance of TREC?

....

....

....

....

9.7 SUMMARY

In this Unit we have dealt with a comprehensive view of the different issues and perspectives of subject indexing. It begins with a brief overview of the origin and development of subject indexing from the ancient to modern times. Then, meaning and purposes of subject indexing have been discussed. It is impossible to understand the subject indexing properly without being aware of the different principles and processes associated with it. For this, different aspects of indexing principles and processes are discussed. Indexing policy with particular reference to exhaustivity and specificity has been explained. Problems associated with subject indexing are highlighted. The quality of indexing is defined in terms of consistency in indexing. This Unit lays major emphasis on the evaluation of indexing system forming an essential component of the information and retrieval system. Different evaluation criteria like recall, precision and relevance are discussed. Major retrieval experiments are discussed briefly at the end of this Unit.

9.8 ANSWERS TO SELF CHECK EXERCISES

1) Subject indexing is a set of interrelated processes associated with the representation of informational content of the document in the form of a subject index. A subject index is a systematic guide to the text of any reading matter or to the contents of documents, comprising a series of entries, with headings arranged in alphabetical or other chosen order and with references to show where each item indexed is located. It is a working tool designed to help the user to find his required documents in a given subject field. It gives subject access to documents irrespective their physical forms.

2) The purposes of subject indexing are to: satisfy subject approach to information; identify pertinent materials on a given subject or topic; enable the enquirer to find materials on related subjects through the network of references; prescribe a standard methodology for constructing uniform subject headings; bring consistency in the choice and rendering of subject entries, using standard vocabulary and according to the given rules and procedures; enable users in accessing any desired document(s) from the catalogue or index; economize the bulk and cost of indexing by deciding on the optimum number of index entries; and provide user-friendly approach in naming the subjects through any vocabulary common to a considerable group of users—specialists or laymen.

3) The origin and development of subject indexing languages is intimately related with the historical development of libraries through ancient and medieval periods to modern days. The specific usage of the term *index* goes back to ancient Rome when term *index* was used for the little slip attached to papyrus scrolls on which the title and sometimes name of the author of the work was written in order to identify the scrolls on the shelves. In the first century A.D., the meaning of the word was extended from "title" to a table of contents or a list of chapters and hence to a bibliographical list or catalogue. During the 13th century alphabetisation by names of authors under systematic subject arrangement was well known with the invention of printing. The practice of supplementing systematic listing with an alphabetical subject index was introduced with the development of university libraries during 15th century. With the rapid growth of knowledge in many fields in the course of the 19th century and the resulting increase in the volume of books and periodicals, subject indexing got a new momentum. In the 19th century, subject access to books was by means of a classification. But it could not go deep enough to the extent of individualizing subjects of documents. Only in the late 19th century, alphabetical subject indexing became widespread and more systematic. It was Charles Ammi Cutter who first gave a generalized set of rules for subject indexing in his *Rules for a Dictionary Catalogue (RDC)* published in 1876. Cutter's rules for specific subject headings for use in a dictionary catalogue seemed to have made an impact on the development of readymade lists of subject headings like Library of Congress Subject headings and Sears List of Subject Headings in 1897 and 1923 respectively. The first quest for logical approach towards solving the problems of compound subject headings is evident in J. Kaiser's *Systematic Indexing* (1911). Kaiser was the first person who gave the idea of categorization of terms: concrete and process. Dr. S. R. Ranganathan was the first to analyse the universe of subjects in depth and developed chain indexing system to supplement the classified catalogue in 1934. Qualities of the classification scheme determined the qualities of the subject headings drawn according to chain procedure in view of its total dependency on the classification scheme. J. E. L. Farradane proposed a new line of thinking in subject indexing system by suggesting the relationship between pairs of concepts instead of categorisation of concepts in Relational Indexing developed in the early 1950s. The contribution of E. J. Coates in subject indexing (1960) was not original in nature. Coates merely synthesised the ideas of Cutter, Kaiser, Ranganathan and Farradane. Preserved Context Index System (PRECIS), developed by Derek Austin in 1969 and applied in BNB in 1971 as an alternative to the chain procedure is considered as the first computerised pre-coordinate indexing system. Postulate-based Permuted Subject Indexing (POPSI) developed in 1969 sought to rectify the defects of chain procedure from entirely different perspective. It is evident from the above discussion that the research on development and use of various subject indexing systems was devoted to techniques of constructing pre-coordinate subject heading. Difficulties stemming from the question pre-coordination of terms led to the development of post-coordinate indexing or simply coordinate indexing systems like Uniterm, Optical Coincidence Card / Peek-a-boo, Edge-Notched Card, etc during 1960s. Computers began to be used to aid information retrieval system in the 1950s. H P Luhn has been credited for the production of computerized permuted title indexes, which he named as Keyword-In-Context (KWIC) index. The publication of Science Citation Index (SCI) by Eugene Garfield in 1963 provided a new approach to the bibliographic file organization. In the arena of computerized indexing, there has been considerable research on the user-interface design, indexing systems using Artificial Intelligence

techniques like Natural Language Processing (NLP), Knowledge Representation Model and Expert System-based subject indexing systems. As a result of the phenomenal growth of content on the Web as an indexing problem, we have seen a continued interest in the development of tools and techniques to index the Web resources. Different search tools and technologies were developed in finding the resources on the Web so far to make computers understand the semantics underlying contents of the Web resources.

4) The indexing process basically consists of two intellectual steps: Conceptual analysis and Translation. Conceptual analysis involves the identification of different component ideas associated with the thought content of the document and establishment of interrelationship between those component ideas. It is carried out by examining the document for establishing its subject content and identifying the principal concepts present in the subject of the document. Translation refers to the expression of principal concepts as identified while analysing the thought content of the document into the language of the given indexing system.

5) The indexing principles operate at four levels: user as the focus, utility, common usage, and specificity.

6) Exhaustivity is the measure of the extent to which all the distinct subjects are discussed in a particular document are recognised in indexing operation. Exhaustivity in indexing allows for the recognition of concepts embodied not only in the main theme of the document but also in sub-themes of varying importance. Specificity is the degree of preciseness of the subject to express the thought content of the documents. It is the measure of the extent to which the indexing system permits the indexers to be precise when specifying the subject of the document. An indexing language is considered to be of high specificity if minute concepts are represented precisely by it. Both Exhaustivity and Specificity are very closely related to recall and precision. A high level of exhaustivity increases recall and high level of specificity increases precision

7) An indexing failure may take place at the following stages of indexing process: (a) failure in establishing concepts during conceptual analysis of the content of a document; (b) failure to identify a topic that is of potential interest to the target user group; (c) misinterpretation of the content of the document, leading to the selection of inappropriate term(s) (d) failure in translating the result of conceptual analysis into the indexing language; (e) failure to use the most specific term(s) to represent the subject of the document; (f) use of inappropriate term(s) for the subject of a document because of the lack of subject knowledge or due to lack of seriousness on the part of the indexer; and (g) omission of important term(s).

8) By evaluation of an Indexing system we mean the measurement of the performance of the indexing system—success or failure, in terms of its retrieval efficiency (ease of approach, speed and accuracy) to the users, and its internal operating efficiency, cost effectiveness and cost benefit to the managers of the system. The purposes of the evaluation of indexing are to (a) identify the level of performance of the given indexing system, (b) understand how well the given indexing system fulfills the queries of the users in retrieving the relevant documents, (c) compare the performance of two or more indexing systems against a standard, (d) identify the possible sources of failure of the given indexing system or inefficiency with a view to raising the level of performance at some future date, (e) justify the existence of the given indexing system by analysing its costs and benefits, (f) To establish a

foundation of further research on the reasons for the relative success of alternative techniques, and (g) improve the means employed for attaining objectives or to redefine goals in view of research findings.

9) The effectiveness of an indexing system can be measured by calculation of recall and precision ratios. Efficiency can be measured by what minimum cost and effort does the indexing system function effectively. Cost factors are calculated indirectly, such as response time (i.e. that is time taken by the system to retrieve the information), user effort (i.e. the amount of time and effort required by a user to interact with the indexing system and analyse the output retrieved in order to get the required information), the cost involved.

10) Recall refers to the index's ability to let relevant documents through the filter. Recall ratio is a ratio of the relevant documents retrieved to the total number of relevant documents potentially available. It measures the completeness of the output by means of a ratio called recall ratio as mentioned below:

Recall ratio = R / C X 100

Where, R = Total Number of relevant documents retrieved against a search

C = Total number of relevant documents to that particular request in the collection.

When the system retrieves items that are relevant to a given query it also retrieves some documents that are not relevant. These non-relevant items affect the success of the system because they must be discarded by the user, which results the wastage of significant amount of time. The term 'Precision' refers to the index's ability to hold back documents not relevant to the user. Precision ratio is a ratio of the relevant documents retrieved to the number of documents retrieved. The formula for calculation of Precision Ratio is:

Precision Ratio = R / L X 100

Where, R = Total number of relevant documents retrieved against a search

L = Total number of documents retrieved in that search

11) The term *pertinence* refers to a relationship between a document and an information need, whereas the term *relevance* refers to a relationship between a document and a request statement (i.e. expressed information need). *Relevance* is consensus judgment, but *pertinence* relates to an individual judgment.

12) Research in information retrieval was limited to small test collections prior to 1991 and there was little transfer of research ideas into commercial systems. The Text REtrieval Conference (TREC), launched in 1991, introduced the first large test collection of full-text documents in order to enable IR researchers to scale up from small test collection of data to larger experiments on a realistic scale TREC is an open testing event conducted annually with common tasks and a standard evaluation scenario. Every year, a set of information retrieval/search tasks is defined. The TREC series of experiments has drawn attention of the LIS professionals all over the world since its inception and has shown that significant research results can be obtained through international efforts and collaboration. A wide range of information retrieval strategies was tested in different TREC experiments (i.e. from TREC 1 in 1992 to TREC 16 in 2007).

9.9 KEYWORDS

Class number	:	Notation that designates the class to which a given item belongs.
Concept	:	A unit of thought, formed by mentally combining some or all of the characteristics of a concrete or abstract, real or imaginary object. Concepts exist in the mind as abstract entities independent of terms used to express them.
Conceptual Analysis	:	A process which involves deciding what a document is about—that is, identification of different component ideas associated with the thought content of the document and establishment of interrelationship between those component ideas.
Exhaustivity	:	The use of enough terms to cover the all topics discussed in a document. It relates to the breath of coverage in indexing. It is sometimes called *depth indexing*.
Identifier	:	A significant term, number, acronym, name, or symbol used alone or with other identifiers to refer to a library resource.
Index	:	A detailed alphabetical or numerical list. List entries represent an aspect of a bibliographic record and are organised into searchable files used to retrieve records in a database or set of records.
Indexing	:	The process of evaluating information entities and creating indexing terms, normally subject or topical terms that aid in finding and accessing the entity. Index terms may be in natural language or controlled vocabulary or a classification notation.
Indexing Term	:	The representation of a concept in an indexing language, generally in the form of a noun or noun phrase. Terms, subject headings, and heading-subheading combinations are examples of indexing terms. Also called *Descriptor*.
Precision Ratio	:	The term *precision* relates to the ability of an indexing system not to retrieve irrelevant items. *Precision ratio* refers to the proportion of retrieved items that are relevant.
Quality of Indexing	:	The ability to retrieve what is wanted and to avoid what is not wanted.
Recall Ratio	:	The term *recall* refers to a measure of whether or not a particular item is retrieved or the extent to which the retrieval of wanted items occurs.

		Recall ratio is the proportion of relevant items retrieved and thus, it is a measure of the completeness of a search in an index file.
Relative Index	:	A type of index in which various aspects of an idea and its relationship with other ideas are shown.
Relevance	:	*Relevance* refers to the ability of an information retrieval system to retrieve material that satisfies the information needs of the users.
Specific index	:	Specific index is an index which shows the broad topics in the form of one-idea-one-entry, i.e. specific context of a specific idea.
Specificity	:	It refers to the use of much smaller number of terms to cover only the central subject matter of a document. The more specific the terms used, the fewer the entries per term on the average. Specificity is the property of the vocabulary used in indexing and it relates to the depth of treatment of the content of a document in indexing.
Subject	:	A systematised or organised body of ideas.
Subject cataloguing	:	The part of cataloguing that provides subject heading/terms and/or classification.
Subject Headings	:	An alphabetical list of words or phrases that represent a concept that is under authority control, e.g., the Library of Congress Subject Headings.

9.10 REFERENCES AND FURTHER READING

Chakraborty, A. R., and Bhubaneswar Chakraborty. *Indexing: Principles, Processes and Products.* Calcutta: World Press, 1984. Print.

Chowdhury, G. G. *Introduction to Modern Information Retrieval*. 2nd ed. London: Facet Publishing, 2004. Print.

Foskett, A. C. *Subject Approach to Information*. 5th Ed. London: The Library Association, 1996. Print.

Ghosh, S. B. and J. N Satpathi., *Eds*. *Subject Indexing Systems: Concepts, Methods and Techniques*. Calcutta: IASLIC, 1998. Print.

Guha, B. *Documentation and Information: Services, Techniques and Systems*. 2nd rev ed. Calcutta: World Press, 1983. Print.

International Standards Organization. *Documentation—Methods for Examining Documents, Determining their Subjects, and Selecting Indexing Terms,* 1985. Print.

Jones, Karen Sparck, Ed. . *Information Retrieval Experiment.* London: Butterworth, 1981.

Lancaster, F. W. *Indexing and Abstracting in Theory and Practice*. 3rd edition. London: Facet Publishing, 2003. Print.

Sarkhel, Juran Krishna. *Information Analysis in Theory and Practice*. Kolkata: Classique Books, 2001. Print.

Sarkhel, Juran Krishna. "Indexing Systems and Techniques". *Information Processing and Retrieval* (MLII-102); Unit 4, p.89-176:, New Delhi: Indira Gandhi National Open University, 2005. Print.

Sarkhel, Juran Krishna. "Evaluation of Indexing Systems". *Information Processing and Retrieval* (MLII-102); Unit 5, p.177-208: New Delhi: Indira Gandhi National Open University, 2005. Print.

Sarkhel, Juran Krishna (1999). "Subject Indexing, Vocabulary Control and Recent Developments in Cataloguing". *Library Cataloguing Theory*. (BLIS 04)New Delhi, Indira Gandhi National Open University, 1999. 180p. Print.

Wellisch, H. H. *Indexing from A to Z. 2nd. ed*. New York: H. W. Wilson, 1995. Print.

UNIT 10 INDEXING LANGUAGES

Structure

10.0 OBJECTIVES

After reading this Unit, you will be able to:

- explain the value of indexing languages;
- appreciate and discuss the differences between indexing language and the language we use in our day to day life;
- describe the structure and attributes of an indexing language;
- explain the meaning, need and objectives of vocabulary control;
- describe the different types of vocabulary control devices;
- develop skills of using different tools of document contents analysis and using subject heading lists and thesauri;
- explain the provisions of Sears List of Subject Headings and Library of Congress (LC) List of Subject Headings; and
- discuss their applicability in libraries for subject indexing purposes.

10.1 INTRODUCTION

Indexing language (IL) is an artificial language made up of expressions connecting several kernel terms and adopted to the requirements of indexing. The function of an IL is to do whatever a natural language (NL) does and in addition organise the semantic content through a different expression providing a point of access to the seekers of information. An IL is a system for naming subjects and has controlled vocabulary. The vocabulary of an IL may be verbal or coded. A classification scheme uses coded vocabulary in the form of notation and authority lists uses verbal vocabulary. It is a prerequisite to understand the features of the language used for the representation of the subject content of the documents in terms of their linguistics structures and functions for the purpose of studying the structure of indexing language. Thus, there are areas of linguistics which are of common interest to information scientists.

10.2 MEANING AND SCOPE

A language is a code through which messages are transmitted. It is a communication medium based on association of thoughts/ideas. In terms of linguistics, all spoken languages (i.e. natural language) consist of three basic elements: vocabulary, syntax and semantics. Vocabulary is a list of terms/words used in a particular natural language. Syntax comprises a grammatical structure or a set of rules that govern the sequence of occurrence of terms/words in a sentence. Semantics refers to the study of what meaning is and how it operates. It is, in other words, a systematic study of how meaning is structured, expressed and understood in the use of a language. Syntax is used to resolve word meaning through the determination of context.

Information systems are concerned with the communication of information about the documents to the potential users of those documents. The means of communication are the subject indexing language or simply an indexing language. An IL is a system for naming subjects of the records of information (i.e. document). It is an artificial language made up of expressions connecting several kernel terms/notations. The function of an IL is to do whatever a NL does and in addition organise the semantic content through a different expression providing a point of access to the seekers of information. Thesauri, the readymade lists of subject headings and classification schemes are the examples of subject indexing languages.

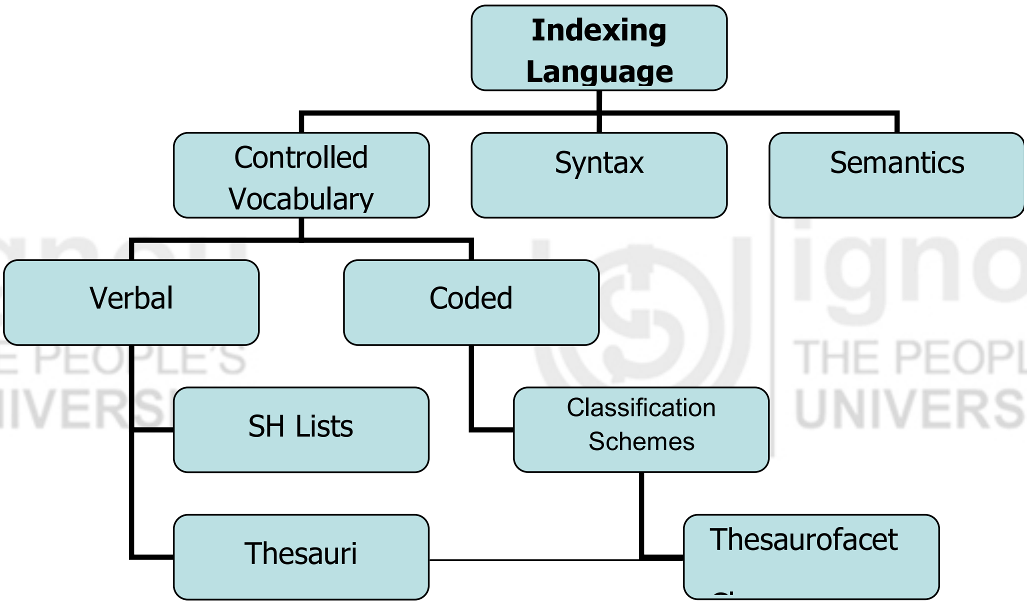
10.3 NATURAL LANGUAGE (NL) VERSUS INDEXING LANGUAGE (IL)

If the terms that appear in the documents are used without required modifications, it is a NL. Since the usage of a NL leads to many problems, such as those arising from the use of different words by different authors to denote the same idea, an alternate to NL is, to use artificial language adapted to the specific needs. The artificial language uses concept indexing rather than term indexing. The terms are representatives of a NL used by authors. The concepts imbibe standard description established in the IL. The NL is flexible and advantageous to authors to use different terms to denote same concept. The indexer, who is more concerned with the ideas conveyed rather than the language niceties, depends upon artificial language. The differences between the natural language and indexing language are furnished below:

Natural Language	Indexing Language
A natural language is a set of codes and their admissible expression used for communication of ideas in speech and writing in our day to day life.	An indexing language is a set of codes and their admissible expression used for representing the content of the documents as well as queries of the users.
A natural language is "natural" in the sense that it grows freely in the lips of human being, totally free from any control whatever.	An indexing language is "artificial" in the sense that it may depend upon the vocabulary of a natural language, though not always, but its syntax, semantics and orthography would be different from the natural language.
A natural language is developed for communication of ideas among human beings in their day to day life.	Indexing language is developed and used for special purpose, i.e. for representation of the thought content of the documents as well as queries of the users.
A natural language is free language and there is no control of synonyms and homographs. One concept may be denoted by more than one term. There is no standardization of terms or words. Anybody can use any words/terms to express her his/ ideas.	An indexing language is a controlled language. There is restriction in using the words/terms in indexing language. Synonyms and homographs are controlled. There is standardisation of terms/words. One concept is denoted by only one term.
Natural language provides auxiliaries like prepositions, conjunctions, etc. to bring out the correct meaning of the sentence.	Such auxiliaries are not available in an indexing language. The order of terms according to the syntactical rules of an indexing language along with the relational symbols like role operators or indicator digits bring the correct meaning of a subject heading.

10.4 STRUCTURE OF INDEXING LANGUAGE

Like natural language, an indexing language consists of three elements: (a) Vocabulary (not free vocabulary, but controlled vocabulary), (b) Syntax, and (c) Semantics. All the structured indexing languages are based upon careful subject analysis. The following figure presents the structure of an indexing language:



10.4.1 Controlled Vocabulary

An indexing language operates with a controlled vocabulary. An IL having controlled vocabulary attempting to indicate the relationship between terms in the index vocabulary is systematically structured. The vocabulary of an IL is either verbal or coded. Subject heading lists and thesauri come within the purview of verbal controlled vocabulary. A classification scheme employs coded vocabulary in the form of its notation. Thus, for example in Colon Classification (CC) Schedule 'Indian History' is rendered as V44. In Sear's List of Subject Headings which employs verbal vocabulary it is rendered as: India - History. There are also controlled vocabularies like Thesaurofacet, Classaurus, etc., which possess both the characteristics verbal as well as coded controlled vocabularies. In any case, selection of terms to be used in each discipline is primary and coding is done at a later stage. The need, objectives, methods of vocabulary control, etc. are discussed in detail under the section 10.6 of this Unit.

10.4.2 Syntax

The etymological meaning of syntax is 'putting things together in an orderly manner'. In the context of an indexing language, syntax refers to a set of rules or grammar which governs the sequence of words in a subject heading, or notations in a classification number.

Most of the subjects treated even in modern macro documents are of compound nature. This means that the name of a subject can no longer be represented by a single word or term. When a number of terms have to be used in representing the subject coextensively, syntax is necessary to put the terms in a most helpful and known searchable order. In other words, we can say that syntax of an indexing language provides pattern of relationship which we recognise between the terms used in the system, i.e. between the terms in the index vocabulary or controlled vocabulary. This recognition is based on a careful subject analysis which is basic to the indexing language.

The order of terms according to the rules of syntax of an indexing language assumes greater importance of presenting correct meaning of a subject heading. Apart from the order of terms prescribed by its rules of syntax, it becomes necessary, at times, to use relational symbols or indicator digits to bring out the correct relations between terms. In this connection, it is to be pointed out here that natural language provides auxiliaries like prepositions, conjunctions, etc. to bring out the correct meaning of the sentence. But in an indexing language such auxiliaries are not available and hence, the correct meaning of a subject heading has to be expressed largely through the order of terms along with the relational symbols like role operators or indicator digits. Syntactical relationship is document dependent relationship.

10.4.3 Semantics

As stated earlier, semantics refers to the systematic study of how meaning is structured, expressed and understood in the use of an indexing language. Various types of semantic relationships are evident in an indexing language. These relationships include equivalence relationships, hierarchical relationships, and associative relationships. Meaning of the term can be derived from its hierarchy. Semantic relationship is document independent relationship. The syntactical rules of an indexing language is also used to resolve the meaning of the term in a subject heading (consisting of string of terms) through the determination of context.

10.5 ATTRIBUTES OF AN INDEXING LANGUAGE

Indexing language is designed for a special purpose. It serves three purposes—representing subject content of documents, organising a searchable file, and representing subject content of the queries of the users while searching the index file. A positive result in searching is achieved only when the content representation of the document by the indexer and that of the queries by the searcher match. This matching of the file is very much dependent on the organisation of the index file in a predetermined order and the awareness of the users of it. Various attributes of an indexing language like vocabulary control, concept coordination, multiple access, syndetic devices, relation manifestation, and structural presentation play very important roles in the successful organisation of the index file and subsequent matching of the index and queries of the users.

10.5.1 Vocabulary Control

The vocabulary of an indexing language is controlled for standardization of terms—i.e. one concept should be denoted by only one term. This is done by controlling synonyms, near-synonyms and word forms, and by distinguishing among homographs. The methods of vocabulary control are discussed in detail under the sub-section 10.6.3 of this Unit.

10.5.2 Concept Coordination

The contents of most of the documents of present days cannot be represented by only one term. Because of the use of multiple terms and multiple relationships among terms to represent the content of the document it has become imperative to make available standard guidelines for coordination of concepts dented by the terms. One of the essential components of an indexing language, i.e. syntax governs the sequence of words in a subject proposition. Natural language provides auxiliaries like prepositions, conjunctions, etc. among the substantive words to bring out the correct meaning of the sentence. But such auxiliaries are not available in an indexing language. The correct meaning of a subject is expressed mainly through the order of terms according to the rules of syntax, sometimes along with the relational symbols like role operators or indicator digits of an indexing language. These rules of syntax will vary from one indexing language to other. Coordination of concepts is carried out by the indexer at the time of indexing (i.e. at input stage) in pre-coordinate indexing and by the searcher at the time searching (i.e. at output stage) in post-coordinate indexing.

10.5.3 Multiple Access

The syntactical rules of the given indexing language help us to determine the order of significance in a linear representation of the subject of a document. It provides only a single access in the searchable index file. Rigidity of the significance order may not meet the approaches of all the users of the index file. In order to satisfy the approaches of all the users, indexing languages have introduced the mechanism for multiple index entries by rotating or cycling of the component terms representing the subject of the document. The rotation is carried out in such a way that each of the component terms gets access position as lead term in the index entries. Each lead term is followed by other component terms in order to maintain the context and correct meaning of the subject proposition. The provision of mechanism for multiple index entries by rotating or cycling of the component terms is a special feature of indexing language. However, it has been observed that even the acceptance of this multiple access mechanism covers only a fraction of the possible number of the total permutations, which in turn, results into the failure of the index file to provide a particular pattern of combinations which the user is looking for. Consequently, a large portion of probable approach points is left uncovered.

10.5.4 Syndetic Device

Syndetic device is an organisational framework in which related subjects are linked together in an underlying classificatory structure.

- **Cross References:** Related and equivalent subjects are linked to each other by a network of references via connecting terms such as *See also* and *See / USE / UF* respectively.
- **Inversion of Headings:** The strict adherence to the natural language order of terms in a subject heading would often lead to headings in which the first word is not the most significant. In such a situation natural language order of terms is inverted in a subject heading.

10.5.5 Relation Manifestations

The range of an indexing language is not simply a matter of vocabulary. Provision for rules of syntax is to be made for the expression of relationships between the terms comprising the vocabulary. These relationships, as conceived by a team lead by J. C. Gardin during the SYNTOL (Syntagmatic Organization Language) programme for the development of a meta language as a common ground between different information retrieval systems in 1960s, are of two kinds: Paradigmatic and Syntagmatic relations.

- **Paradigmatic relationship:** Paradigmatic relations also called semantic or generic relations, usually find expression in the organisation of the vocabulary itself. Thus in classification schedules it is through the successive degrees of subordination that such relations are made explicit. In the readymade lists of subject headings or thesauri the relationships are expressed through the manifestation of hierarchical relationships through the relationship indicators BT and NT. A paradigmatic relationship is document independent relationships because this relationship is established without any reference to a document.

- **Syntagmatic relationship:** In addition to the expression of paradigmatic relationships, rules and facilities are provided for the coordination of terms from the vocabulary in order to express more complex meanings. Syntagmatic relationships, also called syntactical relationships are achieved by means of syntactical rules of the given indexing language. Two major syntactic devices very much common in indexing languages are the use of word or term order and relators or linking mechanisms. Kaiser's Thing—Process, Ranganathan's PMEST, Coates's Thing—Material—Action and Relationship Table are some of the examples of formulae for determining term order and thus standardising and controlling syntax. A major principle underlying term order is significance. Component order of a compound subject heading can be expressed in more than one way. The question of order can be answered by reference to significance—that is, by an analysis of the relative importance to the searcher of the concepts concerned. The result of such analysis is to bring into prominence key concepts. Syntagmatic relationships are document dependent relationships because these relationships are established with reference to the concepts associated with the content of a given document. As for example, a document entitled "Social aspects of literacy among rural women in India" will call for the combination of concepts from Sociology, Education and Geography. They represent the syntagmatic relation in the context of this specific document and an indexing language must have the mechanism to represent these concepts in a subject heading.

10.5.6 Structural Presentation

The basic objective of an indexing language is to provide subject approach to the contents of documents to the users. It is generally agreed that a user-oriented approach may not be confined to the specific subject only. A user who starts a search for looking a specific subject *'Conservation of tiger'* may avoid to notice a document on *'Conservation of wildlife'* which may contain equally valuable information on *'Conservation of tiger'* because he believes that the more specific subject *'Conservation of tiger*' will not be covered in the document *'Conservation of wildlife'*. Similarly, *'Conservation of tiger*' may contain equally valuable information on the *'Conservation of wildlife'*. Thus it appears that the broader as well as narrower subjects may help the user even with a specific search. This situation calls for structuring the indexing language in such a systematic manner that the semantic network of concepts and relationship between concepts are displayed in it. All the indexing languages display such relationships in one way or other and thus all of them are structured. A classification scheme displays such relationship by notation whereas a verbal indexing language like a readymade list of subject headings and thesaurus display such relationship by the relationship indicators BT and NT.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

1) Distinguish between indexing language and the language we use in our day to day life.

....

....

....

....

2) Discuss the functions of the different components of an indexing language?

....

....

....

....

3) What are the different attributes of an indexing language?

....

....

....

....

10.6 VOCABULARY CONTROL

It transpires from the above discussion that it is hardly possible to keep a natural language clear of synonyms and homonyms. If we use natural language for subject indexing, subject matter may be described by any combination of words or phrases, without limitation, such as those occurring in documents themselves. The vocabulary of an indexing language has to be very precise and exact. Both synonyms and homographs are to be controlled. Terms accepted as standard terms (also called descriptors, preferred terms, candidate terms, index vocabulary, etc.) are to be linked with the respective alternative terms (i.e. non-standard terms, non-preferred terms, etc.) by way of elaborate system of references. This will result in complete one to one correspondence between concepts and terms.

10.6.1 Meaning and Need

The term "vocabulary control" refers to a limited set of terms that must be used to index documents, and to search for these documents, in a particular system. It may be defined as a list of terms showing their relationships and indicating ways in which they may usefully be combined to provide specific subject of a document. A certain degree of structure is introduced in a controlled vocabulary so that terms whose meanings are related are brought together or linked in some way. The opposite, an uncontrolled vocabulary (i.e. free vocabulary), is an unlimited set of terms drawn from natural language.

However, certain problems in searching do arise when no control is imposed on the vocabulary. This is because of the fact that a natural language contains a large amount of synonyms, quasi-synonyms, homonyms, acronyms, ambiguous terms, etc. Hence, if vocabulary control is not exercised, different indexers or one and the same indexer at different times is likely to use a different set of terms for representing the thought content of similar documents (i.e. documents dealing with the same subject) on the one hand, and to use a different set of terms for representing the same subject at the time of searching. This, in turn, would result in a "mis-match" and thus affect information retrieval. In short, we can say that the need for vocabulary control arises to overcome the following problems:

- Occurrence of imprecisely defined words,
- Rapidly changing terminology,

- Numerous synonyms for a term, and
- Problem of homographs.

10.6.2 Objectives

The basic objective of controlled vocabularies is to provide a means for organising information. Through the process of assigning terms selected from controlled vocabularies to describe documents and other types of content objects, the materials are organised according to the various elements that have been chosen to describe them.

Controlled vocabularies serve five purposes:

1) **Translation:** Provide a means for converting the natural language of authors, indexers, and users into a vocabulary that can be used for indexing and retrieval.
2) **Consistency:** Promote uniformity in term format and in the assignment of terms.
3) **Indication of relationships:** Indicate semantic relationships among terms.
4) **Label and browse:** Provide consistent and clear hierarchies in a navigation system to help users locate desired content objects.
5) **Retrieval:** Serve as a searching aid in locating content objects.

From the operational point of view, there are basically two fold objectives for having a controlled vocabulary:

1) To promote the consistent representation of the subject matter of documents by indexers and searchers, thereby avoiding the dispersion of related documents, through control of synonymous and nearly synonymous expressions and by distinguishing among homographs, and
2) To facilitate the conduct of a comprehensive search by bringing together in some way the terms that are most closely related semantically.

10.6.3 Methods

The methods of vocabulary control refer to the various means by which the objectives of controlled vocabulary are achieved. The first objective is achieved by controlling the terminology. Control of terminology is achieved in various ways. First the form of term is controlled, whether this involves grammatical form, spelling, singular and plural form, abbreviations or compound form of terms. Second, a choice is made between two or more synonyms, near-synonyms and quasi-synonyms. Third, homographs are distinguished. The control of synonyms is achieved simply by choosing one of the possible alternatives as the "Preferred term" and referring to this term (by using "See" or "Use") from the variants under which certain users may be likely to approach. It should be obvious that the synonyms selected as the preferred term (i.e. the term under which documents will actually be indexed and search for) must be the one under which the majority of users will be likely to look first. Sometimes "quasi-synonyms" are created in the same way as synonyms (i.e. one is chosen and the reference is made from the other). The term "quasi-synonym" is not very precise. Many authors consider the quasi-synonyms as the antonyms that represent opposite extremes on a continuum values. An example is the pair of "roughness" and "smoothness". Clearly "roughness" may be regarded as merely the "absence of smoothness", and vice-versa.

The controlled vocabulary also distinguishes among homographs (i.e. words with identical spelling but the different meaning), usually by means of a parenthetical qualifier or scope note. Thus CRANE (Bird) tells us that this term is to be used exclusively for a type of birds and not as lifting equipment or any other possible context.

By controlling synonyms, near-synonyms and quasi-synonyms, and by distinguishing among homographs, the vocabulary control device avoids the dispersion of like subject matter and the collocation of unlike subject matter. In this way it helps to achieve the object of consistent representation of subject matter in indexing and searching.

The second objective of the vocabulary control device is to link together terms that are semantically related in order to facilitate the conduct of comprehensive searches. A controlled vocabulary will bring together terms that are hierarchically related (in a formal genus-species relationships), and it will also reveal semantic relationship across hierarchies (i.e. terms having non-hierarchical associative relationships). This objective is achieved by means of "See also" references among terms that are most closely related semantically.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

4) What is the difference between paradigmatic and syntagmatic relationships?

....

....

....

....

5) What do you understand by 'Controlled vocabulary'?

....

....

....

....

6) State the objectives of vocabulary control. How are these objectives achieved?

....

....

....

....

10.7 TYPES OF INDEXING LANGUAGES

10.7.1 Classification Schemes

It has already been mentioned in sub-section 10.4.1 of this Unit that controlled vocabulary of an IL is of two types: verbal and coded vocabularies. A classification scheme employs coded vocabulary in the form of its notation. Libraries have long been using notational schemes of library classification to organise information resources on the shelves, and to provide means for locating information resources in the bibliographical tools—such as, catalogues, bibliographies, abstracting and indexing documents. We know that classification is a mental process of grouping of entities in order of their degree of likeness and separating entities according to their degree of unlikeness. All class designations of subjects are the names of subjects irrespective of the fact that whether they are in terms of class numbers or verbal specifications. The assignment of class designation in the notational plane is called class number, and the preparation of tools to be used for this purpose is classification schemes. Modern classification schemes such as Dewey Decimal Classification (DDC), Universal Decimal Classification (UDC), Library of Congress (LC) Classification, Colon Classification (CC), etc. were devised several decades ago or more. Although over the years they have been modified and improved, their main objectives remain unchanged. The Web version of DDC, 22nd edition, i.e. WebDewey includes all updates since its publication in 2003 plus supplemental data. The most important feature of WebDewey is that it gives additional points of access by combining DDC numbers and Library of Congress Subject Headings (LCSH). It also gives access to many pre-built numbers, especially in the Literature class which are not available in the print version Although classification schemes were mainly designed for organising bibliographic items, many researchers have also used classification schemes to organise information resources on the Web.

10.7.2 Subject Heading Lists

Subject heading has been defined as a word or group of words (phrase) indicating a subject under which all materials dealing with same theme is entered in a catalogue or bibliography, or is arranged in a file. A vocabulary control device depends on a master list of words/terms that can be assigned to documents. Such a master list of terms is referred to as a list of subject headings. A subject heading list is an alphabetical list terms and phrases, with appropriate cross references and notes, that can be used as a source of subject headings in order to represent the subject content of a document. A list of other semantically related terms or phrases are displayed under each term or phrase. A printed list of subject headings incorporates the thought and experience of many librarians of various types of libraries.

General Principles

The rules for subject headings in a dictionary catalogue were formulated by Charles Ammi Cutter in 1876 in his 'Rules for a Dictionary Catalog'. The impact of Cutter's principles on construction and maintenance of subject headings is still discernible today. Both, the LCSH and the SLSH adopted the Cutter's principles in assigning subject headings for a document. The general principles that guide the indexers in the choice and rendering of subject headings from the standard lists of subject headings are discussed in the 'following sub-sections.

- **Specificity**

The principles of specific and direct entry require that a document be assigned directly under most specific subject heading that accurately and precisely represents its subject content. If a document is about penguins, it should be entered directly under the most specific heading 'Penguins', not under the heading 'Birds' or even under 'Water Birds' which includes Penguins'. If the name of a specific subject is not available, a broader heading is the most specific authorised heading in the hierarchy that covers the content of the work. In many cases, several headings may be assigned in order to cover different aspects of a subject.

- **Common Usage**

This principle states that the word(s) used to express a subject must represent common usage. There may be problems in the selection of subject headings when the same concept is expressed by two or more terms. According to this principle, subject headings are to be chosen keeping in mind the needs of the users who are likely to use the index file. If a choice between spellings is made for dialectal reasons (for example, between American and British English), the most widely accepted spelling of words, based on users warrant, should be adopted. If a popular and a scientific name refer to the same concept, the form most likely to be sought by the users should be chosen. After deciding on the name of heading, a cross-reference should be made from the non-preferred to the preferred form.

- **Uniformity**

The principle of uniform heading is adopted in order to bring consistency in the use of subject headings. A subject heading list has to be very precise and exact in order to ensure that each concept is represented by a single preferred term. Both synonyms and homographs are to be controlled. It should list the other synonyms and variants as non-preferred terms with USE references to the preferred term. One uniform term must be selected from several synonyms and other variants, and this term must be applied consistently to all documents on the topic. If several meanings are attached to one term (e.g, Crane as a bird / Crane as a lifting equipment) that term must be qualified so that it will be clear to the users for which the meaning is intended.

- **Consistent and Current Terminology**

The principle states that the term(s) chosen as subject headings should be both consistent and current as has already been said regarding the justifications for uniform headings. By principle, common usage prevails when there is a problem of choices among synonymous terms and other variants. Changes in usage also present many practical difficulties. A term chosen on the basis of common usage may become obsolete with the passage of time. Subsequently, a list of subject headings may incorporate current terminology as long as entries pose a problem because of the large number of entries listed under the existing subject headings. In such a situation a subject authority file is to be maintained. Once a heading is changed, every record that was linked to the old heading can be linked to the new heading and this decision is recorded in the subject authority file.

- **Form Heading**

Form headings refer to those words or phrases which represent the literary or artistic form (e.g., Essays, Poetry, Fiction, etc.). These are the words or phrases that follow a subject heading and indicated by a dash. These words or phrases are used to make the subject more specific. Assignment of form headings to individual works as well as to collections and materials about the form enables the libraries to provide access to these kinds of materials to the users. Apart from literary works themselves there are also many kinds of library materials about literary forms that require subject headings. For a document on how to write an essay, the heading "Essay" represents a subject. A topical subject heading and a form headings can be distinguished by using the singular form for the topical subject heading and plural for the form heading (e.g., Short story, Short stories). In addition to the literary form headings, there are some other form headings that are determined by the general format and purpose of the documents, such as Almanacs, Encyclopaedias, Dictionaries, and Gazetteers.

- **Cross Reference**

Cross-references direct the user from term/phrase not used as headings to the term/

phrase that is used, and from broader and related topics to the one chosen to represent a given subject. Three types of cross-references are used in the subject headings structure. These are discussed below:

See (or USE) references: These references guide users from terms that are not used as headings to the authorized headings for the subject in question. 'See' or 'USE' references ensure that inspite of different names for (or different forms of the name of) a given subject a user shall still be able to locate materials on it.

See also (including BT, NT, and RT) references: These references guide users to the headings that are related either hierarchically or associatively and are used as entries in the index file. By connecting related headings, the 'see also' (RT, for related term) references draw the user's attention to material related to his interest. By linking hierarchically related headings, 'see also' (BT, for broader term; NT, for narrower term) references directs the users to search specific deviations or aspects of his subject of interest.

General references: General references direct the users to a group or category of headings instead of individual headings. It is sometimes called a 'blanket reference'. The provision of general references in the standard list of subject headings obviate the need to make long lists of specific references and thus ensure economy of space.

Subject Authority File

A subject authority file consisting of subject authority records ensures uniformity and consistency in subject heading terminology and cross-references. The process of creating subject authority records and maintaining subject authority file is called subject authority control. Subject authority control is exercised at two levels: central and local. At the central level, a central agency (e.g. Library of Congress) maintains the subject authority file (in card or machine-readable form) or subject heading list (in print form) and makes changes to existing headings and cross-references as well as adding new ones. At the local level, a library creates local subject authority records only for headings not yet appeared as established headings in a subject heading list along with needed maintenance information. Thus, the subject authority control at the local level includes correcting erroneous headings and cross-references, updating obsolete headings, and adding or revising cross-references necessitated by new headings. ALA Glossary has defined the subject authority file as *"A set of records indicating the authorized forms of terms used as subject headings in a particular set of bibliographic records; the references made to and from the authorized forms; and the information used, and its sources, in the establishment of the headings and the determination of the references to be made".* (ALA Glossary of Library and Information Science. Chicago: American Library Association, 1983, p.220). This definition suggests that a subject authority record should contain the following items of information: (a) established subject heading; (b) scope notes, if any; (c) cross-references made from it to other headings; and (d) sources or authorities on which the decision on the form of heading was based. A subject authority record is made when subject headings are established and used for the first time.

The functions of a subject authority file are discussed below:

- **Indexing:** The subject authority file serves as the source of indexing vocabulary and as the means of verifying or validating headings assigned to individual indexing records. It helps to ensure that: a) the same heading is assigned to all works on the same subject, b) each heading represents only that particular subject, and c) all headings assigned to indexing records conform to the established forms.

- **Maintenance:** Necessary adjustments to indexing records are needed to be added from time to time as a result of changes in the indexing vocabulary. When existing subject headings are revised or new headings are added; cross-references are often affected and should be adjusted. The subject authority file reflects the most current status of headings and cross-references and thus, serves as the source for verification and validation of subject headings as to the indexing records. It is also useful when a library converts its manual form to the online mode and wishes to have previously existing records reflect current practice.

- **Retrieval:** Subject authority file helps the users in two ways: (1) subject headings displayed in the subject authority file show the user the terminology and form of subject access points in the index file; and (2) the cross-references guide the users to related headings when user's input terms fail to retrieve useful records.

Self Check Exercise

Note: i) Write your answer in the space given below.

ii) Check your answer with the answer given at the end of this Unit.

7) What are the different types of the indexing languages?

....

....

....

....

8) State the principles that guide the indexers in the choice and rendering of subject headings from a standard list.

....

....

....

....

9) Mention the importance of a subject authority file to the cataloguer.

....

....

....

....

10.7.3 Thesaurus

The term *thesaurus* has been derived from Greek and Latin words which mean 'a treasury' and it has been used for several centuries to mean a lexicon or treasury of words. Modern usage may be said to date from 1852, when the first edition of *Thesaurus of English Words and Phrases* was published by Peter Mark Roget. A thesaurus (plural: thesauri) with which we are concerned is meant for information retrieval and is used as a valuable vocabulary control device for indexing and searching in a specific subject area. The journey of the thesaurus from the linguistic domain to information retrieval is evident from the following timeline:

1736: The term 'thesaurus' first appeared in OED. It came from the Greek word 'thesauros' which means 'Treasury or storehouse of knowledge'.

1852: Appeared Peter Mark Roget's Thesaurus. It was a linguistic thesaurus showing the word(s) by which the given idea most fitly and aptly expressed, i.e. Classification of ideas.

1957: Dorking Conference. Miss Helen Brownson first brought the idea of 'thesaurus' in terms if IR through a paper presented there.

1959: H P Luhn gave the idea of the application of thesaurus in IR.

1969: The first thesaurus used in IR system was developed by Du Pont in USA.

Definition

There are many different definitions of thesauri, varying from quite modest definitions that focus on the relations between words without stating which kinds of relations that are meant, to such definitions that state more exactly which relations that are concerned. The definition of *Thesaurus* provided by World Science Information System of UNESCO (known as UNISIST) on the basis of its function and structure seems to be most comprehensive to understand the meaning and scope of the thesaurus:

"In terms of function, a thesaurus is a terminological control device used in translating 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 and generically related terms which covers a specific domain of knowledge".

Purpose

A thesaurus is a semantic network of terms. Its purposes are

a) To provide a map of given field of knowledge, how concepts or ideas about concepts are related to one another, which helps an indexer or a searcher to understand the structure of the field.

b) To provide a standard vocabulary for a given subject field which will ensure that indexers are consistent when they are making index entries to an information storage and retrieval system.

c) To provide a system of references between terms which will ensure that only one term from a set of synonyms is used for indexing one concept, and that indexers and searchers are told which of the set is the one chosen; and to provide guide to terms which are related to any index term in other ways, either by classification structure or otherwise in the literature.

d) To provide a guide for users of the system so that they choose a correct term for a subject search; this stresses the importance of cross references. If an indexer uses more than one synonym in the same index—for example, "abroad", "foreign" and "overseas"—then documents are liable to be indexed haphazardly under all of these; a searcher who chooses one and finds documents indexed there will assume that he has found the correct term and will stop his search without knowing that there are other useful documents indexed under the other synonyms.

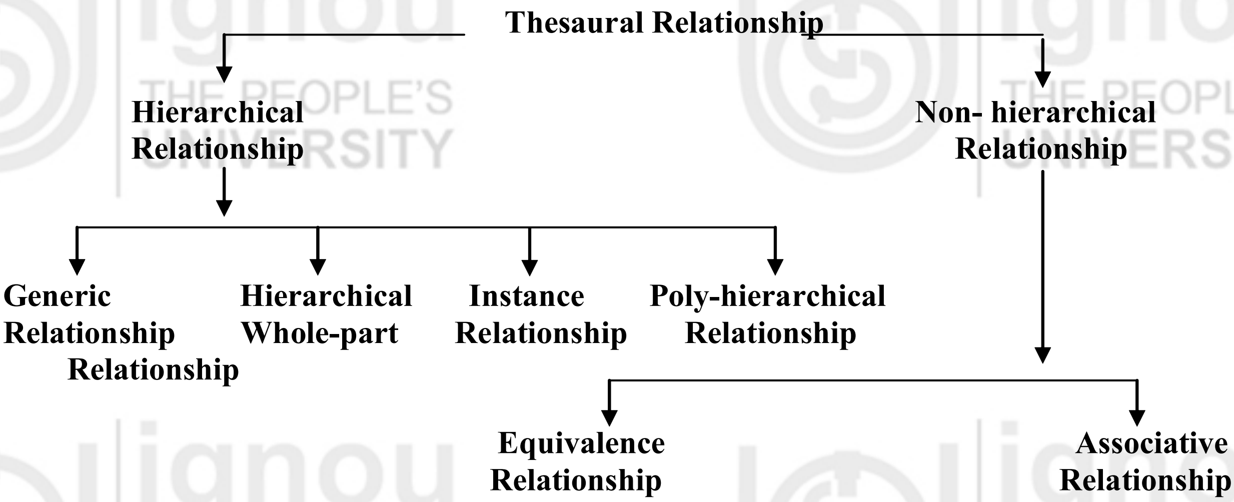
e) To locate a new concept in a scheme of relationships with existing concepts in a way which makes sense to users of the system.

f) To provide classified hierarchies so that a search can be broadened and narrowed systematically, if the first choice of search term produces either too few or too many references to the materials in the store.

g) A desirable purpose, but one which it would be premature to say is being achieved, is to provide a means by which the use of terms in a given subject field may be standardised.

Basic Thesaural Relationships

Basic thesaural relationships or the semantic relationships in a thesaurus refer to two types of relationships: (1) Hierarchical Relationship; (2) and Non-Hierarchical Relationship. The following figure shows the different types of relationships displayed in a thesaurus.



Hierarchical Relationship

Hierarchical relationships are based on degrees or levels of superordination and subordination, where the superordinate term represents a class or a whole, and subordinate terms refer to its members or parts. This relationship is of four types: Genus-Species (Generic) relationship, Whole-Part relationship, Instance relationship and Poly-hierarchical relationship.

Reciprocity in the hierarchical relationships is expressed by the relationship indicators: BT (Broader Term), i.e. a label for the superordinate (parent) term; and NT (Narrower Term), i.e. a label for the subordinate (child) term.

- **Genus-Species (Generic) Relationship** links genus and species and represents the basis of scientific taxonomic system. As for example

Examples of Hierarchical relationship indicator (BT and NT),

Mammals BT Vertebrates	Vertebrates NT Mammals

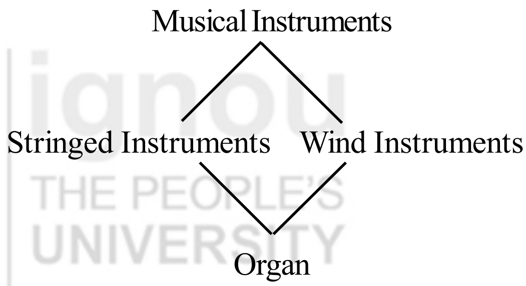
- **Whole-Part Relationship** covers situations in which one concept is inherently included in another, regardless of context, so that the terms can be organised into logical hierarchies, with the whole treated as a broader term. As for example:

Central nervous system NT Spinal cord	Spinal cord BT Central nervous system

- **Instance Relationship** identifies the link between a general category of things or events, expressed by a common noun, and an individual instance of that category, often a proper name. As for examples:

Mountain regions NT Himalayas	Himalayas BT Mountain regions

- **Polyhierarchical Relationship** occurs when some concepts belong, on logical grounds, to more than one category. In the following example, the term *pianos* is assigned to subordinate positions on the basis of its generic relationship to two broader terms—in other words, *pianos* would be an NT to both *stringed instruments* and *wind instruments.*



Non-Hierarchical Relationship

Relationship between terms other than hierarchical is called Non-hierarchical relationship, which may further be grouped as Equivalence (or Preferential) Relationship and Associative Relationship.

Equivalence (or Preferential) Relationship refers to the relationship between preferred and non-preferred terms in which each term is regarded as referring to the same concept. When the same concept can be expressed by two or more terms, one of these is selected as the preferred term. A cross-reference to the preferred term should be made from any "equivalent" entry term. Reciprocity in the equivalence relationships is expressed by the relationship indicators: USE, which leads from a non-preferred (entry) term to the preferred term, and UF or USED FOR, which leads from the preferred entry term to the non-preferred term(s).

Four basic types of equivalence relationship are evident: (a) Synonyms; (b) Lexical variants; (c) near-synonyms; and (d) Generic posting.

- **Synonyms:** Synonymy occurs when a concept can be represented by multiple terms having the same or similar meanings. A thesaurus compensates for the problems caused by synonymy by ensuring that each concept is represented by a single preferred term. It lists other synonyms and variants as non-preferred terms with USE references to the preferred term.

Birds UF Aves	Aves USE Birds
Vaseline UF Petroleum jelly	Petroleum jelly USE Vaseline

- **Lexical variants:** Lexical variants differ from synonyms in that synonyms are different terms for the same concept, while lexical variants are different word forms for the same expression. These forms may derive from spelling or grammatical variation or from abbreviated formats. The following examples indicate the preferred grammatical forms of terms.

Nouns and Noun Phrases: The grammatical form of a term should be a noun or noun phrase. Nouns used as terms are divided into two categories: Count nouns and Noncount (mass) nouns. Count Nouns are names of objects or concepts that are subject to the question "How many?" but not "How much?" These should normally be expressed as plurals. For examples: books, penguins, singers, vertebrates, windows, etc. Mass (noncount) nouns are names of materials or substances that are subject to the question "How much?" but not "How many?" These should be expressed in the singular. Some examples of Singular mass nouns are: milk, water, etc.

Where the singular and plural forms of a term represent different concepts, separate terms for each are entered in the thesaurus. The distinction should be indicated by a qualifier. Some examples are: Bridge (game) / Bridges (structures); Damage (injury) / Damages (law); Wood (material) / Woods (forested areas).

Noun phrases are compound terms that are established as preferred terms if they represent a single concept. Noun phrases occur in two forms: (a) Adjectival noun phrases like Red rose, Marine birds, Cold fusion, Historical drama, etc.; and (b) Prepositional noun phrases like Plaster of Paris, Prisoners of war, Hospitals for children, etc.

Adjectives: Adjectives and adjectival phrases used alone are established as terms in a thesaurus under certain special circumstances. Single adjectives is used in a "nominal" way; that is, the noun is obvious from the context or the adjective is used to describe an attribute of the content object other than topic, such as *colour* or *size*. For examples: small, medium, large, blue, green, red, yellow, etc.

As an alternative to the creation of multiple compound terms, adjectives may appear as separate terms when designed to be precoordinated in indexing or postcoordinated in searching. They should generally not be assigned as indexing terms in isolation. Given the possibility of false coordination in searching (e.g., the linking of an adjective with the wrong noun), adjectival terms should be used sparingly. Some examples of the use of adjectives as terms in pre- and post coordination are: Airborne / Airborne troops; Offshore / Offshore drilling; Mobile / Mobile homes, etc.

Certain noun phrases may be used to modify other nouns, e.g., *high frequency* can modify the noun *waves*.

Adjectives may be used alone in general cross references to direct the user to or from a group of terms beginning with a corresponding noun, e.g., "*cardiac* . . . see also the terms beginning with *heart*." An example of a reference in the opposite direction (noun to adjective) is: "*France* see also the terms beginning with *French* (French art, French language, French literature, French wines)."

Adverbs: Adverbs such as "very" or "highly" should not be used alone as terms. A phrase beginning with such an adverb may be accepted as a term only when it has acquired a specialized meaning within a domain. Some examples of Adverbial phrases are: very high frequency, very large scale integration, very low density lipoproteins, etc.

Abbreviations: Abbreviations are selected as preferred terms only when they have become so well established that the full form of the term or proper name is rarely used, e.g. *AIDS* rather than *Acquired Immune Deficiency Syndrome; Lasers* rather than *Light Amplification by Stimulated Emission of Radiation; UNESCO* rather than United Nations Educational, Scientific, and Cultural Organization; etc. The full form of terms are selected as preferred terms when the abbreviated form is not widely used and generally understood, e.g. *Automated teller machine* rather than *ATM; Prisoners of war* rather than *POW,* etc. Cross-references should be made from the non-preferred forms to preferred form.

Popular and Scientific Names: If a popular and a scientific name refer to the same concept, the form most likely to be sought by the users of the thesaurus should be chosen as the preferred term. For example, *Penguins* is chosen as the preferred term in a nontechnical thesaurus with a cross reference from the scientific equivalent, *Sphenisciformes***.** However, *Sphenisciformes* is selected as the preferred term in a zoological thesaurus with a cross-reference from the popular name, *Penguins.*

- **Near-synonyms:** Near-synonyms are terms whose meanings are generally regarded as different, but which are treated as equivalents for the purposes of a controlled vocabulary. The extent to which terms are treated as near-synonyms depends in large measure upon the domain covered by the controlled vocabulary and its size. Near-synonyms may include antonyms or represent points on a continuum. As for examples, Sea water/salt water [variant terms]; Smoothness/ roughness [antonyms].

- **Generic Posting:** It is a technique in which the name of a class and the names of its members are treated as equivalents, with the broader class name functioning as the preferred term. As for examples, Waxes UF Plant waxes; Plant waxes USE Waxes.

Associative Relationships

This relationship covers associations between terms that are neither equivalent nor hierarchical, yet the terms are semantically or conceptually associated to such an extent that the link between them is made explicit in the thesaurus, on the grounds that it may suggest additional terms for use in indexing or retrieval. The associative relationship used in thesauri is indicated by the abbreviation RT (Related Term). As a general guideline, whenever one term is used, the other should always be implied within the common frames of reference shared by the users of the thesaurus. Either of the following types of terms can be linked by the associative relationship:

a) Those belonging to the same category, and

b) Those belonging to different categories.

- **Relationships between terms belonging to the same category**

Relationships are needed for terms belonging to the same category in various special situations, primarily to guide the user in locating the desired term. Each of the terms belonging to the same category has its own particular meaning, but the boundary between them is often confused with common usage, to the extent that a user checking one of them in the index should be informed of documents indicated by others. As for examples:

Ships RT Boats	Carpets RT Rugs
Boats RT Ships	Rugs RT Carpets

- **Relationships between terms belonging to the different categories**

It is possible to establish many grounds for associating terms belonging to different categories. Related Term references are often made between etymologically related terms, i.e., terms that contain the same root, but which do not represent the same kind of thing. The following are some representative examples of typical relational situations.

a) **Process / Agent**

Temperature control

RT Thermostats

Thermostats

RT Temperature control

b) **Process / Counteragent**

Inflammation

RT Anti-inflammatory agents

Anti-inflammatory agents

RT Inflammation

c) **Action / Property**

Polling

RT Public opinion

Public opinion

RT Polling

d) **Action / Product**

Weaving

RT Cloth

Cloth

RT Weaving

e) **Action / Target**

Harvesting

RT Crops

Crops

RT Harvesting

f) **Cause / Effect**

Cloud

RT Rain

Rain

RT Cloud

g) **Concept or Object / Property**

Poisons

RT Toxicity

Toxicity

RT Poisons

h) **Concept or Object / Origins**

Americans

RT United States

United States

RT Americans

i) **Concept or Object / Units or Mechanisms of Measurement Associative Relationships**

Electric current

RT Amperes

Amperes

RT Electric current

j) **Raw Material / Product**

Wheat

RT Flour

Flour

RT Wheat

k) **Discipline or Field of Study / Object or Phenomenon Study**

Neurology

RT Nervous system

Nervous system

RT Neurology

l) **Discipline or Field of Study / Practitioner**

Mathematics

RT Mathematicians

Mathematicians

RT Mathematics

m) **Antonyms**

Height

RT Depth

Depth

RT Height

n) **Phrases Containing Syncategorematic Nouns and their Apparent Foci**

Ships

RT Model ships

Model ships

RT Ships

o) **Coordinate Ideas**

Hinduism

RT Buddhism

Christianity

Islam

10.7.4 Thesaurofacet

Thesaurofacet: a thesaurus and faceted classification for engineering and related topics was developed from the English Electric Company's *Faceted classification for Engineering,* the first edition of which was published in 1958. Thesaurofacet came about when the third edition of *Faceted Classification for Engineering*, published in 1961, was up for revision. This system was used to organise documents belonging to the libraries of the corporation of English Electric. However, with the growing trends in science and technology and the need for using computer techniques and post-coordinate indexing, a decision was taken in 1967 to commission Jean Aitchison, a member of Classification Research Group, to review the indexing needs of the company and the result of that review was the compilation of a new and improved 4th edition of the faceted classification system called Thesaurofacet, published in 1970. In the 4th edition, the alphabetised index to the classification scheme was replaced with a thesaurus.

Thesaurofacet covers the whole field of science and technology but subjects are treated in varying depth and only engineering and allied fields are covered exhaustively. Full subject coverage includes engineering and fields directly related to engineering like computers, measurement and testing, physics and management. Relevant management-related concepts were borrowed from the "Classification of Business Studies" developed by London Graduate School of Business Studies.

Thesaurofacet is considered as a multi-purpose retrieval language tool because it has classification schedules and a faceted thesaurus. The classification consists of main classes and facets and has a notation system that consists of letters in upper case and numbers from 2 to 9. The faceted thesaurus is the key to the uniqueness of the tool because it offers the user options to identify topics within the system. Because the two are linked, each term in the system appears twice, once in the schedule, and once in the thesaurus, with notation that links the two parts together. However, the information given about the term in the thesaurus is not the same information given about that term in the schedule. The two parts of the system are complementary and should be used conjunctively and not separately. Finally, Thesaurofacet can be used for the arrangement of books on the shelves and arrangement of entries in the subject catalogues. Further, the index terms are intended to be used for indexing and searching.

If we are asked for information on *Documentation,* we turn to the thesaurus and find:

Documentation *use*
Information Science

At *Information Science* we find

Information Science **ZR**

UF Documentation

Librarianship

Library science

RT Communication (Sociology)

Data processing

Information theory

Librarians

We also see the notation to the right and we are told to look for ZR in the classification schedules. At ZR in the classification schedule we find that Thesaurofacet divides *Information Science* using subjects and facets:

Main Class

Subject Field(s)

Fundamental Facets

Sub-Facets

Hierarchies and Arrays

ZR *Information science*

ZR2 **LIBRARIES**

By type:

ZR3 National libraries

ZR4 Public libraries

ZR5 Municipal libraries

ZR6 County libraries

ZRB Educational Libraries

____

____

By management:

ZRP Library management

____

____

By equipment:

____

____

By materials:

____

____

ZT INFORMATION RETRIEVAL

By type of language:

ZT2 **Index languages**

Subdivided by type:

ZT3 **Natural language**

ZT4 Single term natural language

ZT5 Concept natural language

ZT8 **Controlled index languages**

____

____

Here *Information Science* is called Main Class. Main class can be divided into subject fields, in this case LIBRARIES, and INFORMATION RETRIEVAL. Subject fields are broken down into fundamental facets, and are printed by bold typeface (e.g. by type, by management, by equipment, etc.). This is where the schedules start to use facet analysis. If we take a look at one of fundamental facets as an example, Information Retrieval is broken down 'By type of language' into the facet *Index languages.* The facet is broken down into sub-facet *Natural Language* and *Controlled index languages*. Terms are then listed in hierarchies and arrays. The schedules are typically used for a broad subject search and browsing through the possible topics that exist within a class. So a user may start with the main class and navigate through the facets and arrays to locate an ideal or more specific topic.

The classification schedules also allow for the combination or synthesis of topics and notation. For example, there is a note in the main class ZL SOCIOLOGY that tells us that this topic can be combined with ZM PSYCHOLOGY to create the synthesized subject field called ZL/ZM SOCIAL PSYCHOLOGY. Synthesis can indeed be used wherever required, there being no preferred combination order unless there is an instruction in the schedules.

Thus it appears that two types or styles of "faceted classifications integrated with thesauri": First type uses subject fields as main subdivisions, and facet analysis is used to determine the relationships and the second type of faceted thesauri are those in which concepts are first divided by facets.

10.7.5 Classaurus

The vocabulary control device used for POPSI has been designated as *Classaurus*. It is a category-based (faceted) systematic scheme of hierarchical classification in verbal plane incorporating all the essential features of a conventional IR thesaurus—i.e. control of synonyms, quasi-synonyms, etc. RTs are not shown in the classaurus. A scheme of this type, for its application, calls for a complementary alphabetical index giving the address of each term occurring in the systematic part. The purpose for which a classaurus is used does not necessarily warrant any principle-based arrangement of the terms in the array. Even if the terms in each array are arranged alphabetically the purpose is not going to be disturbed. This feature of the classaurus makes it largely amenable to computerisation.

The structure and style of presentation of a classaurus can be systematically presented as follows:

A) **Systematic Part**

A1 Common Modifiers

A1.1 Form

A1.2 Time

A1.3 Environment

A1.4 Place

A2 Inter-subject Relation Modifiers

A3 Discipline and Sub-disciplines

A4 Entities

A4.1 Part

A4.2 Type

A5 Properties

A6 Actions

In respect of systematic part, the following points are to be noted:

a) Each term in the systematic part under each category is enumerated by displaying its COSSCO relationship in a hierarchy of arrays.

b) For each term in the systematic part, the following follows vertically: (a) Definition/ Scope note (if required), and synonyms, quasi-synonyms, and antonyms.

c) No RTs (i.e. non-hierarchically related terms) are enumerated for any term in the classaurus because of its category-based structure. It is assumed that RTs should not be dictated by the designer of the classaurus, rather it should be dictated by the document itself. Any term may be related to any other terms depending upon the nature of the thought-content of the document. Hence, RTs should not be determined beforehand.

d) Each array in the classaurus is open.

e) Each term in the systematic part is assigned a unique address which can be used as a class number.

B) **Alphabetical Index Part**

This part contains each and every term including synonyms, quasi-synonyms, and antonyms occurring in the systematic part along with its address.

Self Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

1) What do you mean by a "thesaurus"?

....

....

....

....

2) Discuss the functions of a thesaurus.

....

....

....

....

3) What are the different types of relationships displayed in a thesaurus?

....

....

....

....

4) What is the difference between lexical variants and synonyms?

....

....

....

....

5) Discuss the essential features of a thesaurofacet?

....

....

....

....

6) Why are the RTs not displayed in a classaurus?

....

....

....

....

10.8 LIBRARY OF CONGRESS SUBJECT HEADINGS

The Library of Congress Subject Headings (LCSH) is used quite widely as a controlled vocabulary for catalogues and bibliographies all over the world. Andrea Crestadoro is usually given credit, who, for the first time in his book *The Art of Making Catalogues* published in 1856, pointed out that the cataloguer should provide a standardised guide to the subject content of a book by giving it a heading. In 1895, the first standard list of subject headings entitled *List of Subject Headings for Use in Dictionary Catalogues* produced by a committee of the American Library Association (ALA) of which C.A. Cutter was a prominent member, was based on Cutter's principles. It went through three editions (1895, 1898, and 1911). In 1910-1914, when Library of Congress (LC) began publishing its list under the title *"Subject Headings Used in the Dictionary Catalogs of the Library of Congress"*, it was found unnecessary to continue the ALA list. LCSH was basically designed for representing the subject and form of books and serials in the Library of Congress collection, now it is not only considered as worldwide authority for subject headings in the English language but has also been translated into several other languages. Later renamed *Library of Congress Subject Headings (LCSH),* the list has been updated in many editions. It is available in various formats including hard copy, microfiche, CD-ROM, and online through the Web (locis.loc.gov.). Weekly lists of additions are available through the Web at www.loc.gov/aba/cataloging/subject/weelylists, replacing the monthly printed version, which was discontinued at the end of 1994. Presently it is in its thirtieth edition (5 volumes, 2007) containing over 280,000 headings and references. *Subject Headings Manual (2008)* is an essential supplementary tool containing the rules and guidelines used by the cataloguer while applying the LCSH correctly. It is issued in loose-leaf form with periodic updates. *Subject Headings Manual* replaces the *Subject Cataloguing Manual: Subject Headings* (5th edition, 4 vols. Loose-leaf. 1996) and all of its updates. In the following sections, we shall study some of important features of the LCSH.

10.8.1 Types of Subject Headings

Subject headings in LCSH are constructed in variety of ways, ranging from single noun to complete descriptive phrases.

- **Single word headings**

A single word heading is usually a noun representing objects, things, persons, or concepts; e.g. **Skating**, **Electrometallurgy**, etc.

Homographs are generally distinguished to avoid ambiguity by providing a noun or a phrase as parenthetical qualifier with the primary term; e.g. **Constructivism (Art)** / **Constructivism (Education)** / **Constructivism (Philosophy)**. Use of parenthetical qualifier, however, has not been restricted to the disambiguation of homonyms. They have also been used to provide general contextual information, e.g. **Cookery (Onions)**; **Marriage (Canon law)**.

- **Adjectival phrase headings**

Adjectival phrase headings start with a modifier followed by a noun or noun phrase. The modifier may be an adjective or it may be a noun used as an adjective. As for examples:

a) Common adjective, e.g. **Administrative law**; **Dental records**.

b) Ethnic, national or geographic adjective, e.g. **African American librarians**; **Spanish literature**.

c) Participial modifiers, e.g. **Applied anthropology**; **Hearing aids**.

d) Common noun used as a modifier, e.g. **Household pets**; **Electron microscopes**.

e) Proper noun used as a modifier, e.g. **Bernstein polynomials**; **Halley's comet**.

- **Conjunctive phrase headings**

Conjunctive phrase headings refer to those subject headings which are composed of two or more nouns, with or without modifiers, connected by '*and'* or ending with '*etc'*. Headings of this type are established only when a work being catalogued discusses a relationship between two topics from both perspectives and in such broad terms that the relationship could not be described by use of a main heading with a subdivision. As for examples: **Wit and humor**; **Libraries and Society**, **Chapters, cathedral, collegiate, etc.**

- **Prepositional phrase headings**

Prepositional phrase headings are used to enable the subject cataloguer to express single but complex ideas for which there is no one word, e.g. **Photography in Psychiatry**; **Radioisotopes in Cardiology**.

- **Inverted Headings**

Inverted Headings serve the alphabetico-classed function of subordinating specific descriptors under their broad generic categories, e.g. **Proverb, Korean**, **Education, Elementary**. Names of geographic features have traditionally been inverted in order to place a significant word in the initial position instead of the generic word. As for example: "Lake Eric" is formulated as **Eric, Lake** so that the distinguishing part of the name, "Eric" appears first.

- **Proper name headings**

Any proper name can be used as a subject heading when that proper name becomes the subject of a work. Names are not, however, all included in the LCSH. The only ones appearing in the LCSH are those that are used as pattern headings, or as examples, or that need special subject subdivisions or instructions printed under them. For examples: **Bermuda Islands**, **Shakespeare, William, 1564-1616**, **Canadian Psychological Association**.

10.8.2 Subdivisions

LCSH requires extensive use of subdivisions as a means of combining a number of different concepts into single subject heading. Subject subdivisions are words or phrases that can follow a subject heading, and are listed below the subject heading and indicated by a dash. These words or phrases are used to make the subject more specific. Complex topics are represented by subject headings followed by subdivisions. Some subdivisions are printed in LCSH but greater number of subdivisions may be assigned according to the rules specified in the *Subject Headings Manual (2008).* Library of Congress has published a separate alphabetical listing of various subdivisions that can be used after any listed subject heading unless otherwise specified. Subdivisions fall into several broadly defined categories: topical and chronological subdivisions specific to particular headings, form subdivisions, geographic subdivisions, free-floating subdivisions, and subdivisions under pattern headings. Valid subdivisions are marked in bold letters. Primary subdivisions are preceded by a single dash. Secondary subdivision—that is subdivisions of subdivisions are preceded with a double dash. Each category is described below:

- **Topical Subdivisions**

Topical subdivisions are used under the main headings or other subdivisions to limit the scope of the concept expressed by the heading to a special subtopic. Such headings are displayed by a long dash in LCSH: if two subdivisions are used, there are two long dashes. For example: **Women—Employment**; **Automobiles—Motors—Carburetors**. The rules for their application are found in the *Subject Headings Manual (2008)* and in general references printed under the generic headings in LCSH.

- **Chronological Subdivisions**

Chronological subdivisions are used to limit a heading, with or without subdivision, to a particular time period, e.g. **Philosophy, French—18th century**. Specific topical subdivisions and chronological subdivisions printed under names of countries and other jurisdictions or regions may be used with them. The date subdivisions displayed under

United States—Economic conditions, **United States—History**, and **United States—Politics and governments** are illustrative.

- **Form Subdivisions**

Form subdivisions are used to indicate the form in which the material on a subject is organised and presented, and as such are added as the last element to any heading. Form and topical subdivisions are both included within a general subdivision category. As for examples: **Art–Bibliography–Periodicals** (a serially issued art bibliography); **Art–Periodicals–Bibliography** (a bibliography of art journals). For combinations of subdivisions not listed in *Subject Cataloguing Manual* or *Free-floating subdivisions*, other methods must be employed. In most subject headings, the form subdivision appears as the last element, following the general pattern of subdivision order, **Topic–Place–Time–Form**.

- **Geographic Subdivisions**

For geographic subdivisions, headings with the code *(May Subd Geog)* may have place names added which are not specifically listed in LCSH. In the LCSH, the heading **Labor supply** *(May Subd Geog)*" tells us that place name may be added with the main heading "Labor supply", e.g. **Labor supply—France**. If a heading contains both a geographic subdivisions and topical or form subdivisions, the location of the geographic subdivision depends on which elements can be subdivided by place, e.g. **Construction industry—Finance—Law and legislation—Italy**. Sometimes geographic headings are subdivided by topic, e.g. **Massachusetts—History**.

- **Free-floating Subdivisions**

There are a number of identical (or nearly so) subdivisions that can be applied in specifically defined situations to primary headings and which won't have separate authority records for each heading and subdivision combination. The five broad groups of free-floating subdivisions are: general, under classes of persons and ethnic groups, under names of corporate bodies, persons and families, under place names, and entries controlled by pattern headings. All free-floating subdivisions are listed in *Free-floating Subdivision: an Alphabetical Index*, published annually by Library of Congress (the latest one is 21st edition, 2009). It is a useful tool for libraries, indicating the various subdivisions that can be used after any listed subject heading unless otherwise specified.

- **Subdivisions Controlled by Pattern Headings**

Certain form or topical subdivisions are common in a particular subject field or applicable to headings in a particular category. Instead of authorising them heading by 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. The applicable subdivisions are displayed under the table of pattern headings in LCSH. The subdivisions listed under a pattern heading may be transferred and used with another heading in the same category even though the combination does not appear in LCSH. The table of pattern headings is arranged alphabetically by category of headings covered by the pattern headings. The specific headings under which the subdivisions will appear in LCSH are listed in the right-hand column. For example, a set of subdivisions has been developed under **Cancer** and **Tuberculosis**, the pattern headings for the category **Diseases**. The subdivisions established under either of these pattern headings may be used as subdivisions under any heading belonging to the category **Diseases** if it is appropriate.

10.8.3 Entry Structure

All preferred primary subject terms and subjects with subdivisions are listed in boldface type, while those in the entry vocabulary only (i.e. references interfiled with the main headings), appear in normal typeface. Each entry in the LCSH is accompanied by all or some of the following:

- **Headings**

Subject headings displayed in the LCSH consist of single word or multiple word headings as discussed under the Section **10.8.1** of this Unit.

- **Class Numbers**

Library of Congress (LC) classification numbers are included in LCSH for many primary subject terms and some subject subdivisions. Class numbers are added only when there is a class correspondence between the subject heading and the provisions of the LC classification schedules, e.g.

Computer software
[QA76.755]

- **Scope Notes**

This is a note under the subject heading that explains and clarifies what is meant and what is not meant in the definition of the term and in its use as a subject heading. It specifies the range of application for a heading or draw distinctions between related terms. Scope notes may follow in separate paragraphs.

- **Reference Structure**

References associated with the headings are then listed in groups under the following codes:

USE = Use this heading instead of another

UF = Used For (do not use this heading)

BT = Broader Topic (less specific)

NT = Narrower Topic (more specific)

RT = ‘Related Topic (similar)

SA = See Also (additional terms)

Equivalence relationship: USE References

The code UF denotes non preferred headings for the given term/phrase. The codes *USE* and *UF* function as reciprocal in the LCSH. As for examples:

Cars (Automobiles)	**Automobiles**
USE **Automobiles**	UF Cars (Automobiles)

In the catalogue, the *USE* and *UF* can be translated into “*see*” references.

Hierarchical Relationships: BT and NT References

Subject headings are linked to other subject headings through cross-references now expressed as Broader Topic (BT) and Narrower Topic (NT). Headings connected by the BT and NT references are all valid headings and the connections use reciprocals. A heading appearing as a BT must be matched by the reversed relationship as an NT, as demonstrated by the following example:

Vehicles	**Transportation**	**Motor vehicles**
BT Transportation NT Motor vehicles	NT Vehicles	BT Vehicles

The BT and NT can be translated into *"see also form"* and *"see also"* references respectively in the catalogue, as demonstrated below by taking the above example:

Transportation *see also* Vehicles	Vehicles *see also* Motor Vehicles

Associative Relationship: RT References

The associative relationship, expressed by the code RT (Related Term), links two headings that are associated in some manner other than hierarchical or equivalence sense. RT references are provided under both terms in a reciprocal way. For examples,

Ornithology	**Birds**
RT Birds	RT Ornithology

General References

A general reference, represented by the code "SA" (see also), is a reference made not to specific individual headings but to an entire group of headings. The code "SA" is translated into "see also" reference in the catalogue. For example,

Wood working industries

SA *names of specific industries, e.g. Furniture industry and trade*

It is expected that each library will make specific references to each individual industry about which the library has documents in its collection.

10.8.4 Filing Order

LCSH's basic arrangement of subject headings is alphabetical, word by word. Numbers given in digits, both Arabic and Roman, precede alphabetic characters in the order of increasing numeric value. Initials separated by punctuation file as separate words. Abbreviations, acronyms, and initials without interior punctuation file as single whole words:

4 - H clubs

35mm cameras

A-5 rocket

A3D bomber

ACI test

Alaska—History

LC subject headings which contain subordinate elements preceded by one or more dashes get arranged as follows:

United States - Foreign relations – To 1945

United States - Foreign relations - 1945

United States - Foreign relations - Executive agreements

United States - Foreign relations - Iran

United States - Foreign relations - Latin America

All subdivisions identified by dashes, file ahead of inverted qualifiers, which are punctuated by commas. Inverted qualifiers, in turn, file ahead of qualifiers in parentheses. Last of all come phrases which start with the primary term. As for examples:

Children

Children—Attitudes

Children—Growth

Children, Adopted

Children, Maori

Children (International law)

Children (Roman law)

Children and Animal

Children in Literature

Children of Working Mothers

10.8.5 Merits and Deficiencies

The strongest aspect of LCSH is that it represents subject headings of a national library, one of the richest national libraries of the world. The administrative and managerial machinery LC has, made it possible for LCSH stand out as an undisputed leader almost on all aspects of library development. In addition to the printed version, LCSH is now available on different formats: microfiche, revised quarterly; CD-ROM, revised quarterly; and online through the LC Internet node. A number of bibliographic utilities allow searching by LCSH. The prominent examples are the Research Libraries Information Network (RLIN) and Washington Library Network's own online catalogue wit access by subject, often with sophisticated search options; most of these libraries use LCSH for subject cataloguing.

With all these strengths and reputation, the Library of Congress has yet to publish a statement of principles for its subject cataloguing system. Without a proper theoretical foundation, it has become a mammoth tool rather becoming difficult to use. The principles that underline the system have to be inverted form practice and policy statements.

Very often LCSH is criticized on grounds of its outdated terminology, illogical syntax and general inefficiency for precise subject retrieval. Its limitations are that it would meet the subject cataloguing requirements only for macro documents. Its use as a bibliographic list, therefore, cannot be extended to index micro documents.

Conscious of all these drawbacks, more intensive efforts are underway at LC to modernise its subject heading system, and systematize the existing tool. Many of its new features recently introduced represent steps in this direction. For the foreseeable future, LCSH gives every sign of retaining its vitality and prominence for subject access to library collection.

Self Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

7) What do you mean by homograph? How are the homographs handled in the LCSH?

....

....

....

....

8) What are the different types of headings that are presented in the LCSH?

....

....

....

....

9) What are the functions of a Pattern Heading in LCSH?

....

....

....

....

10.9 SEARS LIST OF SUBJECT HEADINGS

The *Sears List of Subject Headings (SLSH),* now in its nineteenth edition (2007), is an abridged version of the Library of Congress Subject Headings, named after the first compiler Minnie Earl Sears (1873–1933). It was in response to demand from small and medium-sized libraries for a list of subject headings that Sears compiled. Each edition strives to keep in mind current needs and users' approaches for subject access and the *SLSH* is still a living manifestation of the principles of subject cataloguing put forth by C. A. Cutter (1837–1903). The reforms of the SLSH began in its fifteenth edition with the display of headings and their relationships using standard abbreviations, namely NT, BT, RT, UF, USE, and SA, instead of the confusing x, xx, and "see also" for codified instructions. The thesaurus format conforms to the ANSI/NISO Standards for Thesaurus (1993). Nevertheless, the *Sears List* remains an index for pre-coordinated indexing instead of a true thesaurus. Its terminology is modern, although with an American bias. It is widely used today in and outside USA by small and medium-sized libraries.

Like LCSH, subject headings in the SLSH consist of a variety of forms, ranging from a single word to different kinds of complex descriptive phrases of single concepts in two or more words to complex and compound subjects. SLSH is guided by the general principles such as "Specific and Direct Entry", "Common Usage", "Uniformity" in the formation of its subject hidings as discussed under the Section 10.7.2.1 of this unit. There are four types of Subject headings in SLSH: topical headings, form headings, geographic headings and proper names.

10.9.1 Types of Subject Headings

- **Topical Headings**

Topical subject headings refer to the words or phrases for common things concepts that represent the content of the document.

Single word heading

The single noun is the most desirable form of subject heading if it is specific enough to fit the document in hand and the needs of the library; e.g. **Mythology** or **Neopaganism**, etc. If there is a significant difference between the singular and plural forms, the plural is to be preferred (e.g. "Mouse USE **Mice**"). Sometimes both the singular and plural forms of a noun are used as headings. In such cases, they carry different meaning, e.g. **Essay** (refers to the appreciation as well as technique of writing essays) or **Essays** (refers to the collections of literary essays).

Like LCSH, homographs are resolved by providing a noun or a phrase as parenthetical qualifier with the primary term; e.g. **Seals (Animals)** / **Seals (Numismatics)**. Parenthetical qualifiers are also used as explanatory modifiers used to provide general contextual information, e.g., **Hotline (Telephone counseling)** or **Depression (Psychology)**, etc.

Phrase Heading

A phrase heading consists of two or more words representing the subject of a document. Like LCSH, adjectival phrase headings (e.g. **Computer animation** or **Open access publishing**), conjunctive phrase headings (e.g. **Gods and godesses** or **Anarchism and anarchists**), and prepositional phrase headings (**Animals in art** or **Prisoners of war**) have been used in SLSH. Further phrase heading may be serial (e.g. **"Stories, plot, etc."**), or an intriguing combination of forms (e.g. **Superhero comic books, strips, etc.** or **Life support systems (Medical environment)**.

Form Headings

Form refers to the form of exposition of the content of the document, such as **Almanacs**, **Dictionaries**, **Directories**, **Encyclopaedias**, **Gazetteers,** etc. These headings are assigned to individual document as form subdivisions. Further, any form can also be a topic since it is evident that there are many documents written on dictionaries, encyclopedias, etc. In such a situation, these form headings are used as topical headings. Other form headings are the names of literary forms (e.g. **Fiction**, **Poetry**, **Drama**, **Essays**, etc.) and minor literary forms known as *genres* (e.g. **Science fiction**, **Children's plays**, etc.). Form headings and topical headings in literature are sometimes made by using singular form for topical headings (e.g. **Short story**) and plural form for form headings (e.g. **Short stories**).

Geographic Headings

SLSH prescribes to the documents on geographic areas, countries, cities, etc. under the names of the places. The SLSH does not provide the full list of places, the geographical headings **United States, Ohio** and **Chicago (III)** are provided only as examples. The geographic headings and geographic subdivisions used in SLSH follow the form of abbreviations for qualifying states, provinces, etc. found in Appendix B (Abbreviations) of AACR 2.

Proper Names

Three major types of name headings—personal names, corporate names, and uniform titles are provided in the SLSH. Inverted form is used in case of personal name headings with *See* references from alternate forms, e.g. the heading **Clinton, Bill** would call for a *See* reference from "Clinton, William Jefferson". Names of corporate bodies are entered directly under the corporate name heading as a subject, e.g. **Rockefeller Centre**. Names of sacred scriptures, anonymous literary works, periodicals, motion pictures, radio and television programmes are entered directly under the uniform title as a subject, e.g. **Gone with the wind (Motion picture)** or **Beowulf**. Documents about a literary work with a known author are entered under a name-title heading consisting of the author's name followed by the title, e.g. **Shakespeare, William, 1564-1616. Hamlet.** Name headings are numerous and as such they are not provided in the SLSH. Name headings are to be established by the cataloguer by consulting the sources as suggested in the SLSH.

10.9.2 Subdivisions

The purpose of subject subdivisions is to subdivide topics which are broad in scope or have much written about them. Subdivisions provided by the SLSH indicate a specialised aspect of a broad subject or point of view. There are different types of subdivisions, including topical subdivisions, geographic subdivisions, chronological subdivisions, and form subdivisions. Subdivisions are compounded under a given topic. For examples of the different types of subdivisions [form, key (like LCSH pattern headings), special topic, time, geographic] refer to the heading "**United States—History—1861-1865, Civil War—Medical care"**. The subdivisions in SLSH usually follow the standard order of **[Topical]—[Geographic]—[Chronological]—[Form].**

- **Form subdivisions** are used to indicate the physical form of the exposition of the document, e.g. **...—Bibliography**.

- **Topical subdivisions** provide the aspect of a subject or point of view presented in a particular document, e.g. **Oceanography—Research**.

- **Key headings** in SLSH, like LCSH, serve as patterns for the subdivisions and may be used under all headings of the same type. Some of these *'key headings'* are: **Shakespeare, William, 1564-1616** (to illustrate subdivisions under any author); **President—United States** (to illustrate subdivisions that may be used under presidents, prime ministers, governors, and rulers of any country, state, etc.); **English language** and **English literature** (to illustrate subdivisions that may be used with any language and literature respectively).

- **Chronological subdivisions**, i.e. time divisions, which apply most frequently to history, define a specific chronology for the primary topic, e.g. **Europe—History—1789-1900**. If a chronological period has a given name, this name is included in the heading following the dates, e.g. **United States—History—1600-1775, Colonial period**.

- **Geographic subdivisions** provide geographical aspect of a broad subject. Two forms of geographic divisions are evident in the SLSH: (a) Area—Subject, e.g. **Chicago (III)—Census** and (b) Subject—Area, e.g. **Geology—Bolivia**. Parenthetical instructions are provided with those headings which may be subdivided by place, e.g. **Folklore** (*May subdiv. Geog.*).

10.9.3 References

References in the SLSH are grouped into three main categories: specific See references, specific *See also* references, and general references. With the fifteenth edition of SLSH in 1994 *See* and "*See also*" references and the complementary "x" and "xx" were replaced by the thesaurus symbols UF/USE, BT, NT, RT and SA as explained under the Section 10.8.3 of this Unit.

- **Specific *"See"* References**

The UF symbol stands for *"Used for"*, and it designates those preferred terms or phrases form which *"see"* references are to be made. The most frequent and helpful varieties of *See* references direct the users from:

a) Synonyms or near synonyms, e.g. Chemical geology *USE* **Geochemistry**.

b) The second part of a compound heading, e.g. Motels *USE* **Hotels and motels**.

c) Conjunctive i.e. terms connected by "and", e.g. Religion and Science *USE* **Science and religion**.

d) Variant spellings or initialism to the accepted spelling or full form, e.g. Gipsies *USE* **Gypsies** and ESP *USE* **Extrasensory perception**.

e) Opposites when they are included without being specifically mentioned, e.g. Disobedience *USE* **Obedience**.

f) Singular to plural when the two forms would not file together, e.g. Goose *USE* **Geese**.

- **Specific *"See also"* References**

A *"See also"* reference connects a heading to a related heading or headings designated in the SLSH as *NT* (Narrower term(s)) and *RT* (Related term(s)) if the library has material listed under both headings and it normally move downward from a general term to a more specific term(s), e.g. **Indoor gardening** *See also* **Terrariums; Window gardening**. In this example the general term refers to two more specific terms. The user who pursues the index file by looking under **Window gardening** will find a still more specific downward reference like **Window gardening** *See also* **House plants**. It appears that *See also* references follow the rule that references are made from a general headings to one more specific. Yet there are a high number of bilateral or reciprocal references, where the movement is horizontal (between related subjects) rather than vertically (towards a more specific term from a more general one). These terms are designated as *RT* (Related term(s)) in the SLSH and the *See also* references between related terms are reciprocal in the catalogue. Under the heading **Window gardening**, the users will find **Container gardening** and **Flower gardening** as its RTs, and reciprocal See *also* reference entries are evident in the following examples: **Window gardening** *See also* **Container gardening**; **Container gardening** *See also* **Window gardening**; **Window gardening** *See also* **Flower gardening**; **Flower gardening** *See also* **Window gardening**.

- **General References**

The symbol "*SA*" stands for *"See also"* and introduces a general reference, which refers for one heading to an entire category or class of headings rather than an individual heading. A general explanation or direction is given instead of a long list of individual headings. For example,

Furniture

SA furniture of particular countries. e.g. **American furniture**; types of furniture, and names of specific articles of furniture, e.g. **Tables**; **Chairs**; etc. [to be added]

The results of these general references are so diverse and numerous that it is suggested that specific references be made from a broader term to those narrower term(s) on which the library acquires the materials. If the library acquires materials on Book cases, Chairs, Desks and tables, the reference entries are to be made as follows:

Furniture

See also **Book cases**
Chairs
Desks
Tables

10.9.4 Filing Order

Filing of entries of SLSH follows the ALA Filing Rules (1980). All headings are interfiled disregarding all punctuations. Headings with modifiers given in round brackets and phrase headings appear intermingled with those with dashes and commas. Although this straightforward alphabetical arrangement may pear to be less logical than the one use with punctuations, but the increasing use of computers in libraries and the wish to adhere to standard rules have prompted the practice to adhere to standard rules which have prompted the editors of SLSH to adopt the ALS Rules for Filing.

10.9.5 Merits and Deficiencies

SLSH is comparatively simpler to use than the LCSH. The rules and principles are fairly explicit in their directions, containing scope notes and specific instructions for their use. If followed consistently, SLSH will provide useful reference guide for the user and also to the library reference staff.

The major inadequacies in SLSH as well as LCSH are that these subject heading lists are not backed up by theoretical foundations intrinsic to subject indexing. With the growth of knowledge throwing forth new subjects and intricate facet and phase relations, these enumerated subject heading lists are not equipped to respond to new challenges in subject cataloguing.

Self Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

10) What are the guiding principles that the SLSH follows.

....

....

....

....

....

11) Discuss the different types of *"See also"* cross-references in the Sears List of Subject Headings.

....

....

....

....

12) Describe the filing order of subject headings in the SLSH.

....

....

....

....

13) Prepare the *"See"* and *"See also"* cross-references from the following entry:

Fatigue
UF Exhaustion
Weariness
BT Physiology
NT Jet lag
RT Rest

10.10 SUMMARY

Indexing languages are controlled languages used to facilitate access to information. The present Unit discusses various aspects of an indexing language in holistic manner. The meaning, scope, structure and attributes of an indexing language and how it differs from the natural language have been discussed at the outset. It is followed by the discussion on meaning, need, objectives and method of vocabulary control. Types of indexing languages are discussed. The subject headings list has been explained in the context of the general principles associated with the choice and rendering of subject headings. Subject authority file has also been considered in the light of use and maintenance of the subject heading list. Thesaurus has been discussed with particular reference to its types, functions, structure, nature of relationships between terms, and the important role they play in indexing and searching. Salient features of two extended forms of thesaurus— Thesaurofacet and Classaurus are explained. This unit also looks at the Library of Congress Subject Headings (LCSH) and Sears' List of subject headings (SLSH) in greater detail.

10.11 ANSWERS TO SELF CHECK EXERCISES

1) An indexing language (IL) is a system for naming subjects of the records of information (i.e. document). The language we use in our day to day life for communication of ideas is natural language. A natural language is "natural" as it grows freely in the lips of human being, and is totally free from any control. An indexing language is "artificial" in the sense that its vocabulary, syntax, semantics and orthography differ from the natural language. Vocabulary of a natural language is free language and it has no control over synonyms and homographs. But an indexing language is a controlled language, synonyms and homographs are

controlled there because of the standardisation of terms. Natural language provides auxiliaries like prepositions, conjunctions, etc. to bring out the correct meaning of the sentence. But such auxiliaries are absent in an indexing language. The order of terms in an indexing language is governed by its syntactical rules along with the relational symbols like role operators or indicator digits.

2) Components of an indexing language consist of controlled vocabulary, syntax and semantics. The vocabulary of an indexing language is controlled for standardisation of terms by means of controlling synonyms, near-synonyms and word forms, and by distinguishing among homographs. The order as well as correct relationship between terms in a subject heading is governed by the rules of syntax of a given indexing language. Semantics is used to resolve the meaning of the term in a subject heading through hierarchical, equivalence and associative relationships.

3) The attributes of an indexing language consist of vocabulary control, concept coordination, multiple access, synthetic devices, relation manifestation, and structural presentation. They play very important roles in the successful organization of the index file and subsequent matching of the index and queries of the users

4) Paradigmatic relationship is the semantic or generic relationship and this relationship finds expression in the hierarchical relationship as displayed in the vocabulary of a given indexing language. Syntagmatic relations are based on the syntactical rules of a given indexing language. A paradigmatic relationship is document independent relationships because these relationships are established without any reference to a document. Syntagmatic relationship is document dependent relationship because this relationship is established with reference to the relationship between concepts associated with the content of a given document.

5) Controlled vocabulary refers to an authorised and structured list of terms showing their interrelationships and indicating ways in which they may usefully be combined to represent specific subject of a document.

6) There are basically two fold objectives of vocabulary control: (1) To promote the consistent representation of the subject matter of documents by indexers and searchers, thereby avoiding the dispersion of related documents; and (2) To facilitate the conduct of a comprehensive search by bringing together in some way the terms that are most closely related semantically. The first objective is achieved through control of word forms, synonymous and nearly synonymous expressions by choosing one of the possible alternatives as the "Preferred term" and referring to this term (by using "See" or "Use") from the variants under which certain users may be likely to approach. Homographs are distinguished to resolve ambiguity in the use of terms by means of a parenthetical qualifier or scope note. The second objective is achieved by linking together terms that are hierarchically related in order to facilitate the conduct of comprehensive searches. A controlled vocabulary brings together terms that are semantically related through hierarchical and associative relationships. This objective is achieved by means of "See also" references among terms that are most closely related semantically.

7) Different types of indexing languages are classification schemes, subject heading lists, and thesauri. Thesaurofacet and Classaurus are the extended forms of a thesaurus.

8) The principles that guide the indexers in the choice and rendering of subject headings from the Standard List are: (a) **Specificity**: A work (document) should be entered under the most specific subject heading that accurately and precisely represents

its content; (b) **Common Usage**: The heading chosen to express a subject must represent common usage of the class of readers for whom the material on the subject within which the heading falls is intended; (c) **Uniformity**: One uniform heading must be selected from several synonyms, and this heading must be applied consistently to all works on the topic; (d) **Consistent and current terminology**: Term(s) chosen as subject headings should be both consistent and current as has already been said regarding the justifications for uniform headings; (e) **Form heading**: Form heading should be chosen for the work representing its physical, bibliographic, artistic or literary form and for the work about literary form that requires subject heading; and (f) **Cross References**: Adequate references should be made to direct the user from unused heading to used headings, and from headings referring broader and related topics to the headings chosen to represent a given subject.

9) Subject authority file is an important means of maintaining control over the vocabulary and references. It serves as the source of indexing vocabulary and as the means of verifying or validating heading assigned to individual cataloguing records. By consulting the subject authority file, the cataloguer can make necessary adjustments when existing headings are revised or new headings are added. The subject authority file is particularly useful in ensuring uniformity and consistency in subject heading terminology and cross-references.

10) A thesaurus is a controlled and dynamic vocabulary of semantically and generically related terms used in translating from the natural language of documents, indexers or users into a more constrained 'system language'.

11) The functions of a thesaurus may be summarized as follows: (a) To provide standard vocabulary for a given subject area by exercising control on the vocabulary of terms used in an indexing language; (b) To provide an aid to indexers in selecting term(s) for describing the subject matter of documents; (c) To help users to formulate their queries precisely; (d) To provide display of-relationships between terms to facilitate the conduct of a comprehensive search systematically; and (e) To provide a system of references between terms which will ensure that only one term from a set of synonyms is to be used both for indexing and searching.

12) Basically two types of relationships are displayed in a thesaurus: hierarchical and-non-hierarchical. Hierarchical relationship indicates the interrelationship between concepts in a hierarchy. It expresses degrees or levels of super ordination and subordination of concepts. This relationship may be of four types: Genus-species, hierarchical whole-part, instance and polyhierarchical relationships. Non-hierarchical relationship may be grouped as - equivalence relationship and associative relationship.

13) Lexical variants are different word forms for the same expression. Synonym refers to a word or expression that has the same or nearly the same meaning as another in the given language. Synonymy occurs when a concept can be represented by multiple terms having the same or similar meanings. Lexical variants are derived from spelling or grammatical variation or from abbreviated formats.

14) Thesaurofacet is an extended form of a thesaurus and is used as the vocabulary control device to organise documents belonging to the libraries of the corporation of English Electric. It is considered as a multi-purpose retrieval language tool because it has classification schedules and a faceted thesaurus. It has two parts: the first part is the faceted classification system and the second part consists of a thesaurus. The thesaurus part serves the alphabetized index to the classification

scheme. Because the two parts are linked, each term in the system appears twice: once in the schedule, and once in the thesaurus, with notation that links the two parts together. Thesaurofacet covers the whole field of science and technology with varying depth and only engineering and allied fields are covered exhaustively. Terms relating to the relevant Management-related concepts, borrowed from the "Classification of Business Studies" developed by London Graduate School of Business Studies, are included in the Thesaurofacet.

15) RTs are not displayed in the Classaurus because of its category-based structure. It is assumed that RTs should not be dictated by the designer of the Classaurus, rather it should be dictated by the document itself. Any term may be related to any other terms depending upon the nature of the thought-content of the document. Hence, RTs should not be determined beforehand.

16) A homograph is a word that is spelt like another word but has a different meaning from it. The LCSH distinguishes among homographs to avoid ambiguity by providing a noun or a phrase as parenthetical qualifier with the primary term. Thus the subject heading "Seals (animals)" in the LCSH tells us that this heading is to be used exclusively for a type of animals and not as numismatics or any other possible context.

17) Subject headings in the LCSH are constructed in variety of ways, ranging from single noun to complete descriptive phrases. Different types of headings as presented in the LCSH are: single word headings, adjectival phrase headings, conjunctive phrase headings, prepositional phrase headings, inverted headings, and proper name headings. LCSH has also made provisions for the extensive use of subdivisions as a means of combining a number of different concepts into single subject heading.

18) A pattern heading is a representative heading from a category of terms that would normally be excluded from the LCSH (e.g. names of individuals), included as an example of normal subdivision practice within that category. Certain form or topical subdivisions are common in a particular subject field or applicable to headings in a particular category. Instead of authorizing them heading by heading within the category, they are listed under a chosen or key heading in the category in a separate table with instruction to the cataloguer on how to construct other headings of similar form that have been omitted from the LCSH for reasons of brevity. This key heading then serves as a "pattern heading" of subdivisions for headings in that category. The subdivisions listed under a pattern heading may be transferred and used with another heading in the same category even though the combination does not appear in LCSH.

19) SLSH is guided by the general principles such as "Specific and Direct Entry", "Common Usage", "Uniformity" in the formation of its subject hidings

20) There are two types of "see also" cross-references in the SLSH: Specific and General. A specific "see also" reference is made from a broader term to a narrower term, but not from a narrower term to a broader term. The specific "see also" references are also made between the term following RT symbol and the subject heading and vice versa resulting non-hierarchical two-way references between related terms. The symbol "SA" stands for "see also" and introduces a general reference, which refers from one heading to an entire category or class of headings rather than individual heading.

21) The filing order in Sears List is according to the ALA Rules for Filing. In this system, the punctuation marks are disregarded in filing and all words filed without

any difference between sub-headings, inverted headings with commas and such others.

22)

Exhaustion *See* Fatigue	Weariness *See* Fatigue	Physiology *See also* Fatigue
Fatigue *See also* Jet lag	Fatigue *See also* Rest	Rest *See also* Fatigue

10.12 KEYWORDS

Associative Relationship : A relationship between or among terms in a controlled vocabulary that leads from one term to other terms that are related to or associated with it; begins with the words SEE ALSO or related term (RT).

Broader Term : A term to which another term or multiple terms are subordinate in a hierarchy. In thesauri, the relationship indicator for this type of term is BT.

Classaurus : A category-based (faceted) systematic scheme of hierarchical classification in verbal plane incorporating all the essential features of a conventional IR thesaurus.

Controlled Vocabulary : An authority list of terms showing their interrelationships and indicating ways in which they may usefully be combined to represent specific subject of a document.

Cross Reference : A mechanism available in the indexing language for directing the user from term/phrase not used as headings to the term/phrase that is used, and from broader and related topics to the one chosen to represent a given subject. Two types of cross-references are used in the indexing language: *See* and *See also*.

Descriptor : See *Preferred term*

Entry Term : The non-preferred term in a cross reference that leads to a term in a controlled vocabulary. Also known as "lead-in term." In thesauri, the relationship indicator for this type of term is U (USE); its reciprocal is UF (USED FOR). See also *preferred term.*

Entry Vocabulary : The set of non-preferred terms (USE references) which lead to the preferred terms in a controlled vocabulary.

Equivalence relationship : A relationship between or among terms in a controlled vocabulary that leads to one or more terms that are to be used instead of the term from which the cross-reference is made; begins with the word SEE or USE.

Free-floating Subdivision : A term that can be added to a subject heading in a published list, as needed, whether or not it is written in the published list following that heading. In the LCSH, terms called 'free-floating' have scope notes that express limitations on the use of such terms.

Hierarchical Relationship : A relationship between or among terms in a controlled vocabulary that depicts broader (generic) to narrower (specific) or whole-part relationships; begins with the words broader term (BT), or narrower term (NT).

Homograph : One of two or more words that have the same spelling, but different meanings and origins. In controlled vocabularies, homographs are generally distinguished by *qualifiers*.

Indexing Language : An indexing language is a set of codes and their admissible expression used for representing the content of the documents as well as queries of the users.

Indexing Term : The representation of a concept in an indexing language, generally in the form of a noun or noun phrase. Terms, subject headings, and heading-subheading combinations are examples of indexing terms. Also called descriptor.

Lexical Variant : Lexical variants are different word forms for the same expression resulting from spelling variants, grammatical variation and variation in abbreviated formats.

Narrower Term : A term that is subordinate to another term or to multiple terms in a hierarchy. In thesauri, the relationship indicator for this type of term is NT.

Natural Language : A language used by human beings for verbal communication.

Near-Synonym : A term whose meaning is not exactly synonymous with that of another term, yet which may nevertheless be treated as its equivalent in a controlled vocabulary.

Pattern Heading : A pattern heading is a representative heading from a category of terms that would normally be excluded from the LCSH (e.g. names of individuals), included as an example of normal subdivision practice within that category.

Paradigmatic Relation : A paradigmatic relation, also called semantic or generic relation, is document independent relationship which usually finds expression in the hierarchical organisation of the vocabulary itself. This relationship is document independent relationship because it is established without any reference to a document.

Preferred Term : One of two or more synonyms or lexical variants selected as a term for inclusion in a controlled vocabulary.

Reciprocity : Semantic relationships in controlled vocabularies must be reciprocal, that is each relationship from one term to another must also be represented by a reciprocal relationship in the other direction. Reciprocal relationships may be symmetric, e.g. RT / RT, or asymmetric e.g. BT / NT.

Related Term : A term that is associatively but not hierarchically linked to another term in a controlled vocabulary. In thesauri, the relationship indicator for this type of term is RT.

Relationship Indicator : A word, phrase, abbreviation, or symbol used in thesauri to identify a semantic relationship between terms. Examples of relationship indicators are UF (USED FOR), and RT (related term).

Scope Note : A note following a term explaining its coverage, specialised usage, or rules for assigning it.

Semantics : Semantics is a study of meaning. In an indexing language, semantic relationship refers to the hierarchical and non-hierarchical relationships between the subjects and is governed by *see* and *sees also* references in an index file. Controlled vocabulary serves as the source for *see* and *see also* references.

Subject Heading : A word or group of words (phrase) indicating a subject under which all materials dealing with same theme is entered in a catalogue or bibliography, or is arranged in a file.

Subject Heading List : A subject heading list is an alphabetical list terms and phrases, with appropriate cross references and notes, that can be used as a source of subject headings in order to represent the subject content of a document.

Subject proposition : Name of a subject, i.e. subject heading

Syndetic Device : An organisational framework in which related names, topics, etc. are linked to each other by connective terms Like '*See*' and '*See also*'.

Synonym : A word or term having exactly or very nearly the same meaning as another word or term.

Syntagmatic relation : Syntagmatic relationship, also called syntactical relationship, refers to those relationships which are achieved through the syntactic devices—

word or term order and relators or linking mechanism used in the given indexing language. This relationship is document dependent relationship because it is established with reference to the concepts associated with the content of a given document.

Syntax : The grammatical structure consisting of a set of rules that govern the sequence of occurrence the terms in a subject heading.

Thesaurus : A thesaurus is a controlled and dynamic vocabulary of semantically related terms used in translating from the natural language of documents, indexers or users into an indexing language.

Vocabulary Control : The process of organising a list of terms (a) to indicate which of two or more synonymous terms is authorised for use; (b) to distinguish between homographs; and (c) to indicate hierarchical and associative relationships among terms in the context of a controlled vocabulary or subject heading list. See also *controlled vocabulary*.

10.12 REFERENCES AND FURTHER READING

Bhattacharyya, G. "Classaurus: Its Fundamental, Design and Use". *Studien zur Klassification,* BD. 11; 1982; 139-148. Print.

Chakraborty, A. R. and Bhubaneswar Chakraborty. *Indexing: Principles, Processes and Products.* Calcutta: World Press, 1984. Print.

Foskett, A. C. *Subject Approach to Information*. 5th Ed. London: The Library Association, 1996. Print.

Ghosh, S. B. and Satpathi, J. N., *Eds. Subject Indexing Systems: Concepts, Methods and Techniques*. Calcutta: IASLIC, 1998. Print.

Guha, B. *Documentation and Information: Services, Techniques and Systems*. 2nd rev ed. Calcutta: World Press, 1983. Print.

Lancaster, F. W. *Indexing and Abstracting in Theory and Practice*. 3rd edition. London: Facet Publishing, 2003. Print.

Olson, Hope A. *Subject Analysis in Online Catalogs*. 2nd edition. Englewood: Libraries Unlimited, 2001. Print.

Sarkhel, Juran Krishna. *Information Analysis in Theory and Practice*. Kolkata: Classique Books, 2001. Print.

Sarkhel, Juran Krishna (1999). "Subject Indexing, Vocabulary Control and Recent Developments in Cataloguing". *Library Cataloguing Theory*. (BLIS 04)New Delhi, Indira Gandhi National Open University, 1999. 180p. Print.

Taylor, Arlene G. *Introduction to Cataloguing and Classification*. 10th edition. New Delhi: Atlantic Publishers, 2007. Print.

UNIT 11 INDEXING TECHNIQUES

Structure

11.0 OBJECTIVES

After reading this Unit, you will be able to:

- understand the basic principles of subject indexing techniques;
- appreciate the differences between (a) derived and assigned indexing, (b) pre- and post-coordinate indexing systems;

- trace the major contributions in pre-coordinate indexing systems;
- learn the different stages of intellectual operations and working of some important pre-coordinate indexing systems such as, Chain Indexing, PRECIS and POPSI;
- understand the objective conditions that led to the development of post-coordinate indexing;
- learn the different entry structure such as item entry as well as term entry systems and the methods of post-coordinate indexing with particular reference to Uniterm indexing system and post-coordinate searching devices;
- explain and apply the different varieties of keyword indexing;
- appreciate the differences between the manual indexing and computerised indexing;
- understand the concept of computerised indexing in terms of its features, components, categories, and index file organisation;
- learn different methods associated with the generation of index entries with the aid of computer;
- get acquainted with the indexing internet resources with particular reference to search engine indexing and other associated concepts; and
- develop skills of using different indexing techniques for formulating different types of subject headings.

11.1 INTRODUCTION

Subject approach to information has been an area of intense study and research in the area of organisation of information, resulting in the generation of new theories and the design of the corresponding new indexing techniques based on these theories. Indexing technique actually originated from what is known as the 'back-of-the-book index. Its objective is to show where exactly in the text of a document a particular concept (denoted by a term) is mentioned, referred to, defined or discussed. The 'back-of-the-book index' may be either in the form of specific index or relative index. Specific index presents the broad topics in the form of one-to-one-entry whereas the relative index is one which displays each concept in different context. The best example of such an index is the relative index of Dewey Decimal Classification. But the relative index is usually unique to the text to which it points to and is quite difficult to maintain on a large scale. Subsequently we have seen the development of pre-coordinate indexing model.

Till about the early fifties of the last century, the pre-coordinate indexing models were the only ones that had been developed. In the subsequent decades, the post-coordinate indexing models were designed and developed, the physical apparatus for these index files also changed from the conventional index cards to different formats of post-coordinate indexing models. With the advent of the computers, the keyword index models like KWIC, KWOC, and KWAC were introduced, post-coordinate indexing also became more amenable for computer manipulation. Most of the bibliographic databases today have indexes based on post-coordinate indexing principles. The following sections of this Unit present the major developments in subject indexing techniques for organising the index file.

11.2 DERIVATIVE INDEXING AND ASSIGNMENT INDEXING

Indexing can be either 'derived indexing'/'derivative indexing' or 'assigned indexing'/ 'assignment indexing'. Derived indexing is 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 the subject(s) of the work, for use as headings under which entries are made. It is also known as *extractive indexing*.

In derivative indexing, terms to be used to represent the content of the document are derived directly from the document itself. Here no attempt is made to use an indexer's own knowledge of the subject or other guides, but use only the information which is manifest in the document. Index terms are derived directly from the title or text of the document. It requires least intellectual effort on the part of the indexer. Mechanical devices and computers are used in abundance to carry out the burden of index preparation as well as the tasks of matching these with the questions of users put to the system. Examples of derivative indexing are keyword indexing, citation indexing, automatic indexing, etc.

Assigned indexing is also known as 'concept indexing', because it involves identifying concept(s) associated with the content of each document. It is 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. Here, an indexing language is designed and it is used for both indexing and searching. Some notable examples of assignment indexing are chain indexing, PRECIS, POPSI, classification schemes, etc.

11.3 PRE-COORDINATE INDEXING SYSTEMS

Basically all indexing systems are, by nature, coordinate indexing. The purpose of using a combination or coordination of component terms is to describe the contents of the documents more precisely. Many subjects can be expressed in a single term, e.g. Libraries, Cataloguing, Management, etc. Others are expressed as a combination of these, e.g. Cataloguing in libraries, Management of libraries, etc. When the indexer assigns subject headings representing such compounds and arranges entries in a series of classes according to the subject content of the document, the resulting system is referred to as *pre-coordinate indexing system.* Here, the indexer coordinates the component terms representing compounds at the input stage, i.e. at the time of indexing in anticipation of users' approach. Most of the classification schemes allow a measure of coordination, either by including compound subjects or by providing facilities for creating them out of simple elements. The same is true for readymade lists of subject headings like Sears List of Subject Headings and Library of Congress Subject Headings. Such classification schemes and lists of subject headings can therefore be regarded as pre-coordinate indexing languages. Almost all pre-coordinate indexing models have an *a priori* approach in choosing their semantic paradigms such as categorisation of concepts, role indicators, relational operators, etc. Pre-coordinate indexing involves coordinating (combining, pulling together concepts) followed by engaging in an act of *synthesis* to build the index entries. In such a system, the most important aspect is to determine the order of significance by following the syntactical rules of the given indexing language.

11.3.1 Cutter's Contribution

Charles Ammi Cutter first set forth the rules for alphabetical subject headings in a systematic way in his *Rules for a Printed Dictionary Catalog* in 1876. He used the term 'subject cataloguing' instead of 'subject indexing'. It was C. A. Cutter who gave the idea of specific subject entry. He also tried to systematise the rules for compound subject headings which consisted of more than one term as a phrase and where the subject heading was composed of a name of a subject and the name of a locality and so on. Cutter's concept of specific subject heading was quite different from what we mean today. He had in mind a set of stock subject names and every book had to be accommodated under the most restricted subject which contained the subject of a book. For compound subject heading, Cutter laid down the rule that the order of the component terms in compound subject heading should be the one that is decidedly more significant. But Cutter could not prescribe how one will come forward to decide which one is more significant. The question of significance varies from user to user. The decision in respect of 'significance' was left to the judgment of individual indexer, which was subjective. Some specific rules/guidelines as furnished by Cutter in his *Rules for a Printed Dictionary Catalog* are mentioned below:

1) **Person versus Country**

Entry will be under person in case of single / personal biography. Entry will be under country in case of history, event, etc. For example the biography of Sachin Tendulkar would entered under his name whereas the biography of Indira Gandhi would be entered her name and India- history also.

2) **Country versus Event**

Entry of an event would be under event if it is proper noun, e.g. Jalian Wala Bagh. Entry would be under country if it is common noun. e.g. Freedom fighters in India, entry would be under India.

3) **Subject versus Country**

In scientific subjects, entry will be under subject qualified by place. e.g. Oceanography, India.

It has been stipulated to prepare entry under the name of place qualified by the subject in subject areas like History, Government and Commerce, e.g. India—Moghul Period.

For Humanities, Literature, Art, etc., adjectival form of subject headings were suggested, e.g. Indian Painting, Moghul Architecture, etc.

4) Between overlapping subjects entry will be according to the importance of the subjects. It is to be pointed out here that the decision regarding 'importance' was left to the judgment of the indexer.

5) When there is a choice between different names, Cutter prescribed the followings:

   a) Language—If there are two languages out of which one is English, entry will be under English. Here the rules appear to be biased towards English language.

   b) Synonyms—Entry will be under one word with reference to others.

   c) Antonyms—Entry will be under one word with reference to others.

6) In case of compound subject headings, Cutter prescribed the following rules:

   a) A noun preceded by an adjective, e.g. Organic Chemistry, Ancient History, etc.

   b) A noun preceded by another noun used as an adjective, e.g. War Prisoners, Flower Fertilisation, Death Penalty, etc.

c) Use direct form in case of a noun connected to another noun with a preposition, e.g. Patient with heart disease, Fertilisation of flowers, death penalty, etc.

d) Use direct form in case of phrase or sentence used as the name of a subject, e.g. Medicine as profession.

11.3.2 Kaiser's Contribution

Kaiser started from the point where Cutter left. In 1911, Julius Otto Kaiser (1868–1927), a special librarian and indexer of technical literature, developed a method of subject indexing system known as *Systematic Indexin*g. Kaiser defined indexing as "the process by which our information is collected and made accessible" and claimed that it constitutes "the main work of organising information". The definition highlights a cardinal tenet of his theory of systematic indexing —that, within the framework of a business library with which he was attached, the primary aim should be not to index 'documents' but 'information', which Kaiser took to be the various 'facts and opinions' (i.e., informational units) encoded in the texts of documents. He viewed systematic indexing as a two-step procedure: The first step was to analyse a subject so as to distinguish constituent concepts associated with the content of the given document into two fundamental categories of facets: "Concretes", and "Processes". 'Concrete' refers to things, places and abstract terms, not signifying any action or process; e.g. Gold, India, Physics, etc. "Process" refers to mode of treatment of the subject by the author (e.g. *Evaluation* of IR system, *Critical analysis* of a drama), an action or process described in the document (e.g. *Indexing* of web documents), and an adjective related to the concrete as component of the subject (e.g. Strength of a metal).

Kaiser also laid a rule that if a subject dealing with place, double entry (*'Concrete—Place—Process'* and *'Place—Concrete—Process'*) is to be made. For example, index entries for 'Manufacturing of petrochemicals in West Bengal' would be: PETROCHEMICALS—*West Bengal—Manufacturing; WEST BENGAL—Petrochemicals— Manufacturing.* The second step was to synthesize constituent concepts into indexing statements formulated according to the strict rules of citation order—Concrete is to be followed by Process.

Thus, according to Kaiser's systemic indexing, all indexing terms should be divided into fundamental categories "concretes", "countries", and "processes", which are then to be synthesized into indexing "statements" formulated according to strict rules of citation order. Some examples of subject headings according to Kaiser's systematic indexing are furnished below:

Documents	**Categories**	**Subject Headings**
Indexing of films	Concrete—Process	Films—Indexing
Strike in India	Country—Process	India—Strike
Libraries in Nepal	Concrete—Country	Libraries—Nepal
Manufacturing of copper in Assam	Concrete—Country—Process	Copper—Manufacturing—Assam

Kaiser's systematic indexing did not make any provision for entry under the 'process' term and as a result it failed to satisfy the users' approach by the 'process' term. The concept of 'time' was also left by Kaiser. It is to be pointed out here that Kaiser was perhaps the first person who gave the idea of categorisation in subject indexing.

11.3.3 Chain Indexing

Ranganathan's facet analysis of subject provides a kind of representation of subjects by transforming multidimensional relations of subject into a modulated layer of linear representation. Ranganathan is credited with the invention of chain indexing, an economical system of providing access to the terms in classification schedules without replicating the hierarchical structure of the classification in the alphabetical index. Ranganathan's Chain indexing technique was devised as a complementary and supplementary tool to classification schemes. However, due to the efficiency and economy, this technique can effectively be made use of in deriving alphabetical subject indexes for any indexing/abstracting services.

Definition and Use

A chain is a string of terms organised in a particular sequence based on the classification scheme that the chain adopts. The sequence of terms is pre-coordinated according to the syntactical rules of the given classification scheme. Chain indexing is a method of deriving subject headings from the chain of successive subdivisions leading from general to specific level needed to be indexed.

According to Ranganathan, chain indexing is a

"*procedure for deriving class index entry (i.e. subject index entry) which refers from a class to its class number in a more or less mechanical way.*"

A note is also given with the above definition as

"*Chain procedure is used to derive class index entries in a Classified Catalogue, and specific subject entries, subject analytical, and 'see also' subject entries in a Dictionary Catalogue.*"

Chain indexing was used in *the British National Bibliography* (BNB) in the 1950s and 1960s until it was replaced by PRECIS-indexing.

Chains and Links

The concept of 'chain' is the foundation of chain indexing. A chain is deemed to be a structural manifestation of a subject. The term 'structure' in this context refers to the parts constituting a subject and their mutual interrelationship. It is a modulated sequence of sub-classes or isolates ideas.

Since the chain expressed the modulated sequence of sub-classes more effectively in a notational scheme of classification of subjects, this method takes the class number of the document concerned as the base for deriving subject headings not only for specific subject entry but also for subject reference entries. The nature and structure of the classification scheme used to classify the subject of the document controls the structure of the subject headings drawn according to the chain procedure. The concept of 'chain' becomes operative only after the concept of a set 'links' about the structure of the subject is conceded. A chain should comprise a link of every order that lies between the first link and last link of the chain. The different types of links in chain indexing system are discussed below:

- **Sought Links (SL):** Sought links denote the concepts (at any given stage of the chain) that the user is likely to use as access points.
- **Unsought Links (USL):** Unsought links denote those concepts that are not likely to be used as access points by the user.

- **False Links (FL):** False links are those that really do not represent any valid concept, mostly these are connecting symbols or indicator digits.
- **Missing Links (ML):** Missing links represent those concepts that are not available in the preferred classification scheme, these are inserted by the indexer by means of verbal extension at the chain-with-gap corresponding to the missing isolate in the chain whenever there is such a need.

Steps in Chain Indexing

The following steps are to be followed in chain indexing for deriving different types of subject headings:

1) **Construction of the Class Number of the Subject of the Document**

Classify the subject of the document by following a preferred classification scheme. A class number constructed according to a scheme for notational classification will form the basis for applying the rules for chain procedure for deriving subject headings.

2) **Representation of the Class Number in the form of a Chain**

Represent the class number in the form of a chain in which each link consists of two parts: class number and its verbal translation in standard term or phrase used in the preferred classification scheme.

3) **Determination of Links**

Determine different kinds of links: Sought Links (SL), Unsought Links (USL), False Links (FL), and Missing Links (ML).

4) **Preparation of Specific Subject Heading**

Derive specific subject heading for the specific subject entry from the last sought link and moving upwards by taking the necessary and sufficient sought links in a reverse rendering or backward rendering process. If the subject includes a space isolate, time isolate or a form isolate, break the chain into different part(s) at the point(s) denoting space, time and form in the class number. In such a situation, specific subject heading is to be derived from last SL of first part in reverse rendering process and then by second part, third part, etc., if any, in the similar process. Then, the components of derived subject headings are to be arranged in the sequence of their derivation from each part of the chain of the class number.

5) **Preparation of Subject Reference Headings**

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

6) **Preparation of Subject Reference Entries**

Prepare subject reference entries or 'see also' references from each subject reference heading to its specific subject heading. When a subject heading starts from last sought link denoting space or time or form, prepare 'See' references instead of 'See also' references from subject reference heading(s) to specific subject heading.

7) **Preparation of Cross References, if any**

Prepare cross references (i.e. 'see' references) for each alternative and synonymous term/heading used in the specific as well as subject reference headings.

8) **Alphabetisation**

Merge specific subject entries, subject references (i.e. 'see also' references) and 'see' references and arrange them in single alphabetical sequence.

The above noted steps in chain indexing are demonstrated below with illustrative example:

0) **Subject of the Document**

Researches on Child Psychology in India

1) **Class Number of the Subject of the Document**

Class no.: 155.4072054 [according to DDC, 22nd Edition]

2) **Representation of the Class Number in the form of a Chain**

100	Philosophy, Parapsychology and Occultism
150	Psychology
155	Differential and developmental psychology
155.	———
155.4	Child psychology
152.40	———
152.407	Education, Research related topics
152.4072	Research
152.40720	———
152.407205	Asia
152.4072054	India

3) **Determination of Links**

100	Philosophy, Parapsychology and Occultism [USL]
150	<u>Psychology</u> [SL]
155	Differential and developmental psychology [USL]
155.	[FL]
155.4	<u>Child psychology</u> [SL]
152.40	[FL]
152.407	Education, Research related topics [USL]
152.4072	<u>Research</u> [SL]
152.40720	[FL]
152.407205	Asia [USL]
152.4072054	<u>India</u> [SL]

4) **Preparation of Specific Subject Heading**

Research, Child psychology, India

5) **Preparation of Subject Reference Headings**

Research, Child psychology

Child psychology

Psychology

6) **Preparation of Cross References**

India, Research, Child psychology

7) **Preparation of Index Entries**

Research, Child psychology, India 152.4072054

Bibliographical description and abstracts of the document are to be furnished under the specific subject heading.

Research, Child psychology 152.4072

See also

Research, Child psychology, India

Child Psychology 155.4

See also

Research, Child psychology, India

Psychology 150

See also

Research, Child psychology, India

India, Research, Child psychology

See

Research, Child psychology, India

8) **Alphabetisation**

Arrange the above entries according to single alphabetical order.

Advantages

A major advantage of chain indexing is that it ensures the collocation of aspects of a subject which have been scattered in the classification scheme (i.e. distributed relatives) because last link in the class number is always the first link in the chain of subject index entries.

It offers general as well as specific information to all information seekers by deriving subject headings from the chain of successive subdivisions that leads from the general to most specific level.

It is more or less a mechanical system and it is economic also.

Criticisms

- Too much dependency on classification schemes

Chain indexing could be operative effectively only if they are backed up by a good structured classification scheme. If chain indexing is based on a structurally defective classification scheme, the subject headings drawn according to chain procedure will naturally become defective.

- Disappearance of Chain

Chain disappears in each stage of deriving subject reference entries and thus it results in the loss of full context of the content of the document.

- **Lack of Specificity**

Chain indexing provides only one specific entry and others are subject references.

- **Unsuitability for computerisation**

The formation of 'Chain' is very much a human intellectual operations which is beyond the capability of the computer to manipulate.

- **Running from pillar to post**

It makes direct access to specific subject heading possible only for those who knew the specific subject heading for the given document, but in most cases access to specific subject heading was possible at the cost of running from pillar to post since only one entry is specific subject entry and others are cross references in chain indexing system.

Self Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

1) Distinguish between Derived indexing and Assigned indexing.

....

....

....

....

2) Discuss the techniques by which J. Kaiser has solved the problems of indexing compound subject.

....

....

....

....

3) What is the epithet of the term 'chain' in chain indexing? Why is the class number of the subject taken as the base for deriving subject headings according to chain indexing system?

....

....

....

....

4) Enumerate the steps involved in chain indexing.

....

....

....

....

11.3.4 PRECIS (Preserved Context Index System)

PRECIS is generally considered as one of the most ambitious late twentieth century attempts to create an indexing system from scratch. A major break-through in the field of subject indexing was achieved by Classification Research Group (CRG), London and Derek Austin came out with a new method of subject indexing called PREserved Context Index System (PRECIS) in the early 1970s for the *British National Bibliography (BNB)*. When the BNB began in 1950, it used chain indexing system for about 20 years. However it was not, for various reasons, ideal for a computerised system, and in 1971, when BNB had developed the MARC system in the United Kingdom and was also engaged in using computers for the production of BNB itself, chain indexing was replaced by PRECIS.

What is PRECIS?

PRECIS is a system of subject indexing in which the initial string of terms organised according to the scheme of role operators, is computer manipulated in such a way that each sought term in the string functions as the approach term while preserving the full context of the document. Entries are restructured at every step in such a way that the user can determine from the format of the entry which terms set the approach term into its context and which terms are context dependent on the approach term.

Objectives of PRECIS

a) The computer, not the indexer, should produce all index entries. The indexer's responsibility is to prepare the input strings and to give necessary instructions to the computers to generate index entries according to definite formats.

b) Each of the sought terms should find index entries and each entry should express the complete thought content / full context of the document unlike the chain procedure where only one entry is specific—i.e. fully co-extensive with the subject of the document and others are cross references describing only one aspect of the thought content of the document.

c) Each of the entry should be expressive.

d) The system should be based on a single set of logical rules to make it consistent.

e) The system should be based on the concept of open-ended vocabulary, which means that terms can be admitted into the index at any time, as soon as they have been encountered in the literature.

f) The system must have sufficient references between semantically related terms.

Features of PRECIS

- It is more amenable to automatic manipulation than indexing based on the notational classifications.
- The permuted entries read naturally, which is achieved by the prescribed order of the role operators;
- The terms are linked to a machine-held thesaurus thereby providing possible '*see*' and '*see also*' references;
- PRECIS can be adapted to other languages.
- The indexer determines the meaning of the terms codes the roles and identifies the lead terms, whereas the computer takes care of the permutations.
- Its subject formulation is completely independent of classification, therefore exclusively geared to no classification numbers assigned in the MARC record.
- Context is preserved: It presents the full subject statement at every point of index entry, by gradual inversion of the concept string, thus overcoming the problem of the disappearing chain.

Principles of PRECIS

Two principles are followed in PRECIS:

a) **Principle of Context Dependency**: The "context-dependency" principle may be seen as a combination of context and dependency. When this principle is followed in a PRECIS input string, each term is qualified and sets the next term into its wider context. In other words, the meaning of each term in the string depends upon the meaning of its preceding term and taken together, they all represent the single context. Each term is hence dependent, directly or indirectly, on all the terms which precede it.

b) **Principle of One-to-One Relationship**: When the terms are organised according to the principle of context dependency, they form a one-to-one related sequence—each of the terms in the string directly related to its next term.

Syntax and Semantics

The syntax of PRECIS is based on the role operators, codes and logical rules which act as instruction to the computer. The semantics of PRECIS is handled by linking the terms to a machine-held thesaurus thereby providing possible '*see*' and '*see also*' references.

Role Operator: Role operators consist of a set of alpha-numeric notations which specifies the grammatical role or the function of the indexed term and regulates the order of terms in the input string. Role operators and their associated rules also serve as the computer instruction for determining the format, typography and punctuation associated with each index entry. There are two kinds of role operators: Primary operators and secondary operators. Primary operators control the sequence of terms in input string and determine the format of index entries. Any of the secondary operators is always to be preceded by the primary operator to which it relates.

Codes: Use of codes in the string brings expressiveness in the resulting index entries. Three types of codes are there: Primary, Secondary and Typographic codes.

Input String: A set of terms arranged according to the role operators which act as instructions to the computer for generating index entries.

Schema of Role Operators

Primary Operators		
Environment of core concepts	0	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	Viewpoint-as-form
	5	Selected Instance: study region, study example, sample population
	6	Form of document; Target user
Secondary Operations		
Co-ordinate concepts	f	'Bound' co-ordinate concept
	g	Standard co-ordinate concept
Dependent elements	p	Part; Property
	q	Member of quasi-generic group
	r	Assembly
Special classes of action	s	Roll definer; Directional property
	t	Author-attributed action
	u	Two-way interaction

Schema of Codes

Primary codes	
Theme Interlinks	$x 1st concept in coordinate theme
	$y 2nd/subsequent concept in co-ordinate theme
	$z Common concept
Term Codes	$a Common noun
	$c Proper name (class of-one)
	$d Place name

Secondary codes
Differences
Preceding differences (3 characters)— 1st and 2nd 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
Typographic 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

Entry Structure of PRECIS

The entry structure of PRECIS string consists of a two-dimensional display rather than the one dimensional that we have been accustomed to; instead of putting everything on one line, so that the only relationship which could be shown was that of following or proceeding, PRECIS uses two-line-three part entry structure as follows:

Lead	Qualifier
Display	

Lead is occupied by the approach term, which is the filing word and is offered as the user's access point in the index.

Qualifier position is occupied by the term(s) that sets the lead into wider context (i.e. general to specific). Together, the Lead and the Qualifiers correspond to the Heading. Terms in the heading set down in a narrower-to-wider context order. When the first term of the input string appears in the Lead position, the Qualifier position is usually kept blank.

Display position is occupied by those additional set of qualifying terms of the PRECIS string, which rely upon the heading for their context. When the last term of the input string appears in the Lead position, the Display position becomes empty.

Steps in PRECIS

Let us take the following title of a document for demonstrating the different steps of subject indexing according to PRECIS:

0. **Title**: University libraries in West Bengal

1. **Analysis**: Involves analysis of the thought content of the document and formulating the subject statement in natural language:

Measurement of the performance of university libraries in West Bengal

2. **Preparation of Input String**: Involves the identification of the status or role of each component term denoting key concept in terms of the role operators of PRECIS and assigning the appropriate operators to prepare an input string. The stages of the preparation of input string are furnished below:

[For understanding the stages of the preparation of input string, students are required to consult the schema of role operators and codes as furnished under the sub-section 11.3.4. of this Unit]

a) Identifying the concept signifying an action (if there be any).

   In the present example, the action concept is denoted by the term 'Measurement' and this term should, therefore be prefixed by the role operator (2).

b) Identifying the kind of action represented by this term, i.e. whether transitive action or intransitive action. 'Measurement' is clearly a transitive action since it is capable of taking an object. The object of transitive action is considered as the key system and is coded by the operator (1). In the present example, it is the 'Performance' which is being measured, so the input string should now appear in the form:

   (1) performance

c) Identifying the concept, if any related as property and/or whole-to-part. In this example, 'performance' is the property of 'university libraries' and 'libraries' is the part of the 'universities'. As a result, we will get the following input string:

   (1) universities

   (p) libraries

   (p) performance

   (2) measurement

d) The remaining term 'West Bengal' signifies the environment (i.e. geographical location) in which the whole thing takes place. Accordingly, the operator (0) is to be prefixed to the concept and the resulting input string will be:

   (0) "West Bengal

   (1) "universities

   (p) "libraries

   (p) "performance

   (2) "measurement

   Note: The primary operators have ordinal filing values and the terms in the above input string are sequenced accordingly. Thus, component terms are organised into the above input string according to the principle of context dependency. The secondary operator (p), prefixed with 'libraries' and denoting 'part' of the 'universities', is preceded by the primary operator (1) to which it is related. Similarly, the secondary operator (p), prefixed with ‘performance’ and also denoting ‘property’ of the ‘libraries’, is preceded by the secondary operator (p) meant for ‘libraries’ to which it is related. The terms, except proper name (i.e. West Bengal), are written in lower case initials. Index entries will be generated in upper case initials by the computer. Tick mark (“) is to be provided for the terms which should appear as Lead (access points) in the index entries.

3) **Generation of Index Entries:** The first index entry will be generated by the computer by pushing the first term of the input string into the Lead position and keeping the remaining terms in the Display position. As soon as any term goes to the Lead position, it is printed in bold type face.

West Bengal

Universities. Libraries. Performance. Measurement

The second index entries will be generated by pushing the second term of the input string into the Lead position and thereby replacing the existing Lead term into the Qualifier position, such as:

Universities. West Bengal

Libraries. Performance. Measurement

Similarly other index entries will be generated as

Libraries. Universities. West Bengal

Performance. Measurement

Performance. Libraries. Universities. West Bengal

Measurement

Measurement. Performance. Libraries. Universities. West Bengal

Note: It can now be seen in the above examples that Lead and Qualifier are separated by a full stop and 2-letter space. The standard separator between two terms in the entry is full stop and one space. The terms in the Display position are written leaving 2-letter space from the left. For over-run of Display in the next line, 4-letter space and for over-run of heading in the next line 8-letter spaces are to be left from the margin.

4) **Generation of supporting reference entries**: ‘see’ and ‘see also’ references are generated from semantically related terms taken from a machine-held thesaurus.

5) **Alphabetisation**: All the entries, generated by the process, as stated above, are arranged alphabetically by headings. Under the common heading, displays are organised alphabetically.

Formats of PRECIS Index

Index entries in PRECIS are basically generated in three formats: standard format, inverted format and predicate transformation.

a) **Standard Format**: Index entries in the standard format are generated when any of the primary operators (0), (1), and (2) or its dependent elements appear in the Lead. The process of generation of index entries in the standard format has already been demonstrated under the Section 11.3.4.7 of this Unit.

b) **Inverted Format**: Index entries in the inverted format are generated whenever a term coded by an operator in the range from (4) to (6) or its dependent elements appear in the Lead. The rule relating to the generation of index entries with this format is that—when any of the terms coded either (4), or (5) or (6) or any of their dependent element operators appear in the Lead, the whole input string is read from top to bottom and is written in the Display. However, if the term appearing in the Lead is last term of the input string, then it will be dropped from the Display.

Example: A report on the feminist viewpoint on marriage

Input String:

(2) ✓marriage

(4) ✓feminist viewpoint

(6) ✓reports

Index Entries:

Marriage

– Feminist viewpoint – Reports

Feminist viewpoint

– Marriage – Feminist viewpoint – Reports

Reports

– Marriage – Feminist viewpoint

c) **Predicate Transformation**: When an entry is generated under a term coded (3) that immediately follows a term coded either by (2) or (s) or(t)—each of which introduces an action of one kind or another—the predicate transformation takes place. An input string of this kind is shown below:

Example: Planning of libraries by architect

Input String:

(1) ✓libraries

(2) ✓planning \$v by \$w of

(3) ✓architects

In order to bring expressiveness in the resulting index entries, the connective codes \$v and \$w (see 'schema of codes') are attached to the action concept and it results in a compound phrase. The rule relating to the generation of index entries with this format is: when the term coded (3) goes to the Lead, the computer checks the operator assigned to the next preceding term. If that operator is (2) or (s) or (t), the term coded with any of these operators and the term accompanied by the Code \$w for upward reading connective (if any) are printed in the Display position instead of Qualifier position. Accordingly, the index entries for the aforesaid input string will be:

Libraries

Planning by architects

Planning. Libraries

By architects

Architects

Planning of libraries

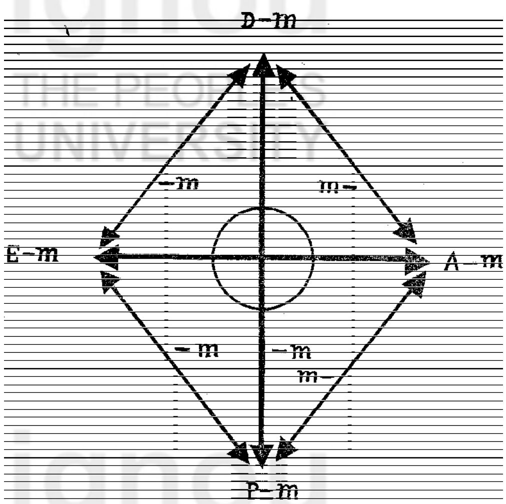
11.3.5 POPSI

All pre-coordinate indexing models are entirely based on the method of facet analysis. Ranganathan pointed out in a paper entitled *'Subject heading and facet analysis' (Journal of Documentation, 20 (3), 1964, p.109-119)* that facet analysis does not depend entirely on notational scheme of classification. The rules of chain procedure, he said, can be so framed as to implement any kind of decision about the sought first heading and the other successive headings in conformity with the principle of local variation. Since then, continuous research on this new line of thinking was going on at Documentation Research and Training Centre (DRTC), Bangalore and a number of papers on Postulate-based Permuted Subject Indexing (POPSI) based on Ranganathan's General Theory of Library Classification came out. Dr. Ganesh Bhattacharyya first explained the fundamentals of subject indexing languages with an extensive theoretical background which ultimately led to the development of newer version of POPSI forming the part of his General Theory of Subject Indexing Languages (GT-SIL). Bhattacharyya developed the POPSI through logical interpretation of the deep structure of subject indexing language (SIL). POPSI drew attention to the helpfulness of adopting a suitable device for ensuring an optimally effective organising classification through the alphabetisation of verbal subject – propositions. It prescribes the use of apparatus words – such as prepositions, conjunctions, participles etc., as and when necessary to communicate the exact meaning of subject propositions. These words are put in parenthesis and they are ignored in alphabetisation. Since the POPSI Index of all verbal entries, filing them in one alphabetical sequence in a unipartite index is made easy.

Major Working Concepts of POPSI

1) **Deep Structure of Subject Indexing Languages (DS-SIL)**

DS-SIL is the logical abstraction of the surface structures of outstanding SILs like Cutter, Dewey, Kaiser and Ranganathan. According to the General Theory of SIL, the structure of a specific SIL has been assumed to be a surface structure of the deep structure of SIL. The DS-SILs has been presented diagrammatically as follows:



It appears from the above diagram that any specific subject may belong to any one of the following elementary categories (D, E, A, P) and modifier:

2) **Elementary Categories and Modifiers**

a) **Discipline (=D)** refers to an elementary category that includes the conventional field of study, or any aggregate of such fields, or artificially created fields analogous to those mentioned above; e.g. Physics, Biotechnology, Ocean science, Library and Information Science, etc.

b) **Entity (=E)** refers to an elementary category that includes manifestations 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.

c) **Action (=A)** refers to an elementary category that includes manifestations denoting the concept of 'doing'. An action may manifest itself as Self Action or External Action. For examples: Function, Migration, etc. are Self Actions; and Treatment, Selection, organisation, and Evaluation, etc. are External Actions.

d) **Property (=P)** refers to an elementary category that includes manifestations denoting the concept of 'attribute'—qualitative or quantitative; e.g. Property, Effect, Power, Capability, Efficiency, Utility, Form, etc.

e) **M=Modifier** refers to a qualifier used to modify any one the elementary categories D, E, A and P. It decreases the extension and increases the intension of the qualified manifestation without disturbing its conceptual wholeness. A modifier can modify any one of the elementary categories, as well as two or more elementary categories. Modifiers are of two types:

- **Common Modifiers:** They refer to Space (e.g. Libraries in India), Time (e.g. Libraries in India 19th Century), Environment (e.g. Desert Birds), and Form (e.g. Encyclopedia of Physics). Common modifiers have the property of modifying a combination of two or more elementary categories.

- **Special Modifiers:** A special modifier is used to modify only one of the elementary categories. It may be of Discipline-based, or Entity-based, or Property-based, or Action-based. Special modifiers can be grouped into two types:

  i) those that require a phrase or auxiliary words to be inserted between the term and thus forming a complex phrase, e.g. Cataloguing using computers; and

  ii) those that do not require auxiliary words or phrase to be inserted in between the terms, but automatically form an acceptable compound term denoting Species/Type, e.g. 'Chemical' in 'Chemical Treatment'.

3) **Organising Classification and Associative Classification**

According to the General Theory of SIL, classification is a combination of both organising classification and associative classification. In other words, an indexing system is a combination of both organising classification and associative classification. The tasks involved in creating an organising classification are the categorisation of concepts and their organising in hierarchies. In organising classification compound subjects are based on genus-species, whole-part, and other inter-facet relationships. Here, classification is used to distinguish and rank each subject from all other subjects with reference to its Coordinate—Superordinate—Subordinate—Collateral (COSSCO) relationships. The result of organising classification is always a hierarchy. In associative classification, a subject is distinguished from other subjects based on the reference of how it is associated with other subjects without reference to its COSSCO relationships. The result of associative classification is always a relative index.

4) **Base and Core**

In the context of constructing compound subject heading, when the purpose is to bring together all or major portion of information relating to a particular manifestation or manifestations of a particular elementary category, the manifestation/category is Base. In case of a complex subject, any one of the subjects can be decided to be the Base subject depending upon the purpose in hand. For example: for a document on 'Eye cancer', 'Eye' is the Base subject in an Eye Hospital Library, and 'Cancer' is to be considered as the Base subject for a Cancer Research Centre.

When the purpose is to bring together within a recognised Base, all or major portion of information pertaining to one or more elementary categories, the category or categories concerned is the Core of the concerned Base. Core lies within the Base, and which one will be the Base or Core depends on the collection or purpose of the library. For example: In DDC, 'Medicine' is the Base, and the 'Human body' and its 'Organs' constitute the Core of the Base.

Features of POPSI'

From the operational point of view, the salient features of POPSI may be grouped under three components: Analysis, Synthesis, and Permutation.

The work of 'Analysis' and 'Synthesis' is primarily based on the postulates associated with the deep structure of SILs for generating organising classification. The task of analysis and synthesis is largely guided by the following POPSI-table. The work of 'Permutation' is based on cyclic permutation of each term-of-approach, either individually or in association with other terms for generating associative classification effect in alphabetical arrangement.

Rules of Syntax

The basic rules of syntax associated with POPSI are:

a) Discipline is followed by Entity, both modified and unmodified.

b) Property follows immediately the manifestation in relation to which it is a Property.

c) Action follows immediately the manifestation in relation to which it is an Action.

d) A Property can have its own Property.

e) An Action can have its own Action.

f) A Species/Type follows immediately the manifestation in relation to which it is a Species/Type.

g) A Part follows immediately the manifestation in relation to which it is a Part.

h) A Modifier follows immediately the manifestation in relation to which it is a Modifier.

The following POPSI Table, like Role Operators in PRECIS, is used in sequencing the component terms for formulating a subject heading

<table>
<tr><td colspan="3">0 Form modifier</td></tr>
<tr><td colspan="3">1 General Treatment</td></tr>
<tr><td colspan="3">2 Phase relation
2.1 General
2.2 Bias
2.3 Comparison
2.4 Similarity
2.5 Difference
2.6 Application
2.7 Influence</td></tr>
<tr><td colspan="2">Common modifiers
3 Time modifier
4 Environment modifier
5 Place modifier</td><td></td></tr>
<tr><td>6 Entity (E)
7 Discipline (D)</td><td>.1 Action (A)
.2 Property (P)
Note: Notations .1 and .2 are preceded by the notation of the manifestation in relation to which it is (A) and (P).</td><td>, Part
. Speciator/Type
—Special modifier
Note: A Species/Type/ Special modifier follows immediately the manifestation in relation to which it is a Species/Type.</td></tr>
<tr><td colspan="3">8 Core I
9 Base (B)
Note: Features relating to Core I and Base (B) are analogous to 6 Entity / 7 Discipline /.1 Action / .2 Property.</td></tr>
</table>

Semantic Relationship in POPSI

Semantic relationship in POPSI is controlled by the vocabulary control device called 'Classaurus'. A classaurus is an elementary category-based (faceted) systematic scheme of hierarchical classification in verbal plane incorporating all the necessary features of a conventional information retrieval thesaurus. Like any faceted classification scheme, a classurus consists of separate schedule for each of the Elementary Categories, namely, Entity, Property and Action, with their Species/Parts and Special Modifiers for each.In each schedule, it displays hierarchical relationship among terms (broader, narrower, and collateral). Classaurus also contains separate schedules for Common Modifiers like Form, Time, Place, and Environment. Like any thesaurus, each of terms in hierarchic schedule is enriched by synonyms, quasi-synonyms, etc. Unlike a thesaurus, a classaurus does not include associatively related terms (RTs) because of its category-based (faceted) structure. The implication of the faceted structure is that a term in one Elementary Category has a high chance of being non-hierarchically related with another term in another category. The task of showing what is non-hierarchically related to what, and how they are related is left to the care of indexing procedure based on the information contained in the document itself. All the component terms associated with the content of the document are brought together under each approach-term by the permutation technique of POPSI. It is assumed that RTs should not be dictated beforehand by the designer of the classaurus, rather it should be dictated by the content of the document itself. Any term may be related to any other terms depending on the content of the document and hence, RTs should not be determined beforehand.

Steps in POPSI

The main steps in applying POPSI are as follows:

1) **Content Analysis**: Involves identification of different component ideas associated with the content of the document with reference to their elementary categories and modifiers.

2) **Formalisation**: Involves preparing the formalized expression of subject statement obtained on the basis of the results of the step 1 (Content Analysis) according to the rules of syntax.

3) **Standardisation:** Deciding the standard term in the formalized expression of subject statement especially for the term(s) having synonym(s), if any. This step calls for the use of Classaurus.

4) **Modulation:** Involves augmenting the standardized subject proposition by interpolating and extrapolating, as the case may be, the successive superordinates of each manifestation by using standard terms with indication of their synonyms, if any. [*Note: A* Classaurus is the tool to guide the operation in steps 3 and 4 with assurance of consistency in practice].

5) **Entry for Organising Classification**: Involves the preparation of entries for organising classification by inserting appropriate notations for Elementary Categories, subdivisions and modifiers from POPSI Table. Modulated subject proposition with appropriate notations from POPSI Table sorted alpha-numerically will produce organising classification effect by juxtaposition of entries in alphabetical sequence.

   Approach-term Selection: Consists of deciding the approach-terms for generating associative classification effect and of controlling synonyms. The selection of approach-terms may vary for one library to another library depending upon the requirement of users.

6) **Preparation of Entries of Associative Classification**: Involves the preparation of the entries under each approach-term by cyclic permutation of sought terms for generating associative classification effect in alphabetical arrangement.

7) **Alphabetisation:** Involves in arranging all the entries in alphabetical sequence.

Demonstration of the Procedure of POPSI-Basic

0) **Title of the document**:

Use of computers for indexing of educational films in university libraries in India

1) **Content Analysis:**

   Library and Information Science = Discipline (D) [Implicit]

   University libraries = Entity (E) [Explicit]

   Educational films = Part of E [Explicit]

   Indexing = A of Part of E [Explicit]

   Use = Application phase relation (PR) [Explicit]

   Computers = E-based Special modifier (Sm) [Explicit]

   India = Common modifier (Cm) of place [Explicit]

2) **Formalisation:**

   Library and Information Science (D), University libraries (E), Educational films (Part of E), Indexing (A of Part of E), Use (Application PR), Computers (E-based Sm), India (Cm)

3) **Standardization:**

   It is assumed that all the terms in the formalized expression of the subject as shown above are standard terms.

4) **Modulation:**

   Library and Information Science (D), Libraries. Academic libraries. University libraries (E), Information sources. Films. Educational films (Part of E), Indexing (A of Part of E), Use (Application PR), Computers (E-based Sm), Asia, India (Cm)

5) **Entry for Organising Classification:**

   Library and Information Science 6 Libraries. Academic libraries. University libraries, Information sources. Films. Educational films 6.1 Indexing 2.6 (using) Computers 5 (in) Asia, India

6) **Approach-term Selection:**

   The following terms are selected as the approach-terms:

   Libraries

   Academic libraries

   University libraries

   Information sources

   Films

   Educational films

   Indexing

   Computers

   India

7) **Preparation of Entries of Associative Classification:**

LIBRARIES

Library and Information Science 6 Libraries. Academic libraries. University libraries, Information sources. Films. Educational films 6.1 Indexing 2.6 (using) Computers 5 (in) Asia, India

The above organising classification will have to be repeated under each of the following approach terms:

Academic libraries

University libraries

Information sources

Films

Educational films

Indexing

Computers

India

8) **Alphabetisation:**

All the entries are arranged according to the alphabetical order, word-by-word.

POPSI-Specific

The steps in POPSI with illustrative examples as demonstrated above fall within the purview of POPSI-Basic. According to Bhattacharyya, there is no single absolute version of organising or associative classification. POPSI tries to find out what is logically basic, and amenable to systematic manipulation to meet specific requirement. The POPSI-Basic is a product of the application of the GT-SIL and it is readily amenable to the systematic manipulation to generate purpose-oriented specific versions known as POPSI-Specific. POPSI-Specific is always a derivation from the POPSI-Basic according to special decisions and rules to meet specific requirements at the local level. It may be noted here that this approach is totally different from that of earlier contributors of different SILs.

If the purpose is to bring together all or a major portion of information pertaining to a specific topic in a discipline manifesting any of the Elementary categories, the above version of POPSI-Basic can be systematically manipulated to generate the required version of POPSI-Specific. This involves the decision about the 'Base' and 'Core'.

For example, our purpose is to bring together all information pertaining 'Educational films' in one place and hence, 'Educational films' is to be considered as the Base. 'University libraries' and 'Indexing' are to be considered as the Modifier and Action to Base respectively. In view of this, we can prepare the following organising classification entry:

Information sources. Films. Educational films — Libraries. Academic libraries. University libraries 9.1 Indexing 2.6 (using) Computers 5 (in) Asia, India.

Self Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

5) Discuss the Principle of Context Dependency.

....

....

....

....

6) Discuss the entry structure and entry format as followed in PRECIS?

....

....

....

....

7) How do you categorise different operational stages of POPSI?

....

....

....

....

8) What are the major steps in formulating index entries according to POPSI?

....

....

....

....

11.4 POST-COORDINATE INDEXING

All indexing systems follow the process of concept coordination to describe the contents of the documents more precisely. We have seen in pre-coordinate indexing, component concepts are coordinated according to the order of significance or citation order by following the syntactical rules of the given indexing language. But the rigidity of the citation order appears to be unsatisfactory to meet the varieties of approaches of all the users. The provision for multiple entries in pre-coordinate indexing by rotating or cycling of the component terms covers only a fraction of the possible number of the total permutations and for this, a large portion of probable approach points is left uncovered. Consequently, the searcher has no choice but to follow the rigid citation order specified by the given indexing language. The above noted problems stemming from the pre-coordination of terms with the rigidity of citation order triggered the development of alternative indexing techniques where the component ideas of a subject are kept separately, uncoordinated by the indexer. Here, concepts/terms are coordinated at the time of searching (i.e. at output stage) by the user. A greater degree of search manipulation is available in post-coordinate indexing system since the search terms can be coordinated almost in any combination or order to retrieve records of information about the documents as required by the users. The indexing systems which are based on this basic principle are called *Post-coordinate Indexing* or simply *Coordinate Indexing Systems*. Based on this basic principle a number of post-coordinate indexing systems like Uniterm, Optical Coincidence Card, etc. were developed. Among the different types of post-coordinate indexing systems, Uniterm system developed by Mortimer Taube is considered as the most popular post-coordinate indexing model.

11.4.1 Pre-Coordinate Indexing versus Post-Coordinate Indexing

Subjects of documents are not simple. There are compound and complex subjects dealing with multiple numbers of concepts. When there is more than one concept in the document the order in which we cite the concepts and their relationship to one another become important. Both pre- and post-coordinate indexing systems are, by nature, coordinate indexing, but the coordination is done in two different stages. The following table furnishes the points of differences between the two systems:

Pre-coordinate indexing system	Post-coordinate indexing system
Coordination of component terms is carried out at the time of indexing (i.e. at input stage) in anticipation of the users' approach.	Component concepts (denoted by the terms) of a subject are kept separately uncoordinated by the indexer, and the user does the coordination of concepts at the time of searching (i.e. at the output stage).
The most important aspect of this indexing system is to determine the order of significance by following the syntactical rules of the given indexing language.	Rigidity of the significance order is very much absent in this system.
It is non-manipulative. The searcher has no choice but to try to predict the citation order specified by the indexer.	It is manipulative. The searcher has wide options for free manipulation of the classes at the time of searching in order to achieve whatever logical operations are required.
In this indexing system, both the indexer and the searcher are required to understand the mechanism of the system—the indexer for arriving at the most preferred citation order and the searcher for formulating an appropriate search strategy—in order to achieve the highest possible degree of matching of concepts.	This indexing system does not require the indexer and searcher to understand the mechanism of the system. However, the operational aspects n need to be understood by them.

11.4.2 Term Entry System and Item Entry System

In *Term Entry System*, we prepare entries for a document under each of the appropriate subject headings, and file these entries alphabetically. Here, terms are posted on the item (i.e. Term on Item System). In this type of post-coordinate indexing, the number of entries for a document is dependent on the number of terms associated with the thought content of the document. Searching of two files (Term Profile and Document Profile) is required in this system. Uniterm and Peek-a-boo are examples of these.

It is possible to take the opposite approach and make a single entry for each item, using a physical form which permits access to the entry from all appropriate headings. A system which works in this way is called an *Item Entry System*. Here, items are posted on the term (i.e. Item on Term System). In this type of post-coordinate indexing, single entry is made for each item. Item entry system involves the searching of one file (i.e. Term Profile) only. Edge-notched Card is an example of item entry system.

11.4.3 Uniterm Indexing

Uniterm indexing system was devised by Mortimer Taube in 1953 to organise a collection of documents at the Armed Services Technical Information Agency (ASTIA) of Atomic Energy Commission, Washington. Uniterm is a post-coordinate indexing system based on term entry principle. Here, component term (uniterm) is independent of all other terms and serves as a unique autonomous access point to all relevant items in the collection.

Indexing Process

The processes involved in Uniterm indexing are as follows:

1) **Preparation of Document Profile**: Here indexer is required to prepare a card for each incoming document to:

   - assign an accession number or 'address' to each incoming document and record the same on the card,
   - identify the different component terms (uniterms) associated with the thought content of the document,
   - select the uniterms and record them on the card,
   - prepare an abstract for each document and record the abstract on the card,
   - write the bibliographical details of the document on the card, and
   - arrange all the cards as prepared for all incoming documents according to the ascending order of the accession numbers of the documents.

Cards so prepared and arranged according to the ascending order of the accession numbers of the documents are called *Documents Profiles*.

2) **Preparation of Term Profile:** Here indexer is required to

   - prepare a card (term card) for each uniterm,
   - divide each term card into 10 equal vertical columns (from 0 to 9),
   - record the accession number of the document relevant to the uniterm according to the system of terminal digit posting. Terminal digit of the accession number determines the column of its posting, and
   - arrange all term cards according to the alphabetical order of the uniterms.

Term cards so prepared as stated above and arranged according to the alphabetical order of the uniterms are called the *Term Profile* of the system.

Searching

The process of searching in Uniterm system involves the following operations:

- Searcher identifies the different component terms/uniterms associated with the content of her/his queries.
- Searcher, after identifying the component terms/uniterms, pull out the pertinent term cards from the alphabetical deck of the *Term Profile.*
- Term cards thus pulled out are matched to find the common accession number(s). The number(s) common in all such Uniterm cards represent the sum total of the component concept of the specific subject.
- With the help of the common accession number(s), relevant card(s) are pulled out from the *Document Profile* where full bibliographical information of the required document(s) is available.

The main advantage of the Uniterm indexing system is its simplicity and the ease with which persons without much knowledge of subject indexing can handle it. The criticisms against Uniterm system centre around: (a) search time: involves much searching time because of the searching of two files—Term Profile and Document profile; and (b) false drops: possibility of retrieving irrelevant documents due to false coordination of uniterms. For example, searching with the uniterms 'Teachers', 'Students' and 'Evaluation' may retrieve documents on both the subjects, 'Evaluation of students by teachers and 'Evaluation of teachers by students', one of which might be irrelevant to a particular user.

To overcome the problem of false drops, the following post-coordinate searching devices have been used:

1) **Use of Pre-coordinated Terms**

   It is the introduction of pre-coordination to some extent in post-coordinate system in which two or more terms in a subject are bound in place of isolated single term/ uniterm to get rid of false coordination.

2) **Links**: Links are special symbols used to group all the related concepts in a document separately, so that inappropriate combinations of terms are not retrieved. Suppose we have a document (accession number: 243) dealing with two different topics—*Classification of non-book materials and indexing of films.* In order to avoid false coordination like *Classification of films and indexing of non-book materials*, alphabetical symbols, which serve as interlocking device, are attached to accession number to indicate different groups:

Classification	243A
Non-book Materials	243A
Indexing	243B
Films	243B

3) **Roles:** Roles are the indicator digits/symbols attached to the terms at the time of indexing to indicate the role or status or use of the term in a particular context. Here, the possible roles of different terms are identified beforehand and terms are

tagged with these role indicators at the time of indexing. For example, roles developed by the Engineering Joint Council, known as EJC role operators may be attached to *'Television'* to distinguish its functions as the product and tool:

Role	Document
2 [Product / Output]	Manufacturing of *television*
3 [Agent / Tool]	Use of television in education

4) **Weighting:** It is the device of allocating quantitative values to the index terms according to their degree of relevance in the document. Different ways for indicating weights have been suggested. A simple system uses numbers 1 to 3, where 3 indicates maximum weight (i.e. the index term is highly specific and covers an entire major subject of the document), 2 and 1 indicate weights of index terms in decreasing order of their values or relevance in the document.

11.5 KEYWORD INDEXING

Keyword indexing is based on the usage of natural language terminology for generating the index entries. The term 'Keyword' refers to a significant or memorable word (also called 'catchword') that serves as a key in denoting the subject taken mainly from the title of the document (so, it is called *title index)* and sometimes from the abstract or text of the document. Common words like articles (a, an, the) and conjunctions (and, or, but) are not treated as keywords because it is inefficient to do so. This system is also known as N*atural or Free Indexing language.* It is to be pointed out here that the concept of keyword indexing is not new and it existed in the nineteenth century as a 'catchword indexing'. With the introduction of computers in information retrieval in the 1950s, Hans Peter Luhn, an IBM engineer, presented a computer-produced index in 1958 that became known as KWIC (Key Word In Context) indexing.

11.5.1 Key Word in Context Indexing (KWIC)

The production of a KWIC index by H. P. Luhn is the earliest example of an automatic index produced using computers to perform repetitive tasks associated with subject indexing. It was a great step forward in the technique of automatic indexing. Utilizing the capabilities of computers, the KWIC method speedily and with a minimum of intellectual effort produces indexes derived solely from the titles of the documents to be analysed. All significant or key words of titles/title like phrases are alphabetised mechanically, and then printed out in turn following a format which emphasises the selected word. The computer uses the 'stop-word' list in order to ignore all syntactical words such as articles; prepositions etc., and select the remaining words in the title as indexing words. The remaining words of the title are arranged to stand in the context of their original appearance. The use of 'stop-word list' reduces the volume of the index.

The result of the machine manipulation is an index of key terms printed in alphabetical order, together with the text immediately surrounding each term. Each significant word as entry point appears in the margin or a designated middle position while the rest of the title printed on either side. The alphabetical filing is done on the basis of the key word printed in bold letters. *Chemical Titles* (of Chemical Abstract Service) and *BASIC* (Biological Abstracts Subjects in Context) are faithful adaptation of Luhn's KWIC indexing.

Let us consider the following title 'Chemical treatment of cancer in the hospitals of Chennai' to demonstrate the index entries generated according to KWIC indexing:

Cancer in the hospitals of Chennai / Chemical treatment of	614
Chemical treatment of cancer in the hospitals of Chennai	614
Chennai / Chemical treatment of cancer in the hospitals of	614
Hospitals of Chennai / Chemical treatment of cancer in the	614

Annotation

a) Title in the above KWIC index has been rotated in such a way that each keyword serves as the approach term and comes in the beginning by rotation followed by rest of the title;

b) Last word and first word of the tile are separated by using a symbol say, stroke [/] (sometime an asterisk '*' is used) in an entry. In some computer-produced KWIC indexes, keywords are positioned at middle of the entry;

c) Keywords are printed in bold type face to bring prominency in the approach term;

d) Identification / location code 614 is given at the right end of each entry; and

e) Entries are arranged alphabetically by keywords.

Advantages and Disadvantages of KWIC

KWIC method offers the following advantages: (1) the detection of formulaic expressions and repeated word patterns; (2) simplicity; (3) speed; (4) maximisation of computer use; and (5) minimisation of the indexer's role.

The most common type of complaint against the KWIC indexing method is the lack of terminology control as it is entirely dependent upon titles/abstract/text of the document. Apart from this, KWIC has basically two problems: (i) KWIC shows sentences which contain distant dependency as different context, and (ii) KWIC also shows sentences which have different word order as different context. The effects of computer's inability to resolve these problems led to the redundancy, scatter of references throughout the index, haphazard groupings and retrieval losses because the user is forced to guess at the terminology the author actually used. The disadvantages of KWIC-type index can be summarized as follows:

— Large number of index entries under a given keyword, which provokes difficulties in searching;

— Lack of significant words in titles (therefore the title and the abstract are often used as a source for indexing to increase the depth and range of indexing);

— No cross references, which make it difficult to find synonyms, spelling variants and inflections in the index;

— Relatively high computing time, due to superfluous non-significant index entries;

— No combination of keywords;

— Lack of consistency in the indexing terms, because different authors can use different form of words to communicate the same idea, or give different meanings to the same word or phrase;

— References are not grouped under a convenient heading;

— Redundancy of the index.

11.5.2 Variations of Keyword Indexing

A number of varieties of keyword index appear in the literature and they differ only in terms of their formats but indexing techniques and principle remain more or less same. Some important versions of keyword indexing are discussed below:

Key Word Out of Context (KWOC)

In KWOC, each index word is extracted from its context and printed separately in the left hand margin with the unmodified title in its normal order printed to the right. For example:

Cancer Chemical treatment of cancer in the hospitals of Chennai 614

Chemical Chemical treatment of cancer in the hospitals of Chennai 614

Chennai Chemical treatment of cancer in the hospitals of Chennai 614

Hospitals Chemical treatment of cancer in the hospitals of Chennai 614

Sometimes, keywords are positioned and printed as heading and the title is printed in the next line under the heading instead of the same line as shown above.

Key-Word Augmented-in-Context Index (KWAC)

It has been observed that the dependency of keyword indexing on titles sometimes fails to represent the thought content of the document co-extensively. In order to solve this problem, KWAC index came into being. In KWAC, the keywords of the title are enriched with additional keywords taken either from the abstract or from the original text of the document and are inserted into the title or added at the end to generate further index entries. KWAC is nothing but the enrichment of KWIC or KWOC. Further enriched KWIC or KWOC gives index entries wherein additional terms are inserted into the title or added at the end. This involves intellectual effort in the selection of additional terms. CBAC (Chemical Biological Activities) of BIOSIS uses KWAC index where title is enriched by another title like phrase formulated by the indexer.

11.5.3 Double KWIC

It is another improved version of KWIC. Double KWIC index is constructed in the following way:

a) The first significant word in a title is extracted as a main index term and replaced by an asterisk (*) to indicate its position in the title.

b) The remaining words in the title are then rotated, so as to permit each significant word to appear as the first word of a wrap-around subordinate entry under the main index term. Steps 1 and 2 are repeated until all of the titles of a given bibliographic listing are processed. The index entries so created are then sorted alphabetically, both with regard to main terms (primary sort) and subordinate terms (secondary sort). Main index terms are not restricted to single words, but may consist of multi-word terms derived from contiguous sets of words in the titles, for example:

ANNUAL REPORT
EDITORIAL:*
1966=……………………………………………………………………………….. F 1
ANNUAL REVIEW
BOOK REVIEW:* OF INFORMATION SCIENCE AND TECHNOLOGY= B3-2.
INFORMATION SCIENCE AND TECHNOLOGY= BOOK REVIEW:*OF B3-2
SCIENCE AND TECHNOLOGY= BOOK REVIEW:* OF INFORMATION B3-2
TECHNOLOGY= BOOK REVIEW:* OF INFORMATION SCIENCE AND B3-2

11.5.4 Other Versions

In addition to the above variations in keyword indexing, a number of varieties of keyword index are available and they differ only in terms of their formats but indexing techniques and principles are more or less the same. They are:

i) **KWWC (Key-Word-With-Context) Index:** In KWWC, only the part of the title, instead of full title, relevant to the keyword is considered as entry term.

ii) **KEYTALPHA (Key-Term Alphabetical) Index:** The KEYTALPHA is just modified form with key terms arranged alphabetically. It is permuted subject index that lists only keywords assigned to each abstract. Keytalpha index is being used in the 'Oceanic Abstract'.

iii) **WADEX (Word and Author Index):** It is an improved version of KWIC index where along with the key words, the names of authors are also treated as keywords and thus indexed accordingly. Thus, it appears that WADEX satisfies both the author and subject. WADEX is used in 'Applied Mechanics Review'. AKWIC (Author and keyword in context) index is another version of WADEX.

iv) **KLIC (Key-Letter-In-Context) Index:** This type of index only takes fragmented word (i.e. key letters), instead of the full word, either at the beginning or at the end of the entry. In this system, the key letters forming the part of the word are specified and the computer retrieves any term containing that letters either at the beginning or at the end of the word. KLIC indexes are almost unknown today, the Chemical Society (London) published a KLIC index as a guide to truncation.

Self Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

9) What do 'Pre-' and 'Post-' signify in Pre-coordinate and Post-coordinate indexing systems?

....

....

....

....

10) What are the search devices used to avoid false coordination of terms in Post coordinate indexing?

....

....

....

....

11) Mention the different varieties of keyword indexing.

....

....

....

....

11.6 COMPUTERISED INDEXING

You already came to know that information retrieval deals with the problems related with the storage, access and searching of information sources by persons in need of information. In this digital network era, information sources are growing at an exorbitant rate, available in many forms and formats, and accessible through various channels. Moreover, recent advancements in Information and Communication Technology (ICT) help in integration of different information sources and process them on a larger scale. The results of ICT applications in library activities are related with the development of wider and efficient information services. Remote database search service (both bibliographic and full-text databases) is possibly one of the most prominent products of ICT-enabled library services.

But you will be surprised to know that investigations of computerised indexing date back to the late 1950s. The earliest and most primitive form of computerised indexing method relying on the power of computers was invented by Hans peter Luhn, an IBM engineer, who in 1958 produced what became known as KWIC (Key Word In Context) indexing. Luhn reported his system in 1960 which has already been discussed under the sub-section 11.5.1. The system was based just on simple, mechanical manipulations of terms derived from document titles. Related forms are the Permuterm Subject Index and the KeyWord Plus known from ISI's citation indexes (this last system is based on assigning terms from cited titles). Computers are now increasingly used in aiding indexing, relying on stored dictionaries of synonyms and homonyms, lists of chemical compounds, plants and animals, etc. They are also used for automatically arranging entries in alphabetical order or subordinating subheadings and cross references in exact sequence under a heading, and performing many other functions that previously had to be done manually and therefore were quite expensive and often subject to errors. After the introduction of Unicode (a 2 Byte-oriented encoding standard that can represent all characters of all scripts of the world) as text encoding standard (in late nineties), computerised indexing systems are able to store, process and retrieve multilingual documents available in different scripts.

11.6.1 Meaning and Features

Computerised indexing can be defined as the process whereby a computer is used to process a natural language text that is already in machine-readable form so that indexing terms are allocated to its content without direct human intervention.

The features of computerised indexing are:

1) Computerised indexing starts with words.

2) Word association prompts the linking of target words in a search statement.

3) Computers scan text and create 'inverted file' which associates words in the file with position in the texts.

4) Matches words in a search statement against 'inverted files' to identify texts that have words in common.

5) Computer algorithms are used to carry out the above operations.

6) Humans do the programming and set the parameters for indexing.

7) Computation techniques used include word frequency and keyword analysis.

8) Computerised indexing cannot replace human or manual indexing, rather complimentary.

11.6.2 Manual Indexing versus Computerised Indexing

Manual Indexing	Computerised Indexing
• Human indexer analyses texts and selects terms for indexing. • Human indexer interprets and encodes text, and makes inferences and judgment in selecting index terms judiciously. • Semantic, syntactical as well as contextual considerations govern the selection of indexing terms • Disagreement among indexers on the determination of the subject of a document as the process of determination of subjects may variy from one indexer to another indexer. • Selected index terms less in numbers. • It is time consuming. • It is expensive. • It is very difficult to maintain consistency in indexing. • Relevancy of results are ensured • Generally maintains a balance in recall-precision.	• Computer analysis of texts has not achieved the reliability of human analysis. Computer analysis of texts is carried out by following the human instructions in the form of a computer programming. Computers can select words by employing statistical technique. • No text interpretation is possible as computer cannot think and draw inferences like human indexer. It can select or match keywords which are provided as input text. • Computer algorithms are drawn to select, or exclude a term by following the rules of semantic, syntactical and contextual connotations, like human indexer. • Determination of the subject of the document is a mechanical process, based on what terms appear frequently and/or prominently in the text—i.e. more frequently a term occurs in a document, the more likely it is that the document is about that term. • Selected index terms are more in number. • It takes less time. • Index entries can be produced at lower cost. • Consistency in indexing is maintained. • Relevancy of results are not always ensured. • Generally shows high recall and low precision.

11.6.3 Advantages and Disadvantages of Computerised Indexing

Advantages of computerised indexing are as follow:

- It is as effective as human indexing;
- It is cost effective compared to expensive human indexing;
- Maintains consistency in indexing;
- Indexing time is reduced;
- Help searchers find information quickly;
- Can be applied to large volumes of texts where human indexing becomes impossible (e.g. Indexing web pages);
- Retrieval effectiveness can be achieved.

Disadvantages are

- Not flexible;
- Not precise when looking at unique materials;
- Not able to adapt new terminology;
- Not able to do the conceptual analysis of the content of the document;
- Not a term occurs several times in a document will always be a significant term.

11.6.4 Components of Computerised Indexing System

A typical computerised indexing system has four major components:

1) **Database**

Database acts as heart of a typical computerised indexing system. Bibliographic databases which deal with metadata of bibliographic entities (e.g. author, title, subject etc.) include two parts. The first part is **sequential file** (a combination of fields —> records —> database) and **inverted file** (indexes to sequential file). Full-text databases, apart from the above two parts also include field-less information entity (source object in different formats like Web page (HTML), PDF file, Doc file etc).

2) **Search Process**

Database determines what can be retrieved, whereas search mechanism determines how information stored in databases can be retrieved. Efficiency of search process is very important factor for a computerised indexing system. Search process of a typical computerised indexing system provides two sets of retrieval techniques – basic retrieval techniques (Boolean operators – AND, OR, NOT; Relational operators - >, <, =, >=, <=; and Positional search operators – NEAR, ADJ, NEARx etc.) and advance techniques (Weighted searching, Fuzzy searching etc.).

3) **Language of Indexing**

Search mechanism determines what retrieval techniques will be available to searchers, whereas computerised indexing language determines the flexibility in (1) document representation; and (2) query representation. Computerised indexing language may be grouped as natural language and controlled vocabulary (classification, subject heading and thesauri).

4) **User Interface**

It is a layer of interaction between users and computerised indexing system. Efficiency of user interface depends on mode of interaction, display features, online help, provision of feedback etc. It is considered as the human dimension of computerised indexing system.

11.6.5 Categories of Computerised Indexing Systems

Since the time of H.P. Luhn (i.e. 1950s) different computerised indexing systems have been evolved to address information demands of different user groups. These indexing systems may be categorised as follows –

Category I: Online Database Indexing

These computerised indexing systems allow users to search databases located in remote places. Here computer technologies are applied to process, store, retrieve records and communication technologies help in accessing records from centralised databases. Generally, these IR systems include bibliographical, numeric, full-text and multimedia information bearing objects. The online IR systems played a major role in the development of computerised indexing systems over the years.

Advantages: Sophisticated retrieval techniques, cross-database searching, use of integrated vocabulary control devices and many more.

Limitations: Heavy initial investments, need of intermediary in searching etc.

Category II: Optical Disk based Database Indexing

These computerised indexing systems emerged from online IR systems. These IR systems offer subset of data (bibliographical, numeric, full-text, and multimedia) through the optical media (like CDROM, DVDROM etc.). These IR systems adopted almost all the retrieval techniques of online IR systems but the end users, rather than the intermediary, does most of the searching in optical disk based IR.

Advantages: Low running cost, end-user friendly, browsing and searching facilities, sharing of databases on LAN etc.

Limitations: Delayed updating (updating frequency ranges from quarterly to bi-annually) and restricted access (remote searching is not possible beyond LAN).

Category III: OPAC based Indexing

An OPAC is outer / external form of library catalogue. OPACs are presently considered as IR systems with their own characteristic features. For example, an OPAC is essentially local (serve library resources of one or more institutes) but can act as global information entity by integrating other IR systems e.g. CDROM databases, e-journals access, Z 39.50 based searching of other library catalogues etc.

Advantages: Supports field-level search (i.e. search by author, title etc.), and subject access points, provides sophisticated retrieval techniques, and facilitates integration of different information resources into a single search interface.

Limitations: Lack of federated search options, absense of multi-lingual user interface etc.

Category IV: Indexing Internet Resources

The first three computerised indexing systems are dealing with structured objects (fields and field values) whereas the last one deals with unstructured information bearing objects (textual objects without metadata). Internet based computerised indexing systems are generally based on automatic harvesting devices such as robots, spiders, crawlers etc. for finding and gathering of information resources available in publicly indexable Internet (see section 11.7).

Advantages: Free indexing services like search engines, subject directories amd meta search engines, quick access to huge information sources, supports end-user searching, simple to retrieve documents.

Limitations: High recall and very low precision, cross-disciplinary semantic drift, relevancy of results not ensured etc.

11.6.6 Comparison of Computerised Indexing Systems

In the previous section we discussed different categories of computerised indexing systems by covering features of the major systems including the relative advantages and disadvantages of those indexing systems. In this section we are going for a comparison of major computerised indexing systems on the basis of a set of defined parameters.

Features	Online Databases	Optical Disk based Databases	OPAC	Internet
Contents	Mainly textual objects; Bibliographic metadata	Bibliographic metadata, Fulltext and Multimdia	Bibliographic metadata	Text, Images and Multimedia
Retrieval approach	Searching	Searching and browsing	Searching and browsing	Searching and browsing
Indexing	Metadata and Keywords in abstract	Metadata and Keywords in full-text	Metadata and Keywords in TOC	Limited metadata, Keywords in full-text
Retrieval techniques	Basic and Advanced (Limited)	Basic and Advanced (Limited)	Basic and Advanced (Full)	Basic and Advanced (full) including Fuzzy
Search modification	Yes	Yes	No	Yes
Search method	Indexing based (expert search)	End-user search	End-user search	End-user search
Output / Display	Traditional ranking	Traditional ranking	Traditional ranking and Limited modern ranking	Different modern ranking facilities
Use of controlled vocabulary	Yes	Yes	Yes	No
Quality control mechanisms	Yes	Yes	Yes	No
Multilingual search	No	No	Yes	Yes

However, the trend of in the domain of IR system is convergence. CDROM based systems are integrated with OPACs, OPACs are linking online databases, document delivery services, and other resource discovery services. MARC 21 bibliographic format includes field 856 for encoding URLs of Internet resources. The Web is becoming the platform for convergence of different IR systems e.g. Web-OPACs are linking open databases (information mashup) and acting as the gateways for local and global information resources to support users.

11.6.7 Index File Organisation

You must have noticed that all four major computerised indexing systems depend on database as core component. Database, on the other hand depends on two basic parameters – method for handling sequential file and method for handling inverted file. Sequential file management ensures organisation of records in a database whereas inverted file management helps in retrieving records from database against queries. Therefore, design and development of the computerised indexing system mainly depends on:

1) Method used for organising records in the file (file organisation), and

2) Method used for searching a record in the file (search techniques).

Further, these two factors are related in a sense that more efficient the file organisation, more efficient the searching.

In computerised indexing, each document is represented by a record that a computer can read and manipulate. Typically, these records contain data identifying an item and a set of index terms to provide subject access. A basic problem in computerised indexing is storing of files consisting of records, each record having an identical format. A record format consists of a list of fields, with each field processing a number of characters and having a fixed data type. A record also consists of values for fields. The complexity of organising a file for storage depends on the operations we intend to perform on the file.

A file can be categorised as a logical file or a physical file. A logical file is the one perceived by an application program, it may be different from the one which is stored in storage units. Logical files are thus abstracts groups of data. Actual data are stored using physical files. A physical file is a set of data stored on a physical medium such as disk, tape, etc. It contains a number of data subsets, called physical records, which have an identical layout.

Some of the important logical file organisations are discussed below:

a) **Sequential File**: The most common organisation of records in a file is the sequential file organisation. It is the document file, which contains document records in their normal form—the form in which they are sequentially entered into the database. Here, document records are stored one after another in the computer memory—this is actually the virtual structure of the database file.

b) **Inverted File**: An inverted file is a computer file in tabular format, in which rows represent documents and columns represent words. Intersections of rows and columns are marked when certain documents contain certain words. At the point of retrieval, the computer scans the entire inverted index for documents which contain the words in the search query. In an inverted file organisation, two files are always maintained: sequential file and inverted file. An inverted file contains all the potential index terms arranged alphabetically, drawn automatically from the document records according to indexing technique adopted for the purpose. Each index term in the inverted file is associated with the record number(s) in which the index term occurs and it links with the list of records that represents documents. Thus, for each index term in the database the inverted file contains an entry along with a reference list which specifies position(s) in the database where the term appears. Each term may occur in a number of documents.

c) **Indexed Sequential Files**: An indexed sequential file is also an inverted file in which every record in the source file/document file.

d) **Chained Files:** It's a special type of inverted file organisation which supports dynamic and multiple record linking to ensure simple updating process and quick response against query.

e) **Tree Structured Files**: Here records in a file based on the keys can be organized using a tree structure. If the record is too large, only the keys (in the directory) are organized in the tree structure.

f) **B-Tree**: Bayer and McCreight suggested a method of storing keys on a disk by developing what is called a page. Here a block of storage of fixed size used to transfer information between main storage and direct access storage. This is called B-Tree indexing system that allows easy retrieval, insertion, and deletion of records.

11.6.8 Methods of Computerised Indexing

The first step in indexing, both manual and computerised indexing, is to decide on the subject matter of the document. In manual indexing, the indexer would consider the subject matter in terms of answer to a set of questions such as "Does the document deal with a specific product, condition or phenomenon?" Computerised indexing follows a set of processes of analysing frequencies of word patterns and comparing results to other documents in order to assign to subject categories. This requires no understanding of the material being indexed therefore leads to more uniform indexing but this is at the expense of the true meaning being interpreted. A computer program will not understand the meaning of statements and may therefore fail to assign some relevant terms or assign incorrectly. Human indexers focus their attention on certain parts of the document such as the title, abstract, summary and conclusions, as analysing the full text in depth is costly and time consuming. A Computerised system takes away the time limit and allows the entire document to be analysed, but also has the option to be directed to particular parts of the document.

The second stage of indexing involves the translation of the subject analysis into a set of index terms. This can involve extracting from the document or assigning from a controlled vocabulary.

Statistical Method

Statistical method involves taking words directly from the document. It uses natural language and lends itself well to automated techniques where word frequencies are calculated and those with a frequency over a pre-determined threshold are used as index terms. A stop-list containing common words like articles, conjunctions, prepositions and pronouns are excluded as index terms using a 'stop word' file. Automated extraction of terms may lead to loss of meaning of terms by indexing single words as opposed to phrases. Although it is possible to extract commonly occurring phrases, it becomes more difficult if key concepts are inconsistently worded in phrases. The following statistical methods are adopted in measuring the word significance:

a) **Term Frequency method**: In this method, terms (other than common words) to be indexed are those occurring either very frequently (indicating concepts dealt with) in a text or very seldom (indicating a topic mentioned expressly only once or twice in the title or first paragraph but then being referred to by 'it' or 'this' and the like). This method is based on the frequency of occurrence and co-occurrence of terms, using probabilistic model.

b) **Relative Frequency Method:** Terms that occur infrequently may be highly significant for example a new drug may be mentioned infrequently but the novelty of the subject makes any reference significant. One method for allowing rarer terms to be included and common words to be excluded by automated techniques would be a relative frequency approach where frequency of a word in a document is compared to frequency in the database as a whole. Therefore a term that occurs more often in a document than might be expected based on the rest of the database could then be used as an index term, and terms that occur equally frequently throughout will be excluded.

c) **Term weighting method**: Not all the words have the same significance level. Words occurring with high frequency in a document are better discriminators than words of low frequency. Hence, term frequency method (how often a term is use) is sometimes coupled with term weighting in which different degrees of importance is assigned to terms on the basis of what terms are used in a search request or on the basis of where and how terms appear (e.g. in the title, in an abstract, or in the first and/or last paragraph of a text). For this, term weighting schema is prepared. This method is based on statistical principle.

Linguistic Method

A quite a different approach to computerised indexing is by syntactic and semantic analysis of text. The syntactical analysis identifies the grammatical role and relation among the words in the sentence. It is concerned with automatic recognition of significant word order in a phrase or sentence and with inflections, prefixes, and suffixes that indicate grammatical relationships. Semantic analysis approach seeks to analyse noun phrase automatically with the aid of stored dictionaries and other linguistic aids. Syntactic and semantic analyses are often used in conjunction. Natural Language Processing (NLP) is based on these two methods. Linguistic ontologies are used in NLP to assist in the analysis of natural language text.

Artificial Intelligence (AI) based Indexing System

The application of artificial intelligence in information retrieval research basically involves processing of source text to identify the roles of words and phrases and the relationships between them. The results of this processing are used to identify appropriate indexing expressions. The following indexing system researches based on AI are evident:

a) **Natural Language Processing (NLP) based Indexing System:** The basic idea is to process the text of documents to generate indexing terms. Methods focusing on the lexical level, attempting to identify grammatical classes or parts of speech of individual words together with machine readable dictionaries to index documents. At the syntactic level, there have been attempts to examine the interconnections between words. It uses both co-occurrences of pairs of terms and threshold distances (number of words between two terms). The ideas of co-occurrence and threshold distance may be seen as an attempt to address the issue of coordination (relationships between subject descriptors). This approach is based upon a network of expertise or knowledge, the elements of which are associated with the documents represented in the database.

b) **Expert System based Indexing System**: Expert systems have been used in information retrieval research in an attempt to replace what is usually referred to as the expert intermediary. The expertise referred to in this case is that necessary to construct a search appropriate to the user's needs and the functioning of the system. The components therefore include knowledge of how to interrogate the system, how to identify the appropriate search terms and how to link these terms. An expert system shell consists of an appropriate conceptual hierarchy and a dictionary of the appropriate subject area. This is created by analysis of subject information by experts in the subject area. Associated with each node in the hierarchy is relationship information for navigating the system and document references. Searching basically consists of browsing the network. The network nodes are linked by different kinds of relationship.

Self Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

12) Why should we use computerised indexing system in libraries?

....

....

....

....

13) What are the major components and categories of computerised indexing system?

....

....

....

....

14) Distinguish between Optical Disk based Indexing and Online Database Indexing.

....

....

....

....

15) What are the different methods used for organising records in the index file?

....

....

....

....

16) Highlight the different methods used in computerised indexing.

....

....

....

11.7 INDEXING INTERNET RESOURCES

You already know in sub-section 11.6.5 that Internet is evolving as the largest information system in the world and Internet based indexing system is now considered as a major category under computerised indexing system. The global growth in popularity of the World Wide Web (WWW) has been enabled in part by the availability of browser based search tools which in turn have led to an increased demand for indexing techniques and technologies. Information resources accessible through the Web are very much different from the bibliographic records of a conventional system. The Web presents information access problems orders of magnitude greater than any encountered before. The major defect of the Internet as an information source, apart from its sheer size, is the fact that it lacks any form of quality control. Current chaotic situation caused by the "every man his own publisher" phenomenon. Publishers of scholarly books and journals apply reviewing/refereeing procedures that are, at least to some extent, effective in eliminating the most worthless of what is written. The published indexing and abstracting services provide the next level of quality filtering, mostly by choosing the journals, report series, or other publications that they cover on a regular basis. It is obvious that human professional indexing of the entire Web is completely impractical. Selective professional indexing, of course, is possible.

There have been several attempts to organise the resources on the WWW. Some of them have tried to use traditional Library Classification Schemes such as the Library of Congress Classification, the Dewey Decimal Classification and others. However there is a need to assign proper subject headings to them and present them in a logical or hierarchical sequence to cater to the need for browsing.

When we talk about indexing Internet resources, we really mean indexing Web resources or simply Web indexing. It centres round on the following:

a) search engine indexing of the Web,

b) creation of metadata,

c) organisation of Web links by category, and

d) creation of a Website index that looks and functions like a back-of-book index.

11.7.1 Search Engine Indexing

Internet search engines are special sites on the Web that are designed to help people find information stored on other sites. A search engine searches a database of information on the web. It is a tool to help users to locate information available via Internet. There are differences in the ways various search engines work, but they all perform three basic tasks:

- They search the Internet — or select pieces of the Internet — based on important words.
- They keep an index of the words they find, and where they find them.
- They allow users to look for words or combinations of words found in that index.

To find information on the hundreds of millions of Web pages that exist, a search engine employs special software robots, called spiders (other names for these programs are crawler, worm, wanderer, gatherer, and so on), which traverse the web, following links between pages. It builds the list of the words found on Web sites. When a spider is building its lists, the process is called Web crawling. In order to build and maintain a useful list of words, a search engine's spiders have to look at a lot of pages.

How does any spider start its travels over the Web? The usual starting points are lists of heavily used servers and very popular pages. The spider will begin with a popular site, indexing the words on its pages and following every link found within the site. In this way, the spidering system quickly begins to travel, spreading out across the most widely used portions of the Web.

When the spider looked at an HTML page, it took note of two things:

- The words within the page
- Where the words were found

Words occurring in the title, subtitles, meta tags and other positions of relative importance were noted for special consideration during a subsequent user search. Meta tags allow the owner of a page to specify key words and concepts under which the page will be indexed. The meta tags can guide the search engine in choosing which of the several possible meanings for these words is correct.

Once the spiders have completed the task of finding information on Web pages, the search engine must store the information in a way that makes it useful. There are two key components involved in making the gathered data accessible to users: (1) The information stored with the data, and (2) The method by which the information is indexed.

In the simplest case, a search engine could just store the word and the URL where it was found. To make for more useful results, most search engines store more than just the word and URL.

Search Engine Categories

Search engines may be divided into the following main categories:

- **General Search Engine**: It covers a range of services and compiles their own searchable databases on the web. Examples: Google, Alta Vista, etc.
- **Regional Search Engine**: It refers to country specific search engine for locating varied resources region-wise. Examples: Euro Ferret (Europe), Excite uk (UK), etc.
- **Meta-search Engine**: A meta search engine does not use crawler for compiling their own searchable database. These search engines utilise databases maintained by other individual search engines. When a query is put before this type of search engine, it forwards that query to other search engines. Examples of such search engines are: Dog pile, Ask Jeeves, Inference Find, MetaCrawler, Profusion, Surfwax.
- **Subject Specific Search Engine**: It does not attempt to index the entire Web. Instead, it focuses on searching for Websites or pages within a defined subject area, geographical area or type of resource. Examples: Geo index (Geography / Environmental science), Biochemistry Easy Search Tool (Biochemistry). Because this specific search engine aims for depth of coverage within a single area, rather than breadth of coverage across subjects, they are often able to index documents that are not included even in the largest search engines databases. Some examples of subject specific search engines are: , Bioweb (Biotechnology), Scirus (Sc. & Tech.), Medical world search, Health A to Z (Medical Sc.), Math Search

(Mathematical Sc.), Agri Surf (Agricultural Sc.), Law Crawler (Law), KHOJ (India specific search engine), etc.

- **Directory-based Search Engines**: Due to explosion of information over Internet it is felt that the results fetched by search engines are often mixed with unwanted ones. Subject directories, unlike search engines, are created and maintained by human editors, not electronic spiders or robots. Directory editors typically organise directories hierarchically into browsable subject categories and sub-categories. The resources they list are usually annotated. Directories tend to be smaller than search engine databases, typically indexing only the home page or top level pages of a site. They may include a search engine for searching their own directory. It enables the searcher to move from menu to menu, making one selection after another until he gets to the level where the chosen sites are enlisted. These directories offer access to the information that has been classified into certain categories. Examples of such search engines are: Yahoos, Google, Look smart, Magellan, Open Directory, and Project.

Salient points of differences between Subject Directories and Search Engines are as furnished below:

	Subject Directory	**Search Engine**
Coverage	It is a categorised list of websites with brief description, based on submission by Web site owners, scrutinised and edited by professional editors. Contents coverage in a subject directory are fewer as compared with a search engine.	A search engine indexes all the information on the entire web automatically. It deals with specific piece of information, not categories. Hence, contents coverage in a search engine is higher than the search engine.
Browse	Allows searchers to browse resources in the home page of a Web site by predefined subject categories. From there, searcher can explore the site for more specific information.	Does not allow searchers to browse resources by predefined subject categories. It takes the searcher to the exact page on which the words and phrases he/she is looking for appear.
Searching	Searches titles and annotations of categorised resources.	Searches full-text of web pages.

11.7.2 Subject / Information Gateway

With the advent Internet many libraries are looking forward to go online with Internet. Often they are finding the information available over Internet is enormous. For the same they have devised subject-based portals which are known as Subject Gateways. Major idea of Subject gateways came with the inefficiency of search engines as they fail to give pin-pointed information. They employ subject experts and information professionals to select, classify and catalogue Internet resources to aid search and retrieval for their users.

A subject/information gateway is a web site that provides searchable and browsable access to online resources focused around a specific subject. Subject gateway resource descriptions are usually created manually rather than being generated via an automated process. There are two kinds of gateways: library gateways and portals. Library gateways are collections of databases and informational sites, arranged by subject that have been assembled, reviewed and recommended by specialists, usually librarians. These gateway collections support research and reference needs by identifying and pointing to recommended, academically-oriented pages on the Web. Some notable examples of library gateways are Librarians' Index to the Internet, Internet Public Library, Academic Information, WWW Virtual Library, Digital Librarian, Infomine, etc.

11.7.3 Semantic Web

Search engines usually employ statistical methods like frequency of occurrence of words, co-occurrence of words, etc. which retrieve a number of irrelevant hits against a search on the Web. Though some search engines like Google and Yahoo use human edited entries; still they come up with a large number of wrong hits. Tim Berners-Lee introduced the concept of Semantic Web to extend the web with semantic information. The idea behind Semantic Web is to develop such technologies that make the information more meaningful for the machines to process, which in turn makes search and retrieval of information more effective for searchers. The conception of Semantic Web is characterized by developing tools and technologies like languages, standards and protocols so that the Web becomes meaningful.

Most of the technologies involved in the development of the Semantic Web are still in their infancy. Some of them already in use are the URIs (for identifying documents uniquely and globally), XML (for structuring the data semantically), RDF (to base the structure of the documents on a common model base), Ontologies (to define the objects/ entities and the interrelations between these objects/entities), etc.

11.7.4 Taxonomies

Taxonomies can be considered another tool useful for organising web-based information. For example, taxonomies provide an excellent means for organising subject-specific information into an easily navigable format. The utility of taxonomies for displaying web information can also be found when examining Yahoo's site – which uses a taxonomy/ subject hierarchy to classify its indexed information.

Self Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

17) What do you mean by indexing Internet resources?

....

....

....

....

18) What are the different types search engines used in indexing Internet resources?

....

....

....

19) Discuss the methods of search engine indexing.

....

....

....

20) What is Semantic Web?

....

....

....

11.8 SUMMARY

In this Unit we have dealt with different techniques of subject indexing. It begins with a brief discussion on derivative indexing and assignment indexing and is followed by the discussion on different types of pre- and post-coordinate indexing systems. We cannot understand the pre-coordinate indexing properly without being aware of the contributions of C.A. Cutter and J. Kaiser. For this, principles and processes of subject indexing techniques as enunciated by Cutter and Kaisers are discussed. Major pre-coordinate indexing systems like Chain Indexing, PRECIS, and POPSI are discussed with reference to their principles, syntactical and semantic aspects, entry structure and system of references. Objective conditions that led to the development of post-coordinate indexing and its differences with pre-coordinate indexing are explained. Term entry system and item entry system forming parts of the post-coordinate indexing system are also discussed with an emphasis on the operational stages of Uniterm indexing system. Different varieties of keyword indexing are explained. Computerised indexing techniques are explained in terms of its meaning, features, differences with manual indexing, advantages and disadvantages, components, categories, index file organisation and different methods associated with the generation of index entries with the aid of computers. Indexing internet resources with particular reference to search engine indexing and other associated concepts are discussed briefly at the end of this Unit. All indexing techniques are demonstrated with illustrative examples.

11.9 ANSWERS TO SELF CHECK EXERCISES

1) Derived indexing is 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 also known as *extractive indexing*.

   Assigned indexing is also known as 'Concept indexing', because what we are trying to do is to identify concept(s) associated with the content of each document. It is a method of indexing in which a human indexer selects one or more subject headings or descriptors from a list of controlled vocabulary, instead of the title and/or text of a document, to represent the subject(s) of a work.

2) J. Kaiser in his "Systematic Indexing", published in 1911, solved the problems indexing compound subjects through classificatory approach. He pointed out that compound subjects might be analysed by determining the relative significance of the different component terms of compound subject in terms of two fundamental categories: (1) Concrete and (2) Process. According to Kaiser, Concrete refers to Things, place and abstract terms, not signifying any action or process; e.g. gold,

India, Physics, etc. Process refers to (a) Mode of treatment of the subject by the author; e.g. *Cataloguing of films*; (b) An action or process described in the document; e.g. *Cultivation of rice*; and (c) An adjective related to the concrete as component of the subject; e.g. Strength of metal. Kaiser laid a rule that a 'P*rocess'* should follow *'Concrete'.*

3) The concept of 'chain' as the foundation of chain indexing is a structural manifestation of a subject, which refers to the parts constituting a subject and their mutual interrelationship. It is a modulated sequence of sub-classes or isolate ideas.

   Since the chain expressed the modulated sequence more effectively in a notational classification of subjects, this method takes the class number of the document concerned as the base for deriving subject headings.

4) The basic steps in chain indexing are (1) Construction of the class number of the subject of the document; (2) Representation of the Class Number in the form of a Chain; (3) Determination of links like Sought Links (SL), Unsought Links (USL), False Links (FL), and Missing Links (ML); (4) Preparation of Specific Subject Heading; (5) Preparation of Subject Reference Headings; (6) Preparation of Subject Reference Entries; (7) Preparation of Cross References, if any; and (8) Alphabetisation.

5) Principle of context dependency in PRECIS is seen as a combination of context and dependency. When this principle is followed in a PRECIS input string, each term is qualified and sets the next term into its wider context. In other words, the meaning of each term in the string depends upon the meaning of its preceding term and taken together, they all represent the single context. Each term is hence dependent, directly or indirectly, on all the terms which precede it.

6) Two-line-three part entry structure is followed in PRECIS. The first line is occupied by two parts—Lead and Qualifier, which together constitute the Heading. Lead is occupied by the approach term and Qualifier position is occupied by the term(s) that sets the Lead into wider context. The second line is occupied by the third part, i.e. Display position which consists of those additional set of qualifying terms, which rely upon the heading for their context.

   Index entries in PRECIS are basically generated in three formats: (a) Standard format— entries are generated when any of the primary operators (0), (1), and (2) or its dependent elements appear in the Lead; (b) Predicate transformation— entries are generated under a term coded (3) that immediately follows a term coded either by (2) or (s) or (t); and (c) Inverted format— entries are generated whenever a term coded by an operator in the range from (4) to (6) or its dependent elements appear in the Lead.

7) Different operational stages of POPSI may be categorised under three components: Analysis, Synthesis, and Permutation. The work of 'Analysis' and 'Synthesis' is primarily based on the postulates associated with the deep structure of SILs for generating organising classification. The task of analysis and synthesis is largely guided by the following POPSI-table. The work of 'Permutation' is based on cyclic permutation of each term-of-approach, either individually or in association with other terms for generating associative classification effect in alphabetical arrangement.

8) The major steps in formulating index entries according to POPSI are (1) Content Analysis, (2) Formalisation, (3) Standardisation, (4) Modulation, (5) Preparation of Entry for Organising Classification, (6) Approach-term Selection, (7) Preparation of Entries of Associative Classification, and (8) Alphabetisation.

9) In pre-coordinate indexing system, coordination of component terms is carried out before the users come to search the index file, i.e. at the time of indexing by the indexer (i.e. at input stage) in anticipation of the users' approach. But in post-coordinate indexing, component terms of a subject are kept separately uncoordinated by the indexer, and the user does the coordination of concepts at the time of searching (i.e. at the output stage).

10) Search devices used to avoid false coordination of terms in Post coordinate indexing are: (a) Use of Pre-coordinated Terms, (b) Links, (c) Roles, and (d) Weighting.

11) Different varieties of keyword indexing include Key Word in Context (KWIC) Indexing, Key Word Out of Context (KWOC) Indexing, Key-Word Augmented-in-Context (KWAC) Indexing, Double KWIC Indexing, KWWC (Key-Word-With-Context) Indexing, KEYTALPHA (Key-Term Alphabetical) Indexing, KLIC (Key-Letter-In-Context) Indexing, and WADEX (Word and Author Index).

12) Library is a collection of databases. For example, Accession register is a database, Card catalogue is a database, Circulation records constitute a database, Member register is a database. We need to add, edit and maintain all these databases regularly. The advantages of computerised indexing system are as follows: Redundancy can be reduced; Inconsistency can be avoided; Data can be shared; Standards can be enforced; Security restrictions can be applied; Integrity can be maintained and conflicting requirement may be solved.

13) A typical computerised indexing system has four major components: 1) Database, 2) Search process, 3) Language of indexing, and 4) User interface.

Computerised indexing systems can be categorised into following four groups on the basis of the varieties of information demands of different user groups: 1) Online Database Indexing, 2) Optical Disk based Database Indexing, 3) OPAC based Indexing, and 4) Indexing Internet Resources.

14) A comparative study of online database indexing and Optical Disk based indexing may be done as follows:

Optical Disk based Indexing	**Online Database Indexing**
1) Locally searchable. No requirement for communication channel.	1) Remotely searchable. Communication network is a must.
2) Cost based procurement.	2) Renewal based procurement.
3) End user searching.	3) Intermediary searching.
4) Offline access. Generally updated on monthly basis.	4) Online access. Updating is daily and sometimes hourly.
5) Contains multimedia database to reference databases.	5) Bibliographic databases those are large and frequently updated.
6) Suitable for homogeneous user environment.	6) Suitable for large number of simultaneous users.

15) Design and development of the computerised indexing system mainly depends on the methods used for organising records in the file. Some of the important logical file organisations are: (1) Sequential File, (2) Inverted File, (3) Indexed Sequential Files, (4) Chained Files, (5) Tree Structured Files, and (6) B-Tree.

16) Different methods of computerised indexing include: (1) Statistical Method, which may consist of (1.1) Term Frequency method, (1.2) Relative Frequency Method, and (1.3) Term weighting method; (2) Linguistic Method, (3) Artificial Intelligence (AI) based Indexing System, which may consist of (2.1) Natural Language Processing (NLP) based Indexing System, and (2.2) Expert System based Indexing System,

17) Indexing Internet resources or Web indexing means the following:

    a) search engine indexing of the Web,

    b) creation of metadata,

    c) organisation of Web links by category, and

    d) creation of a Website index that looks and functions like a back-of-book index.

18) Different types search engines used in indexing Internet resources are: (1) General Search Engine, (2) Regional Search Engine, (3) Meta-search Engine, (4) Subject Specific Search Engine, and (5) Directory-based Search Engines.

19) In order to find information from the Internet, a search engine employs special software robots, called spiders (also called crawler, worm, wanderer, gatherer, etc.), which traverse the web, following links between pages. It builds the list of the words found on Web sites, called Web crawling. Here, the spider begins with a popular site, indexing the words on its pages and following every link found within the site. In this way, the spidering system quickly begins to travel across the Web and takes note of the (a) words within the page, and (b) location of the words - title, subtitles, meta tags and other positions of the Web page. Meta tags allow the owner of a page to specify key words and concepts under which the page will be indexed. The meta tags also guide the search engine in choosing which of the several possible meanings for these words is correct. After completion of finding information on Web pages by the spiders, the search engine stores the information in a way that makes accessible to users. In the simplest case, a search engine just stores the word and the URL where it was found. To make for more useful results, most search engines store more than just the word and URL.

20) The "semantic web" is an approach to extend the web with semantic information to avoid wrong hits. The conception of Semantic Web is characterised by developing tools and technologies like languages, standards and protocols so that the Web becomes meaningful. Technologies involved in the development of the Semantic Web are the Uniform Resource Identifier (URI) for identifying documents uniquely and globally, XML (eXtensible Markup Language) for structuring the data semantically, RDF (Resource Description Framework) to base the structure of the documents on a common model base, Ontologies (to define the objects/entities and the interrelations between these objects/entities), etc.

11.10 KEYWORDS

Action : An elementary category associated with POPSI which refers to an idea denoting the concept of 'doing'. An action may manifest itself as Self Action or External Action.

Assigned Indexing : The process 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. Also known as *Assignment Indexing* and *Concept Indexing*.

Associative Classification : It refers to a classification in which a subject is distinguished from all other subjects based on the reference of how it is associated with other subjects, without reference to its COSSCO relationships. The result of associative classification is always a relative index.

Back-of-the-book Index : An index which shows where exactly in the text of a document a particular concept (denoted by a term) is mentioned, referred to, defined or discussed.

Base : It is a particular manifestation or manifestations of a particular elementary category under which all or major portion of related information are brought together.

Boolean Operators : AND, OR, and NOT. Used to combine search terms. AND finds only records that contain both terms. OR finds records that contain either term. NOT finds records that contain the first term but not the second term.

Chain Indexing : The process of deriving subject index entries based on the extracted vocabulary of a notational scheme of classification. It retains all necessary context but removes unnecessary context.

Classaurus : It is an elementary category-based (faceted) systematic scheme of hierarchical classification in verbal plane incorporating all the necessary features of a conventional information retrieval thesaurus. It is used as vocabulary control device in POPSI.

Coextensive Subject Index : A subject index entry, in which a term, phrase, or a set of terms define precisely the full thought content of the document. Here extension and intension of the ideas are equal to the thought content of the document.

Common Modifier	:	It refers to the name of a place (space), Time, Environment), and Form.
Computerised Indexing	:	A method of indexing in which an algorithm is applied by a computer to the title and/or text of a work to identify and extract words and phrases representing subjects, for use as headings under which entries are made in the index.
Concrete	:	An elementary category suggested by Kaiser to refer to things, place and abstract terms, not signifying any action or process.
Content Designation	:	The act of making a bibliographic record machine readable by encoding its various elements according to a specified scheme.
Core	:	It is a particular manifestation or manifestations of one or more elementary category under which all or major portion of related information are brought together within a recognised Base.
COSSCO Relationship	:	It is a relationship in which COordinate—Superordinate—Subordinate—COllateral (COSSCO) relationships of a subject are shown.
Deep Structure of Subject Indexing Languages (DS-SIL)	:	DS-SIL refers to the logical abstraction of the surface structures of outstanding SILs like Cutter, Dewey, Kaiser and Ranganathan.
Derived Indexing	:	The process of indexing in which terms to be used to represent the content of the document are derived directly from the document itself. Also known as *Derivative Indexing*.
Discipline	:	An elementary category associated with POPSI that includes the conventional fields of study, or any aggregate of such fields, or artificially created fields.
Entity	:	An elementary category associated with POPSI which includes manifestations having perceptual correlates, or only conceptual existence, as contrasted with their properties, and actions performed by them or on them.
False Drops	:	Retrieval of unwanted documents because of the false coordination of terms at the time of searching.
Input String	:	A set of terms arranged according to the role operators which act as instructions to the computer for generating index entries.
Item Entry System	:	A type of post-coordinate indexing system in which It takes the opposite approach to term entry system and prepares a single entry for each

document (item), using a defined physical form, which permits access to the entry from all appropriate headings. Here, items are posted on the term.

Keyword : A term that is chosen, either from the actual text or from the queries of the searcher, that is considered to be a 'key' to finding certain information.

Keyword Indexing : The process of using significant words from a title or an abstract or sometimes from the text of the document as index entries.

KWIC Indexing : Key Word In Context, format for showing index entries within the context in which they occur.

KWOC Indexing : Key Word Out of Context, the use of significant word from titles for subject index entries, each followed by the whole title from which the word was taken.

Meta Search Engine : a program that allows to search across many search engines at once.

Modifier : It refers to a qualifier used to modify any one the elementary categories D, E, A and P associated with POPSI.

Nesting : Grouping terms within parentheses to specify the order in which they will be combined. Terms in the innermost parentheses will be combined and searched first. Without parentheses, terms will be combined in left-to-right order.

Ontology : A formal specification of a representational vocabulary for a shared domain of discourse—definitions of classes, relations, functions, and other objects. Ontologies define data models in terms of classes, subclasses and properties to enhance the functioning of the Web.

Organising Classification : In organising classification compound subjects are based on genus-species, whole-part, and other inter-facet relationships. Organising classification distinguish and rank each subject from all other subjects with reference to its COordinate—Superordinate—Subordinate—COllateral (COSSCO) relationships.

Post-Coordinate Indexing : An indexing model in which terms associated with the content of the document are kept separately in the index file by the indexer and the searcher coordinate coordinates the terms at the time of searching or output stage. Also known as 'coordinate indexing'.

PRECIS : PREserved Context Index System, a subject indexing technique in which an open-ended vocabulary can be organised according to a scheme of role operators, usually for computer manipulation.

Pre-Coordinate Indexing : An indexing model in which terms associated with the content of the document are coordinated by the indexer by following the syntactical rules of given indexing language at the time of indexing or input stage for use in the retrieval of information collection on compound and /or complex concepts.

Process : 'An elementary category suggested by Kaiser to refer to mode of treatment of the subject by the author, an action or process described in the document, and an adjective related to the concrete as component of the subject.

Property : An elementary category associated with POPSI which refers to the idea denoting 'attribute'.

Role Operator : Role operators consist of a set of alpha-numeric notations which specifies the grammatical role or the function of the indexed term and regulates the order of terms in the input string. Role operators and their associated rules also serve as the computer instruction for determining the format, typography and punctuation associated with each index entry.

Search Engine : A retrieval tool on the World Wide Web that, in general, matches keywords input by a user to words found at websites. The more sophisticated search engines may allow other than keyword searching.

Semantic Web : The "semantic web" is an approach to extend the web with semantic information to avoid wrong hits by developing tools and technologies like languages, standards and protocols so that the Web becomes meaningful.

Special Modifiers : A special modifier refers to a qualifier which is used to qualify/modify only one of the elementary categories associated with POPSI.

Stop-word List : The 'stop-word' list refers to a list of words, which have no value for indexing / retrieval. These may include insignificant words like articles (a, an, the), prepositions, conjunctions, pronouns, auxiliary verbs together with such general words as 'aspect', 'different', 'very', etc. Each major search system has defined its own 'stop list'

Subject Analysis : The process of identifying the different component ideas associated with the thought content of the document and establishing the interrelationships among those component ideas.

Subject Gateways : Organized lists of web pages, divided into subject areas by human indexers.

Subject Heading : A word or group of words representing the subject of a document.

Subject Index : A tool that exhibits the analysed contents of the collection of documents (either in the library or database).

Subject Indexing : The process of representing the informational content of the document by analysing its content and translating the result of analysis into an indexing language for creating a surrogate record for it, especially subject access points, in an index.

Term Entry System : A type of post coordinate indexing system in which index entries for a document are made under each of the component terms associated with the thought content of the document. Here, terms are posted on the item.

Web Indexing : Web indexing means providing access points for online information materials, which are available through the use of World Wide Web browsing Software.

World Wide Web (WWW) : a network of many thousands of servers linked together by a common protocol.

11.11 REFERENCES AND FURTHER READING

Austin, Derek. *PRECIS: A Manual of Concept Analysis and Subject Indexing*. 2nd ed. London: British Library Bibliographic Services Division, 1984. Print.

Chakraborty, A. R. and Bhubaneswar Chakraborty. *Indexing: Principles, Processes and Products.* Calcutta: World Press, 1984. Print.

Chowdhury, G. G. *Introduction to Modern Information Retrieval.* 2nd ed. London: Facet Publishing, 2004. Print.

Dykstra, Mary. *PRECIS: A Primer*. Rev. Reprint. Metuchen, NJ & London: Scarecrow Press, 1987. Print.

Foskett, A. C. *Subject Approach to Information*. 5th Ed. London: The Library Association, 1996. Print.

Ghosh, S. B. and J. N Satpathi., *Eds*. *Subject Indexing Systems: Concepts, Methods and Techniques*. Calcutta. IASLIC, 1998. Print.

Guha, B. *Documentation and Information: Services, Techniques and Systems*. 2nd rev ed. Calcutta: World Press, 1983. Print.

International Standards Organization. *Documentation—Methods for Examining Documents, Determining their Subjects, and Selecting Indexing Terms,* 1985. Print.

Jones, Karen Sparck, Ed. *Information Retrieval Experiment.* London: Butterworth, 1981. Print.

Lancaster, F. W. "Do Indexing and Abstracting have a Future"? *Anales de Documentation*, no 6, 2003, 137-144. Print.

Olson, Hope A. *Subject Analysis in Online Catalogs*. 2nd edition. Englewood: Libraries Unlimited, 2001. Print.

Sarkhel, Juran Krishna. *Information Analysis in Theory and Practice*. Kolkata: Classique Books, 2001. Print.

Sarkhel, Juran Krishna. "Subject Indexing, Vocabulary Control and Recent Developments in Cataloguing". *Library Cataloguing Theory*. (BLIS 04)New Delhi, Indira Gandhi National Open University, 1999. 180p. Print.

Taylor, Arlene G. *Introduction to Cataloguing and Classification*. 10th edition. New Delhi: Atlantic Publishers, 2007. Print.

Wellisch, H. H. *Indexing from A to Z.* 2nd. ed. New York: H. W. Wilson, 1995. Print.

Wilson, T. D. *An Introduction to Chain Indexing*. Hamden, Conn.: Linnet Books, 1971. (Programmed texts in library and information science). Print.

BLOCK 4 RECENT DEVELOPMENTS

Introduction

This Block is devoted to the developments taking place in organising and managing information. Knowledge organisation, as a bibliographic control oriented concept, has a long history. Originating in ancient time, it has developed in stages reflecting changing societal information demands. It began as a straight forward list of items, often without any order. Later it took the shape of inventory lists for recording the stock against serial numbers followed by adding the location marks for retrieval. Cataloguing as collocating device was introduced in the 19th century. From the second half of the 20th century information technology has taken the lead, backed by computerised database, the Internet technology, Web resources, and new Metadata tools, in achieving the mission and objectives of Knowledge Organisation. Today, in the 21st century a social dimension has been added to take care of the personal informational requirements of individuals, besides the collective demands of the human societies.

Unit 12, titled **Conceptual Changes: Impact of Technology** is spread over six different topics touching upon different aspects of Knowledge Organisation Systems (KOS). These include: Origin and types of KOS, Analysis and planning of KOS, Methods of linking dispersed digital resources, Methods of accessing heterogeneous networked resources, Contributions of W3C in developing standards like Description Framework (RDF), Web Ontology Language (OWL), and particularly, Simple Knowledge Organisation Systems (SKOS). Lastly, the future of semantic activities is discussed in the context of some identified problems yet to be resolved.

Unit 13 **Online Cataloguing: Design and Service** begins by reviewing the development of library catalogue from its early stage up to its latest manifestation in networking environment. It is followed by a discussion on cataloguing standards. The internal structure of a MARC record is examined taking the requirements of bibliographic fields in view. Next, we observed the functionality of few metadata tools in building online catalogues, including AACR2/RDA. Then, OPAC as user interface is discussed in the context of advancing Web 2.0 technologies. Towards the end the requirements of 'copy cataloguing' and 'original cataloguing' and the features of various utilities, offered by network services for generation, conversion, validation of MARC records, and other supports have been explained in detail. This gives an exposure to the different tasks associated with cataloguing of networked resources, apart from awareness of the utility services as such.

The last Unit, no. 14 in this Block is **Overview of Web Indexing, Metadata, Interoperability and Ontologies**. The concept of web indexing and its importance has been discussed in the Unit followed by a description of its types. The concept of metadata has been explained, delineating its different types. Interoperability is very important keeping in view the disparities in the systems used and their frequent interaction and sharing of information with each other. The Unit is devoted to a discussion on its need, methods and protocols for achieving interoperability.

UNIT 12 CONCEPTUAL CHANGES: IMPACT OF TECHNOLOGY

Structure

12.0 OBJECTIVES

After reading this Unit, you will be able to:

- express the meaning and purposes of Knowledge Organisation Systems (KOS);
- explain the analytical principles for organisation of the intellectual content of information records;
- describe the methods of creating and providing access to records; and
- organise knowledge and information resources meaningfully and purposively from the perspective of a networked environment.

12.1 KNOWLEDGE ORGANISATION: AN OVERVIEW

Knowledge Organisation (KO) is a field of investigation closely related to Information Retrieval (IR). An introductory discussion on KO demands a clear understanding about knowledge, information and data, and their hierarchic relationship.

12.1.1 Knowledge Hierarchy

Information is one of the most frequently used words in our society, and widely differs in its meaning – from news to intelligence, data to knowledge. However, in the context of library and information science, information is understood as something more than data. Data is unrelated whereas information is related.

Information reflects an organisation of data. Characteristically, it provides verifiable statements of facts that can be either true or false. Knowledge, on the other hand, belongs to a higher plane of organisation, assumes a verified status of truth or falsity, and comprehends generalised pattern of information.

The following diagram represents the transitions from data, to information, to knowledge, and ultimately to wisdom. Transition from a lower stage to higher up is carried on through 'understanding', which has no separate level of its own. The structural and functional relationships between data, information, knowledge, and wisdom are represented in the following Data-Information-Knowledge-Wisdom (DIKW) model:

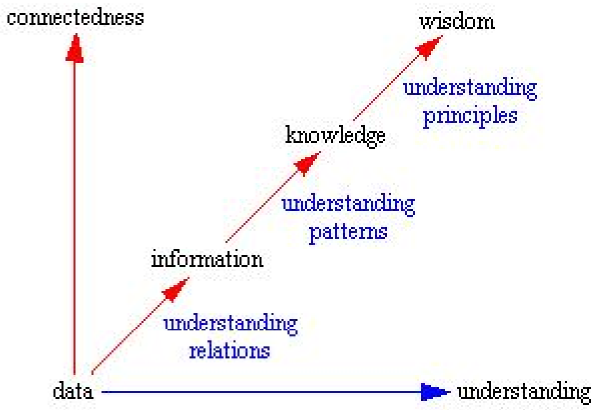


Fig. 12.1: DIKW model

Source: Gene Bellinger http://www.systems-thinking.org/dikw/dikw.htm

12.1.2 Knowledge Organisation : Concept

Knowledge Organisation, often referred to as KO, is a field of investigation within Library and Information Science (LIS). As already mentioned, KO is directly related to Information Retrieval (IR) – the science of searching for documents, for information within documents and for metadata about documents. KO investigates into the nature and order of knowledge, and primarily concerned with *grouping of like things*, documents, for information within documents and for metadata about documents. KO investigates into the nature and order of knowledge, and primarily concerned with *grouping of related things*.

In context of LIS, the word, *things,* may mean:

- Physical documents, or

- Parts of the documents, or
- Representations (Surrogates) of the documents, or
- Representations (Surrogates) of the parts of the documents, as well as
- Concepts (Metadata) used to characterise documents.

Any grouping of objects is based on conceptualisations of the things to be classified.

Example

Suppose we classify *Peacock* as 'National birds', *and not* as 'Peafowl'. In that case, documents about peacocks are grouped with all other documents on National birds, of variant species. Moreover, following the order of knowledge organisation, we may expect to find groupings on national animals, national flowers, and other national *things* in the neighbourhoods.

The example shows:

- The distinction between classifying *things* and classifying documents or concepts is of little significance theoretically. By implication, KO is about the grouping of related *concepts*.
- Our worldview is always dependent on our pre-understandings and conceptualisations of the objects in the world (e.g. National birds, National flowers). We interact with our world with pre-conceived ideas and objects.
- Conceptualisation of things is a culture-specific process, dependant on the cultural outlook and domains of the people. This implies that people from different walks of life with different cultural background tend to classify things differently.
- From socio-cognitive perspective, concepts are *shared classifications.* When people imbibe a culture, learn a language, or embrace a profession, they inherit some common outlook for classifying some part of the world, e.g. public transports.
- Concepts facilitate user interaction with natural and cultural environment. Concepts are generally functional, and they serve pragmatic ends.
- The important question from KO perspective is – which concepts should be preferred in attaining the goal of the information system.
- The success or failure in attaining the goal rests on the selection of controlled vocabularies, and classifications for a certain concept (or meaning) and repression of the alternative concepts/ meanings/ views.

Knowledge Organisation Helps Information Retrieval

As discussed already, Information Retrieval (IR) and KO are closely related fields within LIS. IR relies heavily on some form of KO. It is interesting to note that we create some KO when information is stored (e.g. Class Number, Book Number, Subject Headings), and some KO is created on the fly. For instance, computer information retrieval generates frequency of usage of the search terms that helps to search for information. IR is an intellectual device for providing access to anything stored anywhere, yet the order of which might not be evident. KO, by navigating natural orders, and creating and imposing useful orders helps IR to reach its goal.

Self Check Exercises

Note: 1) Write your answers in the space given below.

2) Check your answers with the answers given at the end of this Unit.

1) What is the continuum of data, information and knowledge? How those are mutually related?

....

....

....

....

2) What is a KOS? What are the common characteristics of KOSs?

....

....

....

....

12.2 KNOWLEDGE ORGANISATION SYSTEMS: ORIGIN AND TYPES

12.2.1 Knowledge Organisation in the Pre-Digital Age

Knowledge organisation, as a bibliographic control oriented concept, has a long history. Knowledge organisation originated in ancient time. Vedas (c 500-400 BCE) is perhaps the oldest specimen of categorisation of human knowledge in four categories. In the Han Dynasty, a library classification system was formulated. In the 19th century, Thomas Jefferson devised his classification system on Baconian methods. Anthony Panizzi's cataloguing principles and Charles Ammi Cutter's Rules for Dictionary catalogues culminated in Paris Principles, and AACR2.

Hodge (2000) pointed out that there has been always a need in a traditional library situation to store a single item at a single location on a shelf. All the bibliographic classification schemes, like LC, DDC, CC, and UDC were developed to serve this purpose. To provide multiple access points libraries use subject heading schemes such as LCSH, Sears, or other specialised schemes at micro level for specific subject areas. Libraries developed Authority Files to control variant form of personal, organisational, and geographic names for searching and browsing local collections.

All these semantic schemes, initially constructed to control print-dominated resources, made enormous contribution in developing Knowledge Organisation Systems to make the networked resources accessible.

Five Main Stages in History of Knowledge Organisation

By examining the trends in scholarly persuasions under the socio-economic and technological influences, we have traced the stages of KO's development with their characteristic features.

- Antiquity: Lists

- Middle Ages: Inventories
- Seventeenth Century: Finding lists
- Nineteenth Century: Collocating devices
- Twentieth Century: Automation and Codification
- Twenty-first Century: Social and Collaborative Tagging

Each represents a turning point that reflects changing societal information demands and the development of new technologies. At the beginning, there were only straight forward list of items, often without any order. During the Middle Ages, making of inventory lists were introduced for recording the stock against serial numbers. Then in the next phase, the location marks were added to the inventory for retrieval. Cataloguing as collocating device was introduced in the 19th century. From the second half of the 20th century information technology has taken the lead, backed by computerised database, the Internet technology, Web resources, and new Metadata tools, in achieving the mission and objectives of Knowledge Organisation. Today, in the 21st century a social dimension has been added to take care of the personal informational requirements of individuals, besides the collective demands of the human societies.

Knowledge Organisation: More than Bibliographic Control

Today the term 'Bibliographic Control' is gradually fading out for more than one reason. First, it stems from a print-dominated book-oriented world. Second, the conventional bibliographic tools are losing their relevance to networked resources. Though, KOs and Bibliographic tools are essentially same, they are different in certain respects:

- Knowledge Organisation is much broader than simply bibliographic control.
- Knowledge Organisation is concerned with understanding how knowledge is generated and used.
- Such knowledge helps us employ relatively more sophisticated approaches to information retrieval.

Knowledge Organisation Systems (KOS)

KOS refers to the semantic tools that present the organised interpretation of knowledge structures. In this broad sense, libraries, encyclopaedias, academic disciplines and such other knowledge organisation systems may serve as examples of KOS. However, it is all important for the development of knowledge organisation, as an intelligent system, to know how far the traditional semantic tools and schemes are relevant and effective in networked resource environment.

The term 'Knowledge Organisation Systems' was coined by *the Networked Knowledge Organisation Systems Working Group* at *the ACM Digital Libraries 098 Conference* in Pittsburgh, Pennsylvania. KOS does not include anything more than what KO does, other than its emphasis on *system*. In a general way, KOS refers to the semantic tools that present the organised interpretation of knowledge structures. In this broad sense, libraries, encyclopedias, academic disciplines and such other knowledge organisation systems may serve as examples of KOS.

Gail Hodge (2000), one of the renowned exponents of KOS, writes:

"The term, *knowledge organisation systems,* is intended to encompass all types of schemes for organising information and promoting knowledge management. Knowledge organisation systems include classification and categorisation schemes that organise materials at a general level, subject headings that provide more detailed access, and authority files that control variant versions of key information such as geographic names and personal names. Knowledge organisation systems also include highly structured vocabularies, such as thesauri, and less traditional schemes, such as semantic networks and ontology. Because knowledge organisation systems are mechanisms for organising information, they are at the heart of every library, museum, and archive."

Common Characteristics of Knowledge Organisation Systems

KOS imposes a particular view of the world.

- The same entity can be characterised in different ways depending on the KOS that is used.
- There must be a sufficient commonality between the concept in KOS and the real world objects it refers.
- A person seeking relevant material by using a KOS must be able to connect his or her concept with its representation in the system.

KOS imposes a particular view of the world on a particular collection through:

- Providing a controlled list.
- Controlling synonyms or equivalents.
- Linking DL (digital library) resources to related resources.
- Making semantic relationships explicit.

Knowledge Organisation Systems for Digital Libraries

- KOS is intended to encompass all types of schemes for organising information and promoting knowledge management.
- Includes traditional classification schemes, subject headings, thesauri, etc.
- Includes less traditional schemes such as semantic networks and ontology
- All digital libraries use one or more KOS.

Hodge points out that there can not be a single knowledge classification scheme on which everyone agrees. A single KOS would have been advantageous, if ever be developed. Cultural diversity makes such an ideal KOS unattainable. This is because of the difference in cultural values and social attitudes, something considered meaningful to one community, not necessarily meaningful to another. Therefore, we live in a world of multiple, variant ways to organise knowledge. (Lesk 1997)

Despite their diversity, we can identify some common characteristics of KOS critically important for their use in organising digital libraries.

Knowledge Organisation for Web Resources

Simple Knowledge Organisation System, or SKOS, provides a model representing the basic structure and the content of *concept schemes*, which may be thesauri, classification schemes, and subject heading lists, taxonomies, folksonomies, and other similar types of controlled vocabulary. SKOS is an application of the Resource Description Framework (RDF), and it allows concepts to be composed and published on the World Wide Web, linked with data on the Web and integrated into other concept schemes.

12.2.2 Knowledge Organisation Systems: Types

There are different systems possible for orgainising digital libraries. Given below their descriptions are based on characteristics such as structure and complexity, relationships among terms, and historical function. KOS are grouped into three general categories:

Term Lists

Authority Files

Authority Files are lists of terms that are used to control the variant names for an entity or the domain value for a particular field. Examples include names for countries, individuals, and organisations. The presentation may be alphabetical or organised by a shallow classification scheme. Specific examples of authority files include the Library of Congress Name Authority File, Sears List of Subject Headings, and the Getty Geographic Authority File.

Glossaries

A glossary is a list of terms, usually with definitions. The terms may be from a specific subject field, or those used in a particular work. Examples include the EPA Terms of the Environment, ALA glossary of library and information science.

Gazetteers

A gazetteer is a dictionary of place names. Traditional gazetteers have been published as books or they appear as indexes to atlases. Each entry may also be identified by feature type, such as river, city, or school. An example is the Geographic Names Information Service, The Columbia Gazetteer of the World.

Dictionaries

Dictionaries are alphabetical lists of terms and their definitions that provide variant senses for each term, where applicable. A dictionary also provides sometimes synonyms and, through definitions, related terms. There was, however, no explicit hierarchical structure or attempt to group terms by concept.

Classification and Categorization

Subject Headings

Subject heading lists can be extensive, covering a broad range of subjects. However, the structure of these lists is generally shallow, with limited hierarchy. The subject headings are pre-coordinated, with rules for constructing headings narrowing down the scope of a concept. Examples include the Medical Subject Headings (MeSH) and the Library of Congress Subject Headings (LCSH).

Classification Schemes, Taxonomies

In general, these types of KOSs provide ways to separate entities into *buckets* or relatively broad topics. Some examples provide a hierarchical arrangement of numeric or alphabetic notations to represent broad topics. These types of KOSs lack the explicit relationships presented in a thesaurus. Examples of well-liked classification schemes include LCCS, DDC, and UDC. Subject categories are often used to group thesaurus terms in broad topic sets, outside the hierarchical scheme of the thesaurus.

Taxonomies are increasingly being used in object oriented design and knowledge management systems to indicate any grouping of objects based on a particular characteristic.

Relationship Group

Thesauri

These KOSs are based on concepts, and they show relationships between terms. Relationships commonly expressed in a thesaurus include hierarchy, equivalence, and associative (or related). These relationships are generally represented by the notation of broader term, narrower term, synonym, and associative or related terms. Associative Relationships may be more granular in some schemes. For example, the Unified Medical Language System (UMLS) provides over 40 associative relationships. Entry terms point to the preferred terms that are to be used for each concept. Most were developed for a specific discipline, for example, FAO Aquatic Sciences and Fisheries Thesaurus, NASA Thesaurus for aeronautics and aerospace-related topics.

Very recently, a new kind of thesauri has come up on the Web named *Visual Thesaurus*, which allows you to discover the connections between words in a visually captivating display.

Semantic Networks

Semantic network is one of the most significant developments in the area of natural language processing technology. Here the concepts and terms not positioned as hierarchies but as a network or a Web. Concepts are thought of as nodes with various relationships branching out from them. These include specific relationships of whole-part, cause-effect, parent-child, etc, instead of standard BT, NT and RT. Princeton's WorldNet is a grand example of semantic network.

Ontologies

In general, ontology is the study or concern about what kinds of things exist – what entities there are in the universe. In information technology, ontology is the working model of entities and interactions in some particular domain of knowledge or practices, such as electronic commerce or "the activity of planning." We may define Ontology as specification of conceptualisations, used to help programs and humans share knowledge. Ontologies are being developed as specific concept models by the Knowledge Management community.

Self Check Exercises

Note: 1) Write your answers in the space given below.

2) Check your answers with the answers given at the end of this Unit.

3) Knowledge Organisation Systems have evolved in stages. Describe those stages highlighting their characteristics.

....

....

....

....

4) Name the types of Knowledge Organisation Systems. State briefly what you understand about Semantic Networks and Ontologies.

....

....

....

....

12.3 PLANNING KNOWLEDGE ORGANISATION SYSTEMS

12.3.1 Analysing User Needs

For any type of library, print-dominated or digital, process of planning starts from assessment of users' needs in terms of content and functionality. Everything else, in terms of resources, and service programs, grows out of the users' needs. The digital library is a *control zone*. It means that a digital library " is always a highly selective subset of available information objects, segregated and favoured, to which access is enhanced and to which the attention of client-users is drawn in opposition to objects excluded." (Atkinson 1996)

An important characteristic feature of a digital library is that it has in-house collections of its own and unbound networked resources within its reach. Therefore, apart from the needs for organising the in-house collections, planning must be done for establishing possible links between contents within and outside the digital library walls. Since KOSs are to act as intermediate authority files, links must always show their relationship to user needs.

12.3.2 Locating Knowledge Organisation Systems

After mapping the user needs decisively, the next task is to locate suitable KOSs, if any that exist. It is preferable to use an existing KOS. Because, building KOSs is costly and time-consuming proposition. The value of a KOS comes from its user community when they appreciate it. By creating a control zone, that is, by selecting some objects and excluding others, information professionals are using their expertise to bring users to documents that hold a particular value. The selection policy must acknowledge the users' behaviour. In users' assessments, the sources built by learned societies, professional associations, or standards groups, stand out more trustworthy than those built in-house.

We must recognise the fact that the networked environment has resulted in both an explosion of primary materials, including documents, electronic journals, and databases, and in an equivalent explosion of KOSs on the Web.

12.3.3 Planning the Infrastructure

Closed Zone

In the context of the digital library, the physical location of the KOS is a critical factor for deciding upon the architecture of KOS. It is on the physical location the position of KOS depends – whether it will be held externally or internally. Both have their own advantages and disadvantages. If KOS is held internally, that is in controlled zone; classification, thesaurus, and ontologies need to be hand-picked cautiously by applying selection criteria, professional expertise and users' feedback. We should safeguard against all possible risks due to the following:

- Selection might limit the access.

This may happen due to various reasons; e.g., when a less inefficient classification scheme is chosen for shelving arrangement, when class numbers are pre-coordinated leaving little scope for libraries to accommodate new concepts; when a chosen subject heading scheme fails to provide unaffected distribution of concepts, e.g. disproportionate distribution of world religions, or other cultural bias; when primary search terms and their spellings are not guided by international standards; and the like shortcomings of KOSs.

- Selection might be biased.

  This may happen when there is no well-documented articulate selection policy to follow; or when standard selection criteria are ignored to give preference to a biased judgment.

- Knowledge organisation supports pre-existing concepts, not for new concepts.

  The traditional semantic tools are inflexible and restricted to the limit of possibilities they have had at starting. It must be ensured that KOSs have in-built mechanisms to accommodate new concepts, terminology, and even search features to perform unaffectedly.

- Not "user-oriented" – individual user's needs are different.

  This may be looked upon as most unfortunate failure from the view point of prevailing Library 2.0 principles where importance of individual requirements of a user have been duly stressed. Libraries must take cognition of the personal needs as well as the collective needs of their users while planning for KOSs.

  On the positive side, the KOS is under more local control. Therefore, it may be possible to improve the response time by not accessing the KOS over the Internet. If the KOS is to be used behind the scenes, concerns of speed and integration become more important. If additional modifications (including digitisation) need to be made to the KOS to integrate it with the digital library, it will also be necessary to load the KOS locally.

Open Zone

If the system is available in open zone on the Web, KOS as an external system, its architecture requires a script to locate the resource. One must then launch a query against the resource to obtain the piece of information that will serve as the *key* between the two files. This key could be

- A universal resource locator (URL) or
- Input to another search query.

A query may be necessary if the KOS is stored in a database. The script may transfer log-on information (including user ID and password) from the digital library system to the external KOS, in order to provide access to the Web-enabled database. In the case of a more direct link, the access may be by URL. Another alternative is the use of other Uniform Resource Identification (URI) schemes and the Uniform Resource Name (URN), which can be sent from the newer Web browsers.

The benefits of linking to a remote resource are that:

- The resource will always be up-to-date.
- The maintenance of the KOS is in the hands of the owner, not the digital librarian.

- It may also be more apparent to users that the KOS is not owned by the digital library.

12.3.4 Change, Upgradation and Version Updation

Like any social or intellectual organisation, an unexpected change in the organisation and content of the system poses problems. Therefore, there should necessarily be a planning process to continue behind the scene to predict and implement changes systematically. Few areas warrant constant review. The software or telecommunications route between the digital library server and the KOS may prove unreliable. The KOS may be obtained from the owner and found tricky when loaded locally. In many cases, this requires licensing that may not be required when the KOS is accessed remotely. Loading a KOS locally also involves issues such as maintenance, local system administration, and disk storage. If the KOS uses special software, such as a database management system, loading the KOS locally will require a copy of that software, which may involves added cost toward purchase or licensing. Besides KOS installation, there are more systems related problems, like firewalls and interface design.

For a digital library, an outdated KOS *can be more of a hindrance than a benefit*. When planning for installation, issues related to maintenance of the KOSs should be settled and blue printed. Version control of the KOS is extremely important. Reloading a new version from the system provider is one way to accommodate changes; however, this may not be acceptable if the locally held version differs substantially from the one currently being in operation. If there has been a significant transformation through customisation procedures, it may prove difficult, if not impossible, to reload the original and recreate the changes that have been made. There is also a marked drift toward developing platform-dependent software. Sometimes, platform changes more frequently than does the software.

There is yet another way of updating KOSs, which is known as *transaction-based approach*, whereby only changed components are transferred by the KOS provider to the library. This, however, requires that the system provider have the necessary infrastructural facilities to effect these transactions on time. In fact, this transaction-based approach is becoming the favourite among version updating options. We now find increasing number of KOS publishers regularly report changes that have taken place since the previous version issued. However, the changes are often not indicated with enough detail to support automatic change transactions.

12.3.5 Presenting the Knowledge Organisation System to the User

Textual Representation

KOS architecture must accommodate the character sets of the incoming sources. This is particularly important if a data string, in ASCII or mark-up language, has been used to represent special characters and diacritical marks. Systems that have been developed in Unicode, which generously accommodate all the existing scripts of languages worldwide have decisively removed the age-old constraints of inter-lingual communications. With the support of Unicode standards, all electronic and web-based KOS platforms are now empowered to communicate directly in original language scripts, that is, without resorting to diacritical marks – a clumsy means of phonetic representations based on extended ASCII. Although the Unicode is now universally accepted standard in the software industry, we still find instances of exceptions. Therefore, the Unicode compliancy needs to be ensured even today when going for a new KOS.

Visibility and Presentation

From the user point of view, the ways a KOS presents itself is no less important than its content is. Because, an ill-designed view front fails to convey the content meaningfully, Moreover, a poor presentation may discourage a person to use the KOS. Therefore, while deciding which KOS should be used and what functions it should serve, the digital library will need to determine how to present the KOS to its users.

The KOS can be exposed to the user in different ways. In the website of the digital library, KOS-related themes or categories may be clustered logically for users' view and interactions. The KOS may be used at a higher level to identify specific portals launched by different user communities. If the content of the digital library includes metadata records, the KOS may be displayed as index terms on the records or as an independent navigation tool. KOS may also remain transparent. For example, a database search procedure may employ a thesaurus behind-the-scene to dig up synonyms for using as multiple search keys. This way, KOSs make presentations of real time statistical information keeping the calculation system behind the scene.

Self Check Exercises

Note: 1) Write your answers in the space given below.

2) Check your answers with the answers given at the end of this Unit.

5) What are the possible risks of Knowledge Organisation Systems in Controlled zone? Suggest safeguards against the risks?

....

....

....

....

6) Determine how to present the Knowledge Organisation Systems to its users.

....

....

....

....

12.4 LINKING INTERRELATED DIGITAL RESOURCES

12.4.1 Relations

Efficacy of KOSs largely depends on their ability of interlinking related digital library resources. KOSs are used for linking digital resources to other digital resources or, indirectly, to physical objects. We may define linking in many ways. We can say in plain words that:

A link is a connection from one page to another destination such as another page, or a different location on the same page.

By means of linking device, identification and categorisation of relations are achieved, which is "a necessary requirement for the formal description that makes navigation possible in the bibliographic universe" (Husby 2001).

The relations may be one of these kinds:

a priori – given by the nature of things. A link is an expression of a relation.

There are many ways of expressing relation. Not all are hypertext links, and some follow quite different methods, such as:

- Citing together

  An article cites more than one item; the cited items are related with each other in reference to the citing source.

- Explicitly stating in text

  The traditional 'See' reference for redirection is a specimen.

- Using controlled vocabularies

  Thesaurus uses the code 'USE' to redirect from uncontrolled to controlled search keys

- Data modeling (relational databases)

  In relational database environment, tables (for Names, Titles, Imprints, etc.) are linked with Record IDs or Pointers, as structured in data model.

- Sharing metadata (identifiers)

  Dublin Core (DC) Metadata serve as links to other web pages, say, generated by same creator, same publisher, or other DC IDs.

- Linking in hypertext

  The Universal Resource Locator (URL), Universal Resource Identifier (URI) are used universally as connecting links between HTML pages, between parts within a HTML page.

In the context of digital resources, the basis for this linking is the identification of information that can be extracted and used to search and locate information within a KOS. This being quite an involved area of discussion, we need a more sophisticated definition of linking to avoid ambiguity and confusion, and admit the following one as a better alternative:

- Relation between Journal and Periodicity is a natural relationship.
- *made up by us*

Relation between an uncontrolled keyword, 'Bharat', to controlled keyword 'India' is a made up relation.

- *deduced from statistics*

Total Number of defaulters reported just-in-time based on circulation statistics.

12.4.2 Types of Linking

Reference Linking

The reference linking is the class of links that can be described as *linking* from metadata (reference, citation) to the full-content.

Some common examples of reference links:

- From an A&I database record to the full-text
- From a citation included in a document to the full-text
- From an OPAC record to an e-journal TOC with further linking possibilities.

Reference links usually target one specific copy of the full-content entity. But the user might rather need or prefer:

- Full content from another supplier
- An OPAC holdings description
- A copy ordering / ILL service
- Another metadata description / abstract
- A book review or access to a net bookshop
- A "full web" search.

Expanding Codes to Full-Text

Coding schemes facilitate communication within a defined group of specialist members. Every discipline has one or more coding schemes to help them communicate precisely and economically. A KOS may use links to connect these coding schemes to the full names for which the code stands. This is an example of static link as opposed to dynamic link that works on-the-fly or just-in-time principle.

The examples provided here include links between databank registration codes and the biological sequence data, and between industrial codes and the full name that the code represents.

Linking to Descriptive Records

Entity names, such as personal and corporate names, location, etc. are linked to additional information about that entity. This was one of the first uses of hyper-linking. KOSs such as dictionaries, glossaries, and classification schemes can be used 'to link the entities in one resource to richer descriptions of that entity in another resource'.

Linking Personal Names to Biographical Information

A common type of authority file is the personal name authority, which controls variants of personal names. For example, the Library of Congress Name Authority File (LCNAF) is used to control variant personal names for authors, editors, artists, and others. The Union List of Artist Names (ULAN), developed by the Getty Vocabulary Program, is another example. Name authorities serve as tools for catalogers and indexers. They ensure use of proper form of name and bringing together all works by or about the person.

A name authority file can also be used to link a bibliographic record or document containing the person's name to a variety of other related materials. If the digital library's resource has a standardised form of the name, it can be identified and searched against the authority file to locate variants. The standardised and variant forms can be joined in a search against a variety of other resources that can provide related information.

For example, in the case of a digital library of images of artists' works, biographical or critical text, a name authority file such as the ULAN or the LCNAF can act as an intermediate file to provide additional information.

Linking Individual Industrial Codes to the Full Scheme

The SIC codes have been used by the U.S. government, economists, financial markets, regulators, and procurement offices to identify manufacturing, agriculture, and service sectors of the economy. The digital library can provide related information by using the authority files for the coding schemes as *a linked authority file*. If a company or economic sector mentioned in the digital library's collection can be linked to an SIC or NAICS code, the code can be searched against the official tables of definitions maintained by the U.S. Census Bureau. These files provide definitions of the codes and place each code in the classification scheme with other economic sectors.

The digital library's content can be further enhanced by making a link between the SIC and NAICS codes. If the digital library resource has the SIC code, it can be extracted and searched against the Census Bureau's *1997 NAICS and 1987 SIC Correspondence Tables.* The table returns the corresponding code from the alternate scheme.

Linking Organism Names to Taxonomic Records

Genus-species names are the Latin names for organisms e.g., plants, animals, and microorganisms. In taxonomy, living organisms are studied and classified. Records are created for each of these organisms. Generally, these records are linked relationally to the other organisms in a hierarchy. Beyond the organism name and the information that it and its placement in the hierarchy convey, taxonomic records use other elements to describe the organism.

These may include distribution patterns, the authority for naming and classification, and the date the organism was identified. Scientists base the information on specimens that are retained because they serve as the physical evidence of the description. Natural history museums, private collections, and individual scientists assign number, or codify the specimens in their collections.

Linking Sequence Numbers to Bio-sequence Databanks

National Center for Biotechnology Information is the most frequently used referenced databanks on the Web. They include GenBank and the Research Collaborators for Structural Bioinformatics Protein Data Bank. Each sequence number is different, but all begin with a persistent code identifying the databank. The link between the literature and the databank is made the following ways. Through a search profile, a text analysis program, or keyword indexing, the text is analysed and the sequence databank numbers identified.

An active link consists of a search strategy can be embedded to locate that sequence number in the databank where the actual sequence is stored. When the user clicks on the active link, the script is generated and launched from the user's browser. The Web-enabled database is searched, and the sequence record is returned to the user.

This type of connection exists between the National Library of Medicine's (NLM) search service, PubMed, and GenBank. If a PubMed search hits records that bear GenBank numbers, an automatic search on GenBank is triggered resulting display of the sequence records.

Linking Chemical Names to Molecular Structures

There are competing systems of nomenclature (i.e., that of the Chemical Abstracts Service [CAS] and of the International Union of Pure and Applied Chemistry) as well as common and commercial synonyms.

BIOSIS, the world's largest not-for-profit producer of biological and biomedical databases uses the chemical registry number to link chemical names with molecular structures. In 1993, BIOSIS began processing its bibliographic citations (titles and keywords) to automatically identify chemical names. BIOSIS assigns CAS Registry Numbers (RNs) to the chemical names identified in this process.

Linking Entity Names to Physical Specimens

Sometimes, we may require going beyond linking the related digital resources, and connect entity names in the digital library resources to physical specimens. Exhibition catalogues describe the art exhibits. Museum catalogues describe objects of art, natural history specimens, and cultural objects. When converted into a computerised database, applications of KOSs become critically important for retrieving information about related objects, and for locating the physical objects.

In natural history community, efforts have been made extensively to create and organise databases of photographs of specimens. The records in their database include the Object Identifiers to facilitate retrieval. The publication of identification codes in the journal literature is also changing. The level of specificity of the identification code changes depending on the biological discipline it belongs to. Vertebrate journals provide the code to the specimen level, Botanical journals tend to list only the institution and the catalogue.

To bring about unified global networked resources for providing unconstrained access worldwide, the following linking strategies are recommended:

1) Use persistent identifiers
2) Use open linking architectures
3) Implement extended services

Self Check Exercises

Note: 1) Write your answers in the space given below.

2) Check your answers with the answers given at the end of this Unit.

7) A link is an expression of relation. Describe the ways links express relationship between two elements. In context of digital resources, how do you define a link?

....

....

....

....

8) Why linking to descriptive records is necessary? How Personal names are linked to Biographical Information? Describe with examples.

....

....

....

....

12.5 UNIVERSAL ACCESS TO HETEROGENEOUS NETWORKED RESOURCES

The Web is the world's largest mass of bits and bytes. It is too big, and going bigger by 1.5 million pages added every day. Here is the *Internet Commons* that embraces many formal and informal Resource Description Communities – the place where disparate communities are to communicate. KOSs are means of connecting heterogeneous resources of these disparate communities.

"Web-based access to digitised images and their descriptions, at anytime from anywhere, lowers the barriers for access to information resources." (Antoine Isaac, 2007)

Over many generations, librarians, curators and archivists have developed controlled vocabularies such as thesauri, classification schemes and ontologies. These Knowledge Organisation Systems (KOS) take 'a term-based approach, where terms from natural language are the first-order elements of a KOS'. Therefore, KOS works well only when the semantics and syntactic structure of the terms are known to the users, which is very unlikely condition. By merging collections without taking care of the semantic heterogeneity, KOSs shift the burden of search to users to obtain their desired objects from variant collections.

KOSs can be used to (1) provide *alternate subject access*, (2) *add modes of understanding* to digital library resources, (3) support *multilingual access*, and (4) supply terms for *expansion of free-text* searches in domains that are relatively unknown to the user.

12.5.1 Alternate Subject Access

Alternate subject access refers to the provision of additional subject orientations that make the resources accessible to different audiences. Instead of using one single conceptual vocabulary for querying or browsing the objects of both collections simultaneously, users are expected and required to use the terminology of the first KOS to identify objects of the first collection, and the second KOS to identify those of the second collection.

This approach is particularly valuable when the digital library resources appeal to groups that do not share a common terminology. It can be a system of subject headings, a classification scheme, or any other subject-oriented system. Alternate subject access can be provided by

- indexing or classifying the resources using multiple schemes,
- retaining original schemes from organisations that contribute to the digital library, or
- mapping between the primary scheme and an alternate scheme.

These Alternate KOSs are not interoperable at the semantic level. To enhance the interoperability we need to solve heterogeneity problems of two types:

- **Representational heterogeneity**: Vocabularies may be presented in different formats; for example, in XML and in plain text. The guiding models may not be compatible. Their general information needs (e.g. 'terms' in thesauri, 'classes' in classification schemes), and KOS may use different kinds of labels for identifying conceptual entities.

- **Conceptual heterogeneity**: In two different vocabularies we find similar concepts having identical meanings attached with different labels or names. (e.g. like "Virgin Mary" and "Madonna"). Also, there will be concepts that are more general than others (e.g. like "Mother" and "Virgin Mary"). By determining the relations between the concepts and their variant labels, or naming, an integrated system can provide users with seamless access to the content described by several vocabularies.

Cataloguing/ Indexing with Multiple Schemes

Classifying and cataloguing the resources with multiple schemes is the most direct method for providing alternate subject access to a collection. The method, however, is expensive, as it involves employment of cataloguers knowledgeable in relative schemes, modifications to the cataloguing tools and procedures, and far more processing time. If, however, the task of cataloguing does not involve cataloguing at the level of individual books but of individual classes of books requiring modifications of labels only, this method may be found quite acceptable.

Indexing from Contributors

The resources of a digital library are built up commonly with contributions from external specialty sources. These sources often follow KOSs developed by themselves for organising their collections. For example, the IEEE Computer Society uses a proprietary classification system, which can be borrowed for using as *alternate system* for classifying resources on computer science. NASA database on aeronautics and astronautics includes relevant bibliographic records from U.S. Department of Defense and US Department of Energy, who permit NASA to use their controlled vocabulary terms to create candidate indexing terms for review by NASA's indexers. The terms collected from other organisations can be viewed as an alternate access point, so that at least part of the collection is accessible through another discipline's terminology.

Mapping Multiple Schemes

Mapping one or more schemes is an indirect method for providing alternate subject access. As reported by Gale Hodge, the experience of BIOSIS, the world's largest private sector abstracting and indexing service in life sciences, illustrated best an application of this method. The records that BIOSIS contributes to TOXLINE database of NLM are processed automatically to have appropriate terms added. This is based on a mapping of the natural language terms that occur in the toxicology literature. BIOSIS' normalised natural language keyword indexing with the MeSH terminology. In the new BIOSIS relational indexing structure, BIOSIS builds and maintains authority files that connect natural language disease names to the MeSH-controlled disease terms. When the BIOSIS indexer assigns the free text keyword for the disease name, the appropriate MeSH term is also added to the record as an alternate access point. (BIOSIS 1999). The assignment is based on the development over time of a mapping between the terminology used by BIOSIS and the MeSH-controlled terms.

In addition to providing alternate access points to BIOSIS products, the inclusion of the MeSH terms makes it possible to perform cross database searching on the indexing field with MEDLINE and other databases that include MeSH terms. The inclusion of terms from an alternate KOS, such as MeSH, therefore supports the use of BIOSIS by medical librarians and practitioners who are familiar with MeSH terminology. Unified Medical Language System (UMLS) is a meta-thesaurus developed by the NLM. It is more extensive that links more than 40 separate KOSs from various medical specialties.

Adding New Modes of Understanding to the Digital Library

Many digital library projects remain text-based, or text-as-image-based, some audio, some video, and increasingly more multimedia-dominated as Internet's capability of presenting information in a variety of other modes goes high up in leaps and bound. KOSs can be used to deal with these new dimensions. In the digital library environment, these dimensions can be viewed as layers that can be added on top of one or more objects. Various tools and services can be developed that are geared to a particular mode. For example, the results of a text search can be presented in graphical or visual form for best satisfaction of the users.

A 'geolibrary', which is defined as a digital library holding 'geoinformation', needs a *geospatial dimension* to be added to provide access by place, called *georeferencing*.

Disciplines like ecology, environmental science, political economy, public health and epidemiology, should be greatly benefited by using KOSs that retrieve geoinformation.a digital library with access to such a digital gazetteer service. Through a geospatial KOS, users can see connections between disparate data, because the data are presented in an alternate mode.

Accessing Multilingual Resources

The problems of accessing multilingual text-based resources are of two kinds, the script and the vocabulary. Until very recently, for writing and reading non-English texts, the knowledge workers had to depend on Romanisation with applications of diacritical marks. It is no more necessary today when we can painlessly write and read, using Unicode, any script of the world in its original form, and sort them all in lexical order.

The problems related to the vocabularies are by far more intricate because of the cultural bearings on their semantic and syntactic properties.

KOSs can support the use of digital libraries by disparate communities providing multilingual access. A variety of sources, including multilingual dictionaries and multilingual thesauri, can support this type of access. One of the most extensive multilingual thesaurus efforts is the Generalised Multilingual Environmental Thesaurus (GEMET) from the European Environment Agency (EEA), The GEMET is available in 12 languages, and plans for more to add.

Expanding Free-Text Search Terms

Free-text searching is the most popular method of discovery on the Web, and in view of some experts, it has more potential for versatile applications than controlled-vocabulary-based searching has. This is mainly because, vocabulary control is highly time-consuming, labour-intensive process and it tends to remain outdated at any point of time.

Apart from a relatively small proportion of metadata and controlled vocabulary, the Web is full of natural language elements. However, variations in natural language make free-text searching problematic. The problem becomes more sensitive when search takes place in interdisciplinary areas, or when the user is not quite familiar with the topic. To overcome these terminology differences, KOSs may help selection of free-text keywords.

The Getty Vocabulary Project emphasises support for searching as a significant application of its vocabularies. Harpring (1999) It reports that the vocabularies are increasingly being used in search engines to look for different terms that refer to the same concept. The Getty vocabularies used in the Art and Architecture Thesaurus, the Union List of Artists Names, and the Thesaurus of Geographic Names, are particularly rich in equivalence relationships. Getty developed a prototype called *a.k.a*. to experiment with the use of equivalence terms to broaden or narrow searches across databases on the Web.

KOSs can be very powerful in supporting free-text searching within digital libraries and in integrating Web resources into existing digital libraries. KOSs have generally been developed for a specific discipline, task, or function, or for the indexing of a specific collection or database. Therefore, depending on the domain in which the KOS is being used and the complexity of the system, it may or may not suggest relevant free-text terms. Expanding a search with related terms, rather than pure synonyms, may return hits that are only peripherally relevant to the user. Hodge (2000)

Self Check Exercise

Note: 1) Write your answers in the space given below.

2) Check your answers with the answers given at the end of this Unit.

9) Where is an alternate subject access the most effective approach? How can its interoperability be enhanced?

....

....

....

....

10) What is cataloguing with multiple schemes? How does this method differ from mapping multiple schemes? Describe with illustrations.

....

....

....

....

12.6 FUTURE OF KNOWLEDGE ORGANISATION SYSTEMS ON THE WEB

12.6.1 Semantic Web, Knowledge Organisation, and Conceptualisation of Things

The World Wide Web, or the Web, is an embodiment of human knowledge. Technically speaking, the Web comprises all the resources and users on the Internet that are using the Hypertext Transfer Protocol (HTTP). The Semantic Web is an extension of the current Web designed to help us to find, share, and combine information more easily through the process of Knowledge Organisation (KO).

Classification and KO is about the *grouping of like things*. In LIS, "things" means physical documents or their parts or their representations as well as concepts used to characterise documents. Any grouping of objects is based on conceptualisations of the things. (Birger Hjørland)

From KO perspective, the goal is decided in controlled vocabularies and classifications. For reaching the goal of the information system, a certain concept (or meaning) is selected while other concepts/meanings/views are repressed. Concepts are generally *functional* in facilitating the users' interaction with the natural and cultural environment, and they serve pragmatic purposes. A concept is a way of classifying some part of the world (e.g. plants).

Libraries and museums of the world carry a long cultural heritage, and inherit a distinguished culture of tool making for organising large collections of objects such as books or museum artifacts. These tools are generally referred to as "knowledge organisation systems" (KOS). Different families of KOSs, such as "thesauri", "classification schemes", "subject heading systems", "gazetteers", "lexical databases", "ontologies", and "taxonomies" are widely recognised and applied in both modern and traditional information systems. KOSs attempt to model the underlying semantic structure of a domain. Modern digital information systems afford more options for mapping and presenting alternative orders of information than traditional physical libraries, and offer more possibilities of presenting information from variant points of interests and discourses. "Thus, the challenge is as much intellectual as technical when we want to develop knowledge organisation systems that are useful and meaningful for the end-users operating in complex, interdisciplinary knowledge domains". (Web4lib at ECDL 2006)

12.6.2 Problem Issues Waiting for a Solution

The future of KOS will depend on getting hold of innovative technologies, and collective and collaborated human endeavors to defeat the intricate problems, particularly, focused on the following issues:

- User-centric design strategies for KOS. The question is:
- How to develop understandable and systematic descriptions of concepts and terms?
- How to show and explain relationships? The challenge is to find the appropriate level of explanation, clarity and conciseness.
- How to achieve these in networked situations?
- KOS Interoperability: Cross-browsing and cross-searching between distributed KOS services, mapping between terms, classes and systems, mapping between KOS and ontologies.
- How to achieve semantic interoperability?
- KOS representations and service protocols: A basic infrastructure is needed in order to achieve programmatic access to KOS services. We need to provide protocols for networked access to a variety of vocabularies for different end users and applications. These require standard representations in formats such as RDF/XML.
- What is the appropriate granularity of base services to apply in evolving Web/Grid environments?
- Why and how is the scalable and sustainable management of KOS mappings required?
- Terminology services: We need to identify and specify terminology services for different applications, within a service-oriented approach/architecture, building on the basic infrastructure.

- Social tagging: Participative user-based approaches to knowledge organisation and cataloguing are emerging and attracting significant community support. Social tagging, which is also known as collaborative tagging, social classification, and social indexing, allows ordinary users to assign keywords, or tags, to items. Unlike traditional classification, social tagging keywords are typically freely chosen instead of using a controlled vocabulary. With the development of Web 2.0, where multi-author contribution will be common, collaborative tagging becomes a popular concept. The tagging metadata contributed by users are used for individual-based activities like searching, filtering, navigating, and group-based social networking like finding people with common interest and so on. Tagging data will help to reveal the knowledge sharing and topic networks based on the relations among tagging words.

- What is the role of social tagging and informal knowledge structures versus established KOS?

12.6.3 Semantic Web Activity

The World Wide Web Consortium (W3C), an international consortium is primarily responsible for the creation and development of Web standards and guidelines. With the development of the Resource Description Framework (RDF), Web Ontology Language (OWL), and Simple Knowledge Organisation System (SKOS), the W3C Semantic Web Activity promotes the deployment of technologies for expressing, exchanging and processing metadata in a form processable by machines. The Dublin Core and related vocabularies of the Dublin Core Metadata Initiative (DCMI) represents a crucial contribution to this growing suite of standards.

The W3C's Semantic Web Activity has stimulated a new field of integrative research and technology development, at the boundaries between database systems, formal logic and the World Wide Web. One facet of the Semantic Web vision is the hope of better organising the vast amounts of unstructured (i.e. human-readable) information in the Web, providing new routes to discovering and sharing that information. [SKOS Reference: 2009]

W3C formally defined RDF and OWL as knowledge representation languages; determined their ways of expressing things meaningfully, and provided structures to information already present in the Web. In addition to precise descriptions, application of these technologies over large bodies of information requires the construction of detailed "maps" of particular domains of knowledge.

12.6.4 Simple Knowledge Organisation System

SKOS is a data sharing standard, bridging several different fields of knowledge, technology and practice. SKOS aims to provide a bridge between disperse communities and the Semantic Web, by transferring existing models of knowledge organisation to the Semantic Web technology context.

Looking to the future, SKOS occupies a position between the exploitation and analysis of unstructured information, the informal and socially-mediated organisation of information on a large scale, and the formal representation of knowledge. It is hoped that, by making the accumulated experience and wisdom of knowledge organisation in the library and information sciences accessible, applicable within and transferable to the technological context of the Semantic Web, in a way that is complementary to existing Semantic Web technology, SKOS will enable many new and valuable applications, and will also lead to new integrative lines of research and development in both technology and practice.

The SKOS data model views a knowledge organisation system as a concept scheme comprising a set of concepts. These SKOS concept schemes and SKOS concepts are identified by URIs, enabling anyone to refer to them unambiguously from any context, and making them a part of the World Wide Web.

SKOS concepts can be labeled with any number of lexical (UNICODE) strings, such as "romantic love" or "れんあい", in any given natural language, such as English or Japanese (written here in hiragana). One of these labels in any given language can be indicated as the "preferred" label for that language, and the others as "alternate" labels. Labels may also be "hidden", which is useful e.g. where a knowledge organisation system is being queried via a text index. Lexical Labels for more on the SKOS lexical labeling properties.

SKOS concepts can be assigned one or more notations, which are lexical codes used to uniquely identify the concept within the scope of a given concept scheme. While URIs are the preferred means of identifying SKOS concepts within computer systems, notations provide a bridge to other systems of identification already in use such as classification codes used in library catalogs.

12.6.5 Future

Future must corroborate deployment of Semantic Web methods in support of Knowledge Organisation systems and services. The assumption underlying semantic digital libraries is that full-text search cannot be the entire solution for the massively expanding information space of emerging digital libraries. Next-generation digital library systems must also be able to handle well-defined metadata describing the stored contents and provide machine support for the end users in their search for content. One crucial first step is to organise bibliographic metadata for automated interpretation by machines. This future perspective of KOS was deliberated in Web4lib at ECDL 2006.

Self Check Exercise

Note: 1) Write your answers in the space given below.

2) Check your answers with the answers given at the end of this Unit.

11) What is W3C? Briefly describe the Semantic Web Activities of W3C.

....

....

....

....

12) Who invented SKOS? What it is and what it is supposed to do? Describe SKOS concepts.

....

....

....

....

12.7 SUMMARY

The theme of this Unit is 'conceptual changes in LIS approaches toward knowledge organisation, as an impact of technology'. We discussed six different topics touching upon different aspects of Knowledge Organisation Systems (KOS).

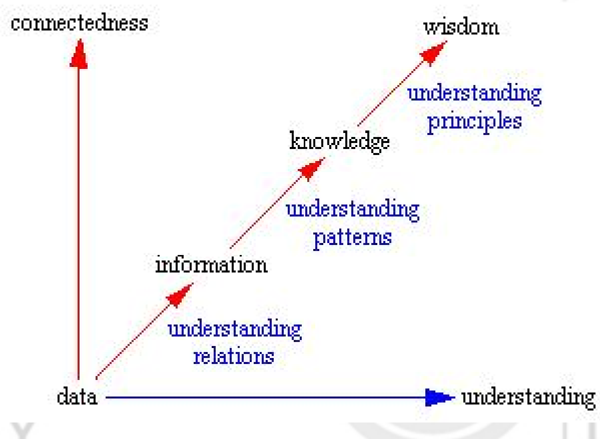
Origin and types of KOS, Analysis and planning of KOS, Methods of linking dispersed digital resources, Methods of accessing heterogeneous networked resources, Contributions of W3C in developing standards like Resource Description Framework (RDF), Web Ontology Language (OWL), and particularly, Simple Knowledge Organisation Systems (SKOS), were highlighted. Lastly, the future of semantic activities is discussed in context of some identified problems yet to be resolved.

12.8 ANSWERS TO SELF CHECK EXERCISES

1) *Information* is one of the most frequently used words in our society, and widely differs in its meaning – from news to intelligence, data to knowledge. However, in context of library and information science, information is understood as something more than data. Data is unrelated, information related.

   Information reflects an organisation of data. Characteristically, it provides verifiable statements of facts that can be either true or false. Knowledge, on the other hand, belongs to a higher plain of organisation, assumes a verified status of truth or falsity, and comprehends generalized pattern of information.

   The following diagram represents the transitions from data, to information, to knowledge, and ultimately to wisdom. Transition from a lower stage to higher up is carried on through 'understanding', which has no separate level of its own. The structural and functional relationships between data, information, knowledge, and wisdom are represented in the following Data-Information-Knowledge-Wisdom (DIKW) model:



2) We may define Knowledge Organisation Systems, in a general way, as semantic tools that present the organised interpretation of knowledge structures. In this broad sense, libraries, encyclopedias, academic disciplines and such other knowledge organisation systems may serve as examples of KOS.

   Common Characteristics in KOS can be described from two view points as follows:

   KOS imposes a particular view of the world.

- The same entity can be characterised in different ways depending on the KOS that is used.
- There must be a sufficient commonality between the concept in KOS and the real world objects it refers.
- A person seeking relevant material by using a KOS must be able to connect his or her concept with its representation in the system.
- KOS imposes a particular view of the world on a particular collection through.
- Providing a controlled list.
- Controlling synonyms or equivalents.
- Linking DL resources to related resources.
- Making semantic relationships explicit.
- Provide a controlled list.
- The KOS imposes a particular view of the world on a collection and the items in it.
- The same entity can be characterised in different ways, depending on the KOS that is used.
- There must be sufficient commonality between the concept expressed in a KOS and the real-world object to which that concept refers that a knowledgeable person could apply the system with reasonable reliability.

3) By examining the trends in scholarly persuasions under the socio-economic and technological influences, we have traced the stages of KO's development with their characteristic features.

Antiquity: Lists

- Middle Ages: Inventories
- Seventeenth Century: Finding lists
- Nineteenth Century: Collocating devices
- Twentieth Century: Automation and Codification
- Twenty-first Century: Social and Collaborative Tagging

Each represents a turning point that reflects changing societal information demands and the development of new technologies. At the beginning, there were only straight forward list of items, often without any order. During the Middle Ages, making of inventory lists were introduced for recording the stock against serial numbers. Then in the next phase, the location marks were added to the inventory for retrieval. Cataloguing as collocating device was introduced in the 19th century. From the second half of the 20th century information technology has taken the lead, backed by computerised database, the Internet technology, Web resources, and new Metadata tools, in achieving the mission and objectives of Knowledge Organisation. Today, in the 21st century a social dimension has been added to take care of the personal informational requirements of individuals, besides the collective demands of the human societies.

4) Generally, KOS may be categorized into three main groups, namely, Term Lists, Classification & categorisation, and Relation Group. The names of KOSs are enlisted below:

   Term Lists

   Authority Files

   Glossaries

   Gazetteer

   Dictionaries

   Classification and Categorisation

   Subject Headings

   Classification Schemes, Taxonomies

   Relationship Group

   Thesauri

   Semantic Networks

   Ontologies

The Semantic Networks and Ontologies are briefly described here:

Semantic Networks

Semantic network is one of the most significant developments in the area of natural language processing technology. Here the concepts and terms not positioned as hierarchies but as a network or a Web. Concepts are thought of as nodes with various relationships branching out from them. These include specific relationships of whole-part, cause-effect, parent-child, etc, instead of standard BT, NT and RT. Princeton's WorldNet is a grand example of semantic network.

Ontologies

In general, ontology is the study or concern about what kinds of things exist - what entities there are in the universe. In information technology, ontology is the working model of entities and interactions in some particular domain of knowledge or practices, such as electronic commerce or "the activity of planning." We may define Ontology as specification of conceptualisations, used to help programs and humans share knowledge. Ontologies are being developed as specific concept models by the Knowledge Management community.

5) Planning for KOS needs to be done carefully to offset all possible risks of unsatisfactory functioning in closed zone environment.

   In the context of the digital library, the physical location of the KOS is a critical factor for deciding upon the architecture of KOS. It is on the physical location the position of KOS depends – whether it will be held externally or internally. There are pros and cons to either approach. If KOS is held internally, that is in controlled zone, Classification, thesaurus, ontologies need to be hand-picked cautiously by applying selection criteria, professional expertise and users' feedback. We should safeguard against all possible risks:

- Selection might limit the access.

This may happen due to various reasons; e.g., when a less inefficient classification scheme is chosen for shelving arrangement, when class numbers are pre-coordinated leaving little scope for libraries to accommodate new concepts; when a chosen subject heading scheme fails to provide unaffected distribution of concepts, e.g. disproportionate distribution of world religions, or other cultural bias; when primary search terms and their spellings are not guided by international standards; and the like shortcomings of KOSs.

- Selection might be biased.

This may happen when there is no well-documented articulate selection policy to follow; or when standard selection criteria are ignored to give preference to a biased judgment.

- Knowledge organisation supports pre-existing concepts, not for new concepts.
- The traditional semantic tools are inflexible and restricted to the limit of possibilities they have had at starting. It must be ensured that KOSs have in-built mechanism to accommodate new concepts, terminology, and even search features to perform unaffectedly.
- Not "user-oriented" – individual user's needs are different.

This may be looked upon as most unfortunate failure from the view point of prevailing Library 2.0 principles where importance of individual requirements of a user has been duly stressed. Libraries must take cognition of the personal needs as well as the collective needs of their users while planning for KOSs.

6) The ways a KOS presents itself is no less important than KOS content. It is true for the textual and graphical presentation on screen, and for the design of the interactive user interface.

Textual Representation

KOS architecture must accommodate the character sets of the incoming sources. This is particularly important if a data string, in ASCII or mark-up language, has been used to represent special characters and diacritical marks. Systems that have been developed in Unicode, which generously accommodate all the existing scripts of languages worldwide have decisively removed the age-old constraints of inter-lingual communications. With the support of Unicode standards, all electronic and web-based KOS platforms are now empowered to communicate directly in original language scripts, that is, without resorting to diacritical marks – a clumsy means of phonetic representations based on extended ASCII. Although the Unicode is now universally accepted standard in the software industry, we still find instances of exceptions. Therefore, the Unicode compliancy needs to be ensured even today when going for a new KOS.

Visibility and Presentation

From the user point of view, the ways a KOS presents itself is no less important than its content is. Because, an ill-designed view front fails to convey the content meaningfully, Moreover, a poor presentation may discourage a person to use the KOS. Therefore, while deciding which KOS should be used and what functions it should serve, the digital library will need to determine how to present the KOS to its users.

The KOS can be exposed to the user in different ways. In the website of the digital library, KOS-related themes or categories may be clustered logically for users' view and interactions. The KOS may be used at a higher level to identify specific portals launched by different user communities. If the content of the digital library includes metadata records, the KOS may be displayed as index terms on the records or as an independent navigation tool. KOS may also remain transparent. For example, a database search procedure may employ a thesaurus behind-the-scene to dig up synonyms for using as multiple search keys. This way, KOSs make presentations of real time statistical information keeping the calculation system behind the scene.

7) There are many ways of expressing relation. Not all are hypertext links, and some follow quite different methods, such as:

- Citing together
- An article cites more than one item; the cited items are related with each other in reference to the citing source.
- Explicitly stating in text.
- The traditional 'See' reference for redirection is a specimen.
- Using controlled vocabularies.
- Thesaurus uses the code 'USE' to redirect from uncontrolled to controlled search keys.
- Data modeling (relational databases).
- In RDB environment, tables (for Names, Titles, Imprints, etc.) are linked with Record IDs or Pointers, as structured in data model.
- Sharing metadata (identifiers).
- Dublin Core (DC) Metadata serve as links to other web pages, say, generated by same creator, same publisher, or other DC IDs.
- Linking in hypertext.
- The Universal Resource Locator (URL), Universal Resource Identifier (URI) are used universally as connecting links between HTML pages, between parts within a HTML page.

In context of digital resources, the basis for this linking is the identification of information that can be extracted and used to search and locate information within a KOS. This being quite an involved area of discussion, we need a more sophisticated definition of linking to avoid ambiguity and confusion, and admit the following one as a better alternative:

A link is a connection from one page to another destination such as another page, or a different location on the same page

8) Linking to Descriptive Records is necessary as because pieces of information related to any particular target item, or entity, are usually stored at different locations, as a part of other records, and those are required to be retrieved by the search system. Entity names, such as personal and corporate names, location, etc. are linked to additional information about that entity. This was one of the first uses of

hyper-linking. KOSs such as dictionaries, glossaries, and classification schemes can be used 'to link the entities in one resource to richer descriptions of that entity in another resource'.

Linking Personal Names to Biographical Information

A common type of authority file is the personal name authority, which controls variants of personal names. For example, the Library of Congress Name Authority File (LCNAF) is used to control variant personal names for authors, editors, artists, and others. The Union List of Artist Names (ULAN), developed by the Getty Vocabulary Program, is another example. Name authorities serve as tools for catalogers and indexers. They ensure use of proper form of name and bringing together all works by or about the person.

A name authority file can also be used to link a bibliographic record or document containing the person's name to a variety of other related materials. If the digital library's resource has a standardised form of the name, it can be identified and searched against the authority file to locate variants. The standardised and variant forms can be joined in a search against a variety of other resources that can provide related information.

For example, in the case of a digital library of images of artists' works, biographical or critical text, a name authority file such as the ULAN or the LCNAF can act as an intermediate file to provide additional information.

9) *Alternate subject access* refers to the provision of additional subject orientations that make the resources accessible to different audiences. This approach is particularly valuable when the digital library resources appeal to groups that do not share a common terminology. It can be a system of subject headings, a classification scheme, or any other subject-oriented system. Alternate subject access can be provided by

- indexing or classifying the resources using multiple schemes,
- retaining original schemes from organisations that contribute to the digital library, or
- mapping between the primary scheme and an alternate scheme
- These Alternate KOSs are not interoperable at the semantic level. To enhance the interoperability we need to solve heterogeneity problems of two types:
- **Representational heterogeneity**: Vocabularies may be presenred in different formats; for example, in XML and in plain text. The guiding models may not be compatible. Their general information needs (e.g. 'terms' in thesauri, 'classes' in classification schemes), and KOS may use different kinds of labels for identifying conceptual entities.
- **Conceptual heterogeneity**: In two different vocabularies we find similar concepts having identical meanings attached with differenr labels or names. (e.g. like "Virgin Mary" and "Madonna"). Also, there will be concepts that are more general than others (e.g. like "Mother" and "Virgin Mary"). By determining the relations between the concepts and their variant labels, or namings, an integrated system can provide users with seamless access to the content described by several vocabularies.

10) **Cataloguing with Multiple Schemes**

Cataloguing the resources with multiple schemes is a direct method for providing alternate subject access to a collection. The method, however, is expensive, as it involves employment of cataloguers knowledgeable in relative schemes, modifications to the cataloging tools and procedures, and far more processing time. If, however, the task of cataloguing does not involve cataloguing at the level of individual books but of individual classes of books requiring modifications of labels only, this method may be found quite acceptable.

Mapping Multiple Schemes

Mapping one or more schemes is an indirect method for providing alternate subject access. As reported by Gale Hodge, the experience of BIOSIS, the world's largest private sector A&I service in life sciences, illustrated best an application of this method . The records that BIOSIS contributes to TOXLINE database of NLM are processed automatically to have appropriate terms added. This is based on a mapping of the natural language terms that occur in the toxicology literature. BIOSIS' normalized natural language keyword indexing with the MeSH terminology. In the new BIOSIS relational indexing structure, BIOSIS builds and maintains authority files that connect natural language disease names to the MeSH-controlled disease terms. When the BIOSIS indexer assigns the free text keyword for the disease name, the appropriate MeSH term is also added to the record as an alternate access point. (BIOSIS 1999). The assignment is based on the development over time of a mapping between the terminology used by BIOSIS and the MeSH-controlled terms.

In addition to providing alternate access points to BIOSIS products, the inclusion of the MeSH terms makes it possible to perform cross database searching on the indexing field with MEDLINE and other databases that include MeSH terms. The inclusion of terms from an alternate KOS, such as MeSH, therefore supports the use of BIOSIS by medical librarians and practitioners who are familiar with MeSH terminology. Unified Medical Language System (UMLS) is a meta-thesaurus developed by the NLM. It is more extensive that links more than 40 separate KOSs from various medical specialties.

11) **Semantic Web Activity**

The World Wide Web Consortium (W3C), an international consortium is primarily responsible for the creation and development of Web standards and guidelines. With the introduction of the Resource Description Framework (RDF), Web Ontology Language (OWL), and Simple Knowledge Organisation System (SKOS), the W3C Semantic Web Activity promotes the deployment of technologies for expressing, exchanging and dealing out metadata in a form that machines can process. The Dublin Core and related vocabularies of the Dublin Core Metadata Initiative (DCMI) represents a crucial contribution to this growing suite of standards.

The W3C's Semantic Web Activity has stimulated a new field of integrative research and technology development, at the boundaries between database systems, formal logic and the World Wide Web. One facet of the Semantic Web vision is the hope of better organising the vast amounts of unstructured (i.e. human-readable) information in the Web, providing new routes to discovering and sharing that information.

W3C formally defined RDF and OWL as knowledge representation languages; determined their ways of expressing things meaningfully, and provided structures to information already present in the Web. In addition to precise descriptions, application of these technologies over large bodies of information requires the construction of detailed "maps" of particular domains of knowledge.

12) **Simple Knowledge Organisation System**

Simple Knowledge Organisation Systems (SKOS) is a data sharing standard, created and developed by W3C. SKOS aims to bridge several different fields of knowledge, technology and practice as well as between disperse communities and the Semantic Web, by transferring existing models of knowledge organisation to the Semantic Web technology context.

Looking to the future, SKOS occupies a position between the exploitation and analysis of unstructured information, the informal and socially-mediated organisation of information on a large scale, and the formal representation of knowledge. It is hoped that, by making the accumulated experience and wisdom of knowledge organisation in the library and information sciences accessible, applicable within and transferable to the technological context of the Semantic Web, in a way that is complementary to existing Semantic Web technology, SKOS will enable many new and valuable applications, and will also lead to new integrative lines of research and development in both technology and practice.

The SKOS data model views a knowledge organisation system as a concept scheme comprising a set of concepts. These SKOS concept schemes and SKOS concepts are identified by URIs, enabling anyone to refer to them unambiguously from any context, and making them a part of the World Wide Web.

SKOS concepts can be labeled with any number of lexical (UNICODE) strings, such as "romantic love" or "रोमांटिक प्रेम", in any given natural language, such as English or Hindi (written here in Nagari). One of these labels in any given language can be indicated as the "preferred" label for that language, and the others as "alternate" labels. Labels may also be "hidden", which is useful e.g. where a knowledge organisation system is being queried via a text index. Lexical Labels for more on the SKOS lexical labeling properties.

SKOS concepts can be assigned one or more notations, which are lexical codes used to uniquely identify the concept within the scope of a given concept scheme. While URIs are the preferred means of identifying SKOS concepts within computer systems, notations provide a bridge to other systems of identification already in use such as classification codes used in library catalogs.

12.9 KEYWORDS

Classification : The operation of grouping elements and establishing relationships between them (or the product of that operation).

Controlled Vocabulary : A collection of preferred terms that are used to assist in more precise retrieval of content. Controlled vocabulary terms can be used for categorising content, building labeling systems, and creating style guides and database schema. One type of a controlled vocabulary is taxonomy

Data Model : A description of data that consists of all entities represented in a data structure or database and

		the relationships that exist among them. It is more concrete than an ontology but more abstract than a database dictionary (the physical representation).
Domain	:	A sphere of knowledge, influence, or activity.
Element	:	An object or concept.
Extensible Markup Language (XML)	:	A W3C standard markup language for documents containing structured information. As opposed to HTML, which is designed specifically for web browsers, XML is designed for much wider use and is extensible to fit each application. XML is the basis for an incredible array of standards that describe everything from messages between systems to security specifications to document structures. The advantage of XML is that it is human understandable and platform independent.
Indexing	:	The intellectual analysis of the subject matter of a document to identify the concepts represented in the document and the allocation of descriptors to allow these concepts to be retrieved. Indexing a large number of documents can be done semi-automatically using software applications.
Metadata	:	A definition or description of data. In data processing, metadata is definitional data that provides information about, or documentation of, other data managed within an application or environment.
Ontology	:	Ontologies resemble faceted taxonomies but use richer semantic relationships among terms and attributes, as well as strict rules about how to specify terms and relationships. Because ontologies do more than just control a vocabulary, they are thought of as knowledge representation. The oft-quoted definition of ontology is "the specification of one's conceptualisation of a knowledge domain."
Relationships	:	A defined linkage between two elements.
Resource Description	:	A W3C standard XML framework for describing.
Framework Framework (RDF)		and interchanging metadata. The simple format of resources, properties, and statements allows RDF to describe robust metadata, such as ontological structures. As opposed to Topic Maps, RDF is more decentralized because the XML is usually stored along with the resources.
Semantic Web	:	The Semantic Web is an extension of the current Web that will allow you to find, share, and combine information more easily. It relies on

		machine-readable information and metadata expressed in RDF.
Taxonomy	:	A classification of elements within a domain
Thesaurus	:	A taxonomy that also includes associated and related terms. It is the most complex type of controlled vocabulary, and is sometimes used to standardise an organisationís terminology and subsequently inform both navigation and search systems.

12.10 REFERENCES AND FURTHER READING

Bean, Carol A. *Relationships in the Organisation of Knowledge.* Ed. Rebecca Green and A. Bean. New York: Springer, 2001.19-35. Print.

Bliss, Henry Evelyn. *The Organisation of Knowledge and the System of the Sciencesý*. New York: H. Holt and Company, 1929. Print.

Fugmann, Robert, ed. "Tools for Knowledge Organisation and the Human Interface: Proceedings". *INDEKS, 1990. 1st International ISKO Conference*, Darmstadt, 14-17 August 1990. Print.

Gilchrist, Alan, ed. *From Classification to "Knowledge Organisation": Dorking Revisited or "Past is Prelude*". The Hague: FID, 1997. Print.

Hodge, Gail M. *Systems of Knowledge Organisation for Digital Libraries: beyond Traditional Authority Files*. Washington DC: Council on Library and Information Resources. Web 24 September 2012. < http://www.clir.org/pubs/reports/pub91/contents.html>

Hjørland, Birger. "Theory of Knowledge Organisation and the Feasibility of Universal Solutions". Presented at the *Eighth International ISKO Conference*, London, July 13-16, 2004. ISKO, 2004. Print.

Husby, Ole. *Linking Resources.* BIBSYS, 2001. Web. 24 September 2012. < http://old.stk.cz/elag2001/Papers/OleHusby/Show.html>

Joachim, Martin D, ed. *Languages of the World: Cataloging Issues and Problems.* New York: Haworth Press, 1994. Print.

Knowledge Organisation and the Global Information Society. Ed. Ia McIlwaine. Wurzburg: Ergon-Verlag, 2005. Print.

Knowledge Organisation and Change. Ed. Rebecca Green. Frankfurt: Indeks Verlag, 1996. Print.

López-Huertas, Mariá J., ed. *Challenges in Knowledge Representation and Organisation for the 21st Century.* Wurzburg: Ergon-Verlag, 2002. Print.

Rowley J.E. and Richard J. Hartley. *Organising Knowledge: An Introduction to Managing Access to Information.* Aldershot: Ashgate, 2008. Print.

Williamson, Nancy Joyce and Clare Beghtol eds. *Knowledge Organisation and Classification in International Information Retrieval.* New York: Haworth Press, 2004. Print.

UNIT 13 ONLINE CATALOGUES: DESIGN AND SERVICES

Structure

13.0 OBJECTIVES

After reading this Unit, you will be able to:

- define an online library catalogue;
- explain its functions and principles;
- discuss changes in its working; and
- design an online library catalogue.

13.1 PHYSICAL CATALOGUE TO OPAC: CHANGING PERSPECTIVES

13.1.1 Physical Catalogue

Around 4000 years ago, libraries were there in Sumer and other ancient regions. Those libraries stored collections of stone tablets, and had their catalogues as well. Early catalogues were mostly inventory lists, all hand written until libraries stared printing their catalogues in book form. The popularity of printed catalogue grew for their easy handling and storage. However, the librarians experienced difficulties in keeping their catalogues up-to-date. Production of printed catalogue was time-consuming, labour-intensive, and expensive.

Librarians found solution in card catalogue. In 1860, Harvard Librarian John Langdon Sibley introduced a card catalogue for public use. Card catalogue stayed for a century as the most effective form of catalogue.

In early 1960s, Library of Congress developed a system of machine-readable catalogue with a new set of rules for machine manipulation. They did it for transferring their massive bibliographic records into electronic format. Although, the development of online catalogue was not their immediate objective, this initiative had paved way to build online catalogue with support of advancing telecommunications computing technology.

It took more than three decades for libraries for reaching the era of online catalogue.

13.1.2 Descriptive Catalogue

Along with the changes in the physical shapes and media, the rules governing the arrangement of bibliographic descriptive elements have been continuously evolving. The First Age of Descriptive Cataloguing began with Panizzi and included Cutter. Michael Gorman labelled it as "the age of the single-author code."

Before, Anthony Panizzi's *Rules for the Compilation of the Catalogue*, printed in 1841, there had been no serious and successful attempt to construct comprehensive code for cataloguing. Many contemporaries of Panizzi like Thomas Carlyle thought of a catalogue as simple listing of authors and book titles. They found his rules unnecessary. The criticism could not prevent him publishing his rules. He thought:

"A reader may know the *work* he requires; but he cannot be expected to know all the peculiarities of different *editions,* and this information he has a right to expect from the catalogues." So here we have two individuals looking at the same object—the book—but seeing different things.

Charles Ammi Cutter his *Rules for a Dictionary Catalogue*, published in 1876, has deeper impact on the development of cataloguing principles and methodologies. We can trace FRBR's development in Cutter's cataloguing code.

Second Age of descriptive cataloguing was "the era of the committee code", as Gorman called. The year 1908 was important in cataloguing: the American Library Association and the Library Association of the United Kingdom published a set of common cataloguing rules. They did not agree on absolutely everything, so separate American and British editions were made, but this was the first set of Anglo-American cataloguing rules.

Third Age of Descriptive cataloguing began in 1967 with the release of A*nglo-American Cataloguing Rules*. It was an international (American, Australian, British, Canadian) standardisation of descriptive cataloguing rules, with a philosophy based on the *Paris Principles*. The *Statement of Principles* passed at the International Conference on Cataloguing Principles held in Paris in 1961. These principles eventually generated a large set of detailed rules into *Anglo-American Cataloguing Rules* (*AACR*) and serve as the basis of other national codes. It was a culmination of most of the major ideas: *an axiomatic approach*, *user needs*, and *standardisation* and *internationalisation*.

Third Age of Descriptive Cataloguing has been continuing *in spirit* till today. In terms of its applicability of its regulatory principles, its future looks little uncertain, particularly in approaching environment of Web 2.0 technology.

13.1.3 Standards

The three cataloguing standards produced during the Third Age, namely MARC, ISBD and AACR2, are presently being critically reviewed by the professional bodies and experts.

The Machine Readable Cataloguing (MARC) format came out after a long trial, with valid promises to carry on bibliographic record sharing at large scale. The International Federation of Library Associations and Institutions (IFLA) published a standard specification for bibliographic description (ISBD) in 1971. ISBD was originally constructed as a means of standardising the presentation of descriptive data so that it could be machine-translated into MARC.

In 1978, AACR2 reached a historic agreement between the North American and the British library associations and published the two sets of rules as a continuous single text document.

These three standards, which serve as the bedrock of cataloguing as of now, had a beginning in card catalogue environment, and embedded specifications well-suited for the physical catalogue of 5" × 3" cards.

13.1.4 Electronic Catalogue

Michael Gorman considered that the confluence of a need (national and research libraries throughout the world needing less expensive and more current cataloguing) and a means (automation and, more specifically, the MARC format) that has brought us nearer to Universal Bibliographic Control (UBC) than anyone would have dreamt possible thirty years ago.

In early 60s, the Library of Congress had developed a schematic design to help transfer its card catalogue records into newly installed computer. The design provided a way to pass on coded instructions to the computer, made of numbers, alphabets and special characters, along with bibliographic data. The computer requires all those for identifying cataloguing data elements, and reproducing LC catalogue cards, and tapes for distribution among the member institutions. This was how MARC began. The communication format of the MARC records, in the beginning, was suitable for encoding the bibliographic elements of books in roman script. During the span of four decades, the Library of Congress, with collaborators enriched MARC in terms of document-types and contents. They enhanced its structural design to accommodate requirements of all kinds of recorded knowledge– from prehistoric artifacts to Web publications.

For efficient management of bibliographic data in networking environment five separate formats are defined to handle five types of data: bibliographic, holdings, authority, classification, and community information.

From day one of mechanisation, the Library of Congress pursues the policy of sharing benefits of their machine-readable catalogue with libraries having no machine to read, by the way of distributing printed catalogue cards in conventional format. This happened to be the primary reason for structuring MARC in compliance with catalogue card production.

13.1.5 Online Catalogue

Definition of an online catalogue is simple. It is a catalogue in electronic (machine-readable) format accessible online. The term 'online' means 'connected to a computer network or accessible by computer'.

The recent technological boom has empowered libraries to exploit computer and communication facilities to their best advantage, and implement MARC standard for interchange of bibliographic information between databases. Besides institutional will, what the libraries need most is the knowledge and skill to develop MARC compliant databases addressing local as well as global requirements.

As we have already noticed, the formatting of bibliographic elements in a MARC record reflects a semblance of catalogue card, when computer processing systems call for a different approach. Nevertheless, "the fact is that there are tens of millions of MARC records in the world; … MARC is the basis for almost all automated bibliographic systems (including commercially produced systems); and, no practically feasible or demonstrably better system has been advocated."

Cataloguers appreciate these days that their working tools are too many yet inadequate to keeping up with what is happening around them. Today things are vastly more complicated:

- Cataloguing costs money and takes time. Sharing cataloguing records will save both, if everyone can agree on how to catalogue things the same way.
- Electronic resources (on computers) are hard to catalogue and manage, and not always easy to make available.
- Everything comes in many formats, and they are hard to catalogue, manage, and make available, too.
- There is more of everything.
- Technology is changing how libraries work, what they have in their collections, and what users need and expect.

13.1.6 Next-Generation Catalogue

The technology of Web 2.0 has opened up alternative ways to develop many different search models for bibliographic database. Quite a few successful catalogue systems already visible online, though still in experimental stage. They:

- give the patron a simple search interface that allows the user to enter vague, broad, and simple searches;
- allow the patron to drill down through the large result list, narrowing it down by whatever criteria they choose, until it is as precise as they want sort the results list so that the most relevant items are at the top of the list; and
- are tolerant of misspellings and unusual word choices in the patron's search.

The list shows what patrons desire to get, and what the next-generation catalogues can meet. Next-generation catalogues give patrons the same tools they already enjoy on websites. Since library databases are generally built in compliance with MARC, they look beyond integrated library system and will have little problem in working with a variety of systems.

Self Check Exercise

Note: 1) Write your answer in the space given below.

2) Check your answer with the answers given at the end of this Unit.

1) What is Next-generation catalogue? What are their special characteristics.

....

....

....

....

13.2 ONLINE CATALOGUE AND MARC DATABASE

13.2.1 Online Catalogue – Definition

Online catalogue is a library catalogue consisting of a collection of bibliographic records in machine-readable format. In other words, it is a machine-readable catalogue, by definition, and something more. It needs to be maintained in a dedicated computer that provides uninterrupted interactive access via terminals or workstations in direct, continuous communication with the central computer. In short, it is a catalogue in electronic (machine-readable) format accessible online. The term 'online' means 'connected to a computer network or accessible by computer'. For example, when you are connected to the Internet, you are online, when you search British Library Catalogue, you search an *online database*.

Online database needs two *common* record formats, a physical format, and a logical format. The common physical format serves as a vehicle to carry records of bibliographic data to remote computers. The common logical format, on the other hand, ensures that the recorded data are stored and interpreted correctly at remote ends.

13.2.2 What is MARC?

The key to the machine-readability is the common record format. MARC follows a physical structure and a set of control mechanism for enabling the machine to identify and process the data elements, whenever needed. This physical data structuring scheme, originally developed as a carrier of MARC data, is now being used as a standard, nationally (ANSI Z39.2) and internationally (ISO 2709), by other communication formats for bibliographic information interchange, like UNIMARC, CCF, national MARCs, etc. They all are implementations of ANSI Z39.2 / ISO 2709 standard.

The LC MARC found its way to US MARC, as an acknowledged national format for bibliographic communications, and after that evolved into MARC 21. MARC 21 is not a new format but a harmonised edition of USMARC, CAN/MARC, UKMARC and AUSMARC brought out as a consolidated scheme. Although, one finds there little change content wise, MARC 21 differs significantly in its vision. It tends to take issues beyond national preferences, and to negotiate with the up-coming events as well. Being its focus shifted from geographic to temporal zone, MARC 21 appears as a MARC version for the 21st century.

13.2.3 MARC Compliant Database

MARC database and MARC compliant database are not same. A MARC database contains a sequential file of hierarchically arranged records following ANSIZ39.2/ ISO2709 specifications and MARC defined content designation for every field in every record; whereas, a MARC compliant database may be a relational, or a different model, holding records comprising elements equivalent to MARC fields by definition. A database in compliance with MARC can transform its records into MARC communication format and exchange records with any MARC database or with another MARC compliant database; otherwise it cannot.

No online library can function without capability of interchanging records. As things stand now, it is almost impossible for a library to maintain a MARC-incompatible catalogue. Because, the population of records in MARC databases is by far greater than the rest of the databases as a whole. When working as stand-alone electronic catalogues, many libraries feel no obligation to follow a common record structure, overlooking any probability of sharing bibliographic data in future. For time being any non-standard computerised catalogue, system can serve the current and local needs satisfyingly. However, the good and committed libraries are expected to move toward standardisation sooner or later.

13.2.4 Resource Sharing Initiative

In manual environment, libraries had no other options but invest huge staff time in creating catalogue entries of their own for every document they acquire. When MARC files became available, individual libraries, with no computer facility, started buying computer-printed cards from the Library of Congress. In due course, more and more libraries installed computers. Those who designed their databases MARC compliant, preferred to download MARC from the Library of Congress or from other bibliographic utilities, like OCLC, WLN, RLIN, and A-G Canada. The agencies extended their on-line services based on their high-power mainframe systems, against fee-plus-communication-cost. Output was a magnetic tape containing bibliographic information tagged in modified MARC 21 format. The academic institutions, having sufficient fund and sophisticated library software, joined bibliographic utilities. They enjoyed benefit of sharing their collective resources by downloading and contributing MARC records. With the advancement of information technology, computer systems are getting more powerful, affordable, and friendlier day-by-day. Libraries have found automated systems are cost-effective as well as efficient, and resource sharing is the key to solve the problem of unleashed proliferations of information and ever-increasing information price. This happens to be the backdrop of coming of MARC into prominence. The scenario is different in less-advanced countries. Although computer power is within easy reach of the libraries, the concept of sharing MARC records is picking up.

13.2.5 Importance and Advantages

It is important for the implementing agencies to note the plus points of MARC, particularly MARC 21, and its advantages over other bibliographic exchange formats, like UNIMARC, CCF, and National MARCs:

- MARC is a collaborative effort. The MARC scheme, with its massive structure and exceedingly ambitious planning for upgrading its capabilities, depends on active participation of expert agencies. Primarily, the Library of Congress is responsible for creation and up keeping of the scheme. The National Agricultural Library, National Library of Medicine, United States Government Printing Office, and National Library of Canada, together with the Library of Congress serve as sources of authoritative cataloguing and provide most of the codes and standards used for MARC cataloguing.

- MARC is one of the very few international schemes that sustain the tempo of development and continuously update its resources. It stands current and serves as a reliable processing instrument for machine-readable cataloguing. The Library of Congress made no revisions in MARC specifications unilaterally. Two Committees were set up to review and update MARC 21 format documentation. One is the MARC Advisory Committee and the other, Machine- Readable Bibliographic Information (MARBI) Committee of the American Library Association (ALA). The MARC Advisory Committee is composed of the representatives from libraries and scholarly associations, as well as from the bibliographic utilities and vendor groups. The MARC authorities welcome proposals from all professional institutions and experts for incorporating changes in specifications in the interest of the user community at large.

- MARC 21 embraces all media of documents in its scheme. In 1970, the Library issued specifications for magnetic tapes containing *monographic* records in *MARC II format,* as they named it. The first document published with USMARC in title appeared in 1990. After a decade a revised edition of USMARC format for bibliographic data, has been produced in collaboration with the British Library and National Library of Canada as MARC 21. As its variant name suggests the revised format aims to meet the challenge of the 21st century.

- There is hardly anything in MARC that falls outside the scope of standardisation. Being a bibliographic information interchange format, MARC in principle does not dictate use of any particular standard, but allow its users to follow any one of the available standards or authorities, especially for data rendering, choice of access points, codes, classifications and choice of subject heading, name authorities, etc. MARC Program for Cooperative Cataloguing (PCC), as an international cooperative effort emphasises on greater flexibility in tailoring local cataloguing practice to local needs and priorities maintaining relative standards.

- MARC databases are multiplying. The Library of Congress, the official depository of United States publications, is a principal source of cataloguing records for US and international publications. The OCLC shared database, another enormous source of MARC records, available online for downloading against fees. These apart, many a public library and institutional library in USA, Canada, Britain and Australia offer MARC records, often at no cost. The population of MARC databases is growing fast. In India, under Government patronage, MARC 21 implementations are in progress in a number of libraries, including the National Library of India. The increase in accessibility of MARC records tends to boost up MARC implementations in turn.

13.2.6 Schematic Design

In early 60s the Library of Congress had developed a schematic design to help transfer its card catalogue records into newly installed computer. The design provided a way to pass on coded instructions to the computer, made of numbers, alphabets and special characters, along with bibliographic data. The computer requires all those for identifying cataloguing data elements, and reproducing LC catalogue cards, and Tapes for distribution among the member institutions. This was how MARC began. The communication format of the MARC records, in the beginning, was suitable for encoding the bibliographic elements of books in roman script. During the span of four decades, the Library of Congress, with collaborators enriched MARC in terms of document-types and contents. They enhanced its structural design to accommodate requirements of all kinds of recorded knowledge – from prehistoric artifacts to web-publications.

Union catalogues comprising MARC records can be generated and maintained efficiently by using MARC 21 holding format. MARC 21 bibliographic format offers full coverage of bibliographic elements related to every possible type of document. Its integrated design allows records of varied types and sizes to be kept together, and searched together.

13.2.7 Bibliographic and other Related Formats

For efficient management of bibliographic data in networking environment five separate formats are defined to handle five types of data: bibliographic, holdings, authority, classification, and community information.

Bibliographic Format

MARC 21 Format for Bibliographic Data is central to MARC system. It is an integrated format defined for the identification and description of different forms of bibliographic material, no more restricted to monographs as it was, but extends specifications in details for encoding elements of documents of any description, shape and material type — such as books, serials, computer files, maps, music, visual materials, web sites, databases and mixed materials too.

Holding Format

Format for Holdings Data serves as a subset of Bibliographic format. It contains that portion of format specifications which a library may require for encoding data elements pertinent to the holdings and physical location of the items the library holds.

Authority Format

This is a format of the Authority Record. Libraries maintain the files of records as cataloguing tools. Authority records help inputting consistently and uniformly personal and corporate names, subject headings, which serve as preferred access points in records. Format specifications for encoding these bibliographic elements standardise the textual representations of these bibliographic elements and ensure data consistency – a precondition for effective search.

Classification format

MARC 21 Format for Classification Data contains format specifications for encoding classification numbers and their descriptive terms. It serves as an in-house tool for maintaining consistency in library classification, and helps catalogue search by class numbers and browsing books on shelves.

Community Information format

Format for Community Information provides format specifications for records containing house-keeping information, events, programs, services, etc. required for running circulation system and other library management programs.

13.2.8 Implementing MARC

The people working with MARC standards, directly or indirectly, stand responsible for success and failure of MARC in playing its primary role, which is to develop sharable bibliographic resources.

The software developers building MARC utility programs, the database vendors distributing MARC records, and the professionals cataloguing MARC records – must know MARC adequately to do their part responsibly. In implementing MARC compliant library automation system, it is important that the people at top acquire clear-cut understanding about MARC to support their decision-making and help them see that their:

- Library Management Package is demonstratively MARC-compliant, and includes efficient MARC utility programs.
- MARC data files, whenever downloaded from external sources, hold no substandard records.
- MARC cataloguers work proficiently adhering to MARC specifications.

Self Check Exercise

Note: 1) Write your answer in the space given below.

2) Check your answer with the answers given at the end of this Unit.

2) What is MARC? State briefly how MARC21 was evolved.

....

....

....

....

13.3 MACHINE - READABLE CATALOGUING: STRUCTURAL DESIGN

13.3.1 Bibliographic Data Vehicle

Cataloguing data elements are transmitted as contents in an electronic file especially designed as per ISO 2709 specifications for bibliographic information interchange. The file stores a continuous stream of records consecutively placed in a single row, each separated by a special character.

record 1▲ record 2▲ record 3▲record 4▲record 5▲ … ▲record n↔

Fig. 13.1: Physical File

Computer singles out the records one by one by tracing the predetermined record-separators, which is a special character rarely used in text, often a non-printable sign. Each physical record as specified in ISO 2709 standard comprises three sections of data:

- Leader or Record Label
- Directory
- Data Content

Leader or Record Label

The Leader consists of a string of 24 characters, 00 to 23. The string includes mostly coded information all about the organisation and features of the record itself, only a few about bibliographic information.

Position: 01 ⟶ ⟸ 24

String: 01218cam 22003134a 45000

Fig. 13.2: Leader

The code may be a letter, number, or a Blank Space, as illustrated in Fig. 2. The meaning of any one-lettered code depends on its relative position in the string. For example, letter 'a' in position 06 indicates that the Record is for a language material.

Directory

The Directory serves as road map of Data Contents area. Directory information is dynamically gathered and stored in a place between the Leader and the Data Contents sections. The Directory is used for locating various fields of data elements each represented by a unique address comprising 12 numeric characters made of Field Tag, position, and field length.

001000900000005001700009008004100026906004500067925004200112955
0213001540100017003670200027003840200022004110400018004330420014
00451050002400465082001600489100001900505245003700524260004200561
30000380060349000250064150000210066650400590068750500710074665000
33008176500012008508000042 00862

Fig. 13.3: Directory

Data Content

MARC cataloguing involves

- logical structure of bibliographic record,
- content designation, and
- data content

Before taking up the logical construct of a MARC record, it should be worthwhile to find out what the components of a logical record are, and how those are organised into a machine-readable record serving as functional library catalogue. We discuss content designators first, because that will provide clues for analysis and interpretation of bibliographic records in communication format.

Content Designation

The data units are clustered in Data Content Section. While reading the continuous character string, computer picks up units of data with the help of predetermined control characters, which separate the physical units as fields and subfields. Hence, Fields and Subfields are actually logical names of physical units and sub-units of data.

By the same logic, the actual text, or value, of a data unit is defined as Field Content. The meaning of a data unit, or field content, is established only by putting it in context of Field Tag. The Field Tags, however, are kept in the Directory section, away from their respective data fields, but programmatically linkable. The 3-digit Field Tag, as indicated earlier, represents the type of bibliographic data contained in the field, e.g. Tag '245' signifies that the field contains *title statement* of the record item. This clarifies that the Field Tag serves a logical purpose of recognising the relationship of the field content with the record item. Like Field Tag, there are other tools to analyse and extract the field values.

Physical Control of Data Content

The two control devices, Subfield Delimiter and Field Terminator, work like traffic signals. The Subfield Delimiters (e.g. '$') embedded in field contents alert computer programs to take the immediate next character as a data identifier code *and not* a part of text. This way, the computer continues to read the text string as a subfield till it encounters another Subfield Delimiter or a Field Delimiter.

The control characters used for Subfield Delimiter and Field Delimiter are special characters, most often nonprintable. In MARC 21 records, an upright solid triangle and an upside-down solid triangle serve as the Field Separator and Subfield Delimiter respectively. When a MARC record is printed in Tagged Format the nonprintable characters are transformed into some legible signs by proxy.

```
LDR 01218cam 22003134a 450
001 13020293
005 20030923103827.0
008 021204s2003 nyua b b 001 0 eng
906 __ $a 7 $b cbc $c orignew $d 1 $e ecip $f 20 $g y-gencatlg
925 0_ $a acquire $b 1 shelf copy $x policy default
955 __ $a jb12 2002-12-04 $c jb12 2002-12-04 $d jb04 2002-12-06 $e jb02 2002-12-06 to Children's $a lb00 2002-12-10 $d lb04 2002-12-12 $a aa07 2002-12-16 $a ps07 2003-08-25 1 copy rec'd., to CIP ver. $f pv06 2003-08-28 CIP ver to CCD
010 __ $a 2002154957
020 __ $a 0516242946 (lib. bdg.)
020 __ $a 0516278819 (pbk.)
040 __ $a DLC $c DLC $d DLC
042 __ $a pcc $a lcac
```

050 00 $a QL737.C23 $b E34 2003
082 00 $a 599.756 $2 21
100 1_ $a Eckart, Edana.
245 10 $a Bengal tiger / $c by Edana Eckart.
260 __ $a New York : $b Children's Press, $c c2003.
300 __ $a 24 p. : $b col. ill. ; $c 16 x 19 cm.
490 1_ $a Animals of the world
500 __ $a "Welcome books."
504 __ $a Includes bibliographical references (p. 23) and index.
505 0_ $a Bengal tigers — Cubs — Roaring — New Words — To find out more.
650 _0 $a Tigers $v Juvenile literature.
650 _1 $a Tigers.
800 1_ $a Eckart, Edana. $t Animals of the world.

Fig. 13.4: Tagged MARC Record for Visual Representation

Functions of Content Designators

As Content Designators, the Field Tag, Indicators-1 and Indicator-2, and Subfield Code — all contribute to computer performance in reading the content of a bibliographic record meaningfully. The main objectives of the content designation are to support computer programming in -

- Searching and retrieving all identifiable bibliographic data elements.
- Formatting of retrieved data for visual presentation on screen and in print.

Field Tag

MARC reserves a number of Field Tags, or 3-digit codes, each represents a particular type of data. The 3-digit codes are, in fact, abbreviated form of the field names used for describing bibliographic units. Field Tag provides bibliographic format with flexibility. This is an indispensable feature for recording bibliographic elements since many data fields, e.g. Title, Author, etc. *vary in length*. Field Tag, being potentially a repeatable device, supports *multiple occurrences* of any field specified as 'Repeatable' in MARC 21 documentation. When a field, e.g. Tag 700, repeats twice it will produce two Alternate Authors fields. The nature of the data content of a field type determines repeatability (R) /non-repeatability (NR) of fields. For example, a bibliographic record supports only one non-repeatable (NR) main entry field 100 for Personal Author.

100 1_ $a Spilsbury, Richard, $d 1963-	NR
020 __ $a 0516242946 (lib. bdg.)	R
020 __ $a 0516278819 (pbk.)	
650 _1 $a Tigers.	R
650 _1 $a Endangered species.	

Fig. 13.5: Examples of Repeatable Field Tags

Indicators

Indicators are two: Indicator 1 and Indicator 2. Both the Indicators work together as Content Designators. Indicator 1 holds the first position at the beginning of a variable data field; the Indicator 2 holds the next position. Each of these provides some supplementary information about the field content, mostly related to visual presentation of the field. Each Indicator holds single-character code. The code may be a numeric, a lowercase alphabetic character, or a blank space. A blank space may mean:

Indicator undefined

- No value provided, or
- A specific meaning assigned.

Here are few examples:

- Value 3 in 1st indicator of field 246 indicates that field content is a parallel title; while value 3 in 2nd indicator indicates that there will be no notes but an added entry under the parallel title.
- A '0' value of 1st indicator in field 245 indicates that the title is the main entry; in field 246 it indicates that the field content is a portion of title.

TAG	INDICATOR	Field CONTENT
246	3	Parallel Title
246	3	No Notes. Added Entry
245	0	Main Entry under Title
245	3	Skip 3 characters in sorting/filing
650	0	LCSH used for Subject Heading

Fig. 13.6 : Changing Values of Indicators in Variant Fields

- A '0' value of 2nd indicator in field 650 indicates that for subject heading LCSH is followed; in field 245 the 2nd indicator value 3 indicates that the title begins with an article with three non-filing characters ('An' and a space). Indicator value 9 is reserved for local implementation.

Subfield Code

Subfield Codes identify data elements within a field for enabling the computer to manipulate each one separately. A Subfield Code is composed of a Subfield Delimiter and a Data Element Identifier. It is a necessary component of Subfield Code, but not a code by itself. A delimiter's function ends with passing a signal to computer predicting the presence of a Data Element Identifier. Data Element Identifier is a code, and has a crucial role to play in analysing bibliographic data content.

Data Element Identifier consists of either a lowercase alphabetic character or numeric character. The character '9' is kept reserved for local use as data element identifier. The order of subfields is specified by the bibliographic standards followed by the cataloguing agency, e.g., cataloguing rules, ISBD, giving priority to MARC 21 specifications wherever available.

245$a	Title statementTitle proper/short title	NR NR
$b	Remainder of title	NR
$f	Designation of vol./issue and/or date	NR
$g	Miscellaneous information	NR
$h	Medium	NR
$i	Display text	NR
$n	Number of part/section of a work	R
$p	Name of part/section of a work	R
$5	Institution to which field applies	NR
$6	Linkage	NR
$8	Field link and sequence number	R

Fig. 13.7: Examples of Subfield Codes with Dollar sign as Subfield Delimiter

Figure 7 illustrates use of numeric character in Subfield Code as processing parameters $6 and $8 besides repeatability (R) and non-repeatability (NR).

The design aspects of MARC, as we examined, cover the physical and logical control of bibliographic elements with the instrumentality of content designators and other control devices. The robust schematic design of MARC as a whole is responsible for the spectacular development of bibliographic network worldwide.

Self Check Exercise

Note: 1) Write your answer in the space given below.

2) Check your answer with the answers given at the end of this Unit.

3) What is Field Tag? What is the use of a Field Tag? Explain with illustration

....

....

....

....

13.4 METADATA TOOLS FOR CATALOGUING NETWORKED RESOURCES

13.4.1 Introduction

Since the time of Panizzi and Cutter, libraries have been evolving a set of tools for gaining bibliographic control of books, journals, and other printed resources. As the volume grows, and formats of resources diversify, principles of organisation undergo change all the time, in order to process their contents and make them accessible. In print-dominated library environment, the principles of organisation were applied manually, and the set of rules they created for 'technical' processing was known as cataloguing. This process is still being used in manual environment. It is effective, but labour-intensive and highly duplicative from library to library, and did not scale well as knowledge expanded. The mechanisation of cataloguing in the 1970s extended the practice of creating bibliographic records manually by providing a means for such records to be shared, thereby reducing costs and duplication. Even today, libraries achieve bibliographic control by means of sharing manual cataloguing of printed resources through electronic distribution.

For networked resources, however, cataloguing is found most inadequate for describing contents and providing access. To take care of the emerging needs of networked resources, a comprehensive and diverse set of controls is created and upgraded continuously. This set is appropriately called METADATA – the data about data.

13.4.2 Metadata

In 2004, NISO defined MEADATA as 'Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage information objects'. It is machine-understandable and support wide range of operations, and may be categorized as Descriptive metadata, Structural metadata, and Administrative metadata. It is believed that 'if librarians are involved at all, their role with respect to metadata will be vastly different for their old cataloguing role'.

Cataloguing Metadata Tools

The Metadata tools for cataloguing networked resources are too many and many of them are dependent on the nature and type of resources. Typology of Metadata can be mapped in the following way:

- Data Structure: Dublin Core, MODS, CDWA, VRA, LOM
- Data Content: AACR2, RDA, FRBR, CCO, CDPDCMBP
- Data Value: LCSH, AAT, TGN, LCTGM, ULAN, W3CDTF
- Data Format: XML, SGML, MARC
- Data Presentation: ISBD, CSS and/or XSLT

Here we briefly introduce the core sets of Metadata developed particularly for online cataloguing, namely AACR/RDA, and MARC, and two more related standards, namely, ISBD and FRBR.

AACR

Anglo-American Cataloguing Rules is a content standard for bibliographic description and access. It covers not just books, but all other print formats as well. Rules for each category of material specify what fields should be used and what data to include in each field. Text strings were originally intended for printed catalogue cards.

The key principles of AACR are:

- One principle entry per resource
- Catalogue from item in hand
- Chief source of information

AACR Timeline. It all begins with the publication of UK and US editions in 1967. The unified second edition, consistent with ISBD specifications, was published in 1978.

The major breakthrough was achieved in 1997 in Toronto conference on AACR2. Next year the Functional Requirements of Bibliographic Records (FRBR) was brought out. In 2005, Resource Description and Access (RDA) was developed.

AACR 2, in its first part deals with bibliographic descriptions: general rules, resource-specific rules, and rules regarding analytical entries. The second part deals with choice and constructions of access points.

The significance of AACR2 as a metadata tool for cataloguing online resources is fading out due to the following shortcomings:

- *Increasingly complex*
- *Lack of logical structure*
- *Mixing content and carrier data*
- *Hierarchical relationships missing*
- *Anglo-American centric viewpoint*
- *Written before FRBR*
- *Not enough support for collocation*
- *Unclear relationship with MARC Format*

RDA

RDA stands for "Resource Description and Access" the new standard that will replace AACR2.

The Aims of RDA are:

- *Rules should be easy to use and interpret*
- *Be applicable to an online, networked environment*
- *Provide effective bibliographic control for all types of media*
- *Encourage use beyond the library community*
- *Be compatible with other similar standards*
- *Have a logical structure based on internationally agreed principles*
- *Separate content and carrier data*

RDA Timeline

The RDA Prospectus was issued along with draft of some chapters in 2005. During next two years further drafts of chapters on description and access were issued. In 2008 IFLA conference a screenshot demo was presented.

Launch of online product has been scheduled in early 2009.

Use of RDA

The purpose of using RDA is primarily to analyse the knowledge resource being described and determine what is the type of content, what is its carrier form, what other resources it relates, which persons, families or corporate bodies is it related, to what concepts, events and places is it related. RDA rules are for applying to all content types and all media types.

RDA aims to be:

Independent of communication formats for bibliographic and non-bibliographic resources:

- UNIMARC, MARC, MARCXML, MODS/MADS
- DC, EAD, ISBD, VRA, MPEG7

Compatible / better aligned with other similar standards

- Archives: ISAD (G)
- Museums: Cataloguing Cultural Objects

MARC

MARC is a communication and exchange format providing a structure for encoding the content of bibliographic and authority data. The MARC format was developed in the late 1960s as a tagging scheme for exchanging cataloguerecords on magnetic tape. It remains the standard way to represent such data. At present, MARC is steadily being converted (slowly) to modern computing formats, e.g., Unicode, XML.

The followings are the primary reasons for the growing dissatisfaction about MARC format:

- A classic legacy system
- Not designed for computer algorithms
- One record per item (poor links between records)
- Tied to traditional materials and traditional practices
- Not Unicode
- 100 of million records at $100 — $10 billion
- Note that the content is designed to be part of a printed cataloguerecord and is not in a convenient format for computer manipulation.

MARC Timeline

In 1960s, Library of Congress designed MAchine-Readable Cataloguing format for developing database of cataloguerecords for producing printed cards. The British Library developed UKMARC in parallel. In 1970s, variant national formats like AUSMARC, DANMARC, were developed. USMARC brought out 8 material formats, Books, Serial, Maps, etc. In 1977, UNIMARC was developed by IFLA to exchange records between national MARCs. Some important changes have taken place in recent time. Those are:

Expansion of USMARC to a family of formats

Bibliographic, Holdings, Authority, Classification, Community Information

Integration of USMARC bibliographic format and previous 8 formats

Widespread adoption of MARC 21

1997 – USMARC and CANMARC become MARC 21

2003/4 – MARC 21 enhanced by UK proposals; British Library adopts MARC 21

2006/7 – MARC 21 enhanced by German proposals: this will enable libraries to move from MAB to MARC21

Modernizing MARC

- Keep the content of the catalogue record
- Convert to Unicode for representing scripts
- Convert to XML for tagging cataloguing metadata.

ISBD

The International Standard Bibliographic Description (ISBD) is a set of rules produced by IFLA to describe a wide range of library materials within the context of a catalogue. One of the original purposes of the ISBD was to provide a standard form of bibliographic description that could be used to exchange records internationally. This would support IFLA's program of universal bibliographic control.

ISBD Timeline

International Standard Bibliographic Descriptions (ISBD) was developed by IFLA 1969 onwards.

The consolidated edition of the ISBD was published in 2007. It superseded earlier separate ISBDs that were published for monographs, older monographic publications, artographic materials, serials and other continuing resources, electronic resources, non-book materials, and printed music. IFLA's ISBD Review Group is responsible for maintaining the ISBD.

ISBD defined seven areas of description, their sequential order, and associated punctuations.

- Title
- Statement of Responsibility
- Edition
- Resource specific information
- Publication details
- Physical description
- Series information
- Notes and standard identifiers

FRBR

FRBR is a product of the logical progression of thought, initiated one and half a century ago, on how to organise the catalogue in this challenging information age.

FRBR Timeline

Functional Requirements of Bibliographic Records (FRBR) was prepared by IFLA in 1998.

From 1992 to 1995, the IFLA Study Group on Functional Requirements for Bibliographic Records developed an entity relationship model as a generalised view of the bibliographic universe, intended to be independent of any cataloguing code or implementation.

FRBR entity-relationship model defines:

Tasks: find, identify, select, obtain

Resourcerelationships:

work, expression, manifestation, item

- Entities: people, corporate bodies (agents)
- Entities: concepts, objects, events, places

FRBR Aims

The idea behind the FRBR conceptual model is that the catalogueis not seen as a sequence of records, but rather as a network of connected data. FRBR clarifies how catalogues should function, and illuminates what information is of the most value to users of the catalogue.

Self Check Exercise

Note: 1) Write your answer in the space given below.

2) Check your answer with the answers given at the end of this Unit.

4) What is Metadata? What are the metadata tools for cataloguing?

....

....

....

....

13.5 OPAC – ONLINE CATALOGUE INTERFACE

13.5.1 What is OPAC

The acronym, OPAC, stands for *Online Public Access Catalogue*. For all practical purpose, we may define OPAC as computer terminal at user end providing a friendly interface for searching, retrieving and viewing the machine-readable cataloguing data in eye-readable text format. Web 2.0 technology helped decoupling the user interface from the back-end systems that support all internal operations of a library including management of databases and services.

13.5.2 Backdrop

Long back in early 1980s, a librarian predicted that "Survival of Library "as an information agency will be dependent on its ability to redefine its procedures and goals in terms of the bibliographic universe as a whole. In doing so, it will be necessary to place its basic tool, the catalogue, in its proper perspective with other access tools."

Today, the scenario remained as it was, if not worsened. After 25 years, OCLC surveyed 396 college students from six countries, on their use of library resources. The survey revealed that 89 percent of them began their information search on a search engine, and only 1 percent on a library catalogue.

It is, however, realised now that the "persistent problems of the catalogueexist less with its business modules and more with its front-end" as R. David Lankes stated at the ALCTS 50th anniversary conference in June 2007.

13.5.3 We Need a New Interface

In an environment of Web 2.0 technology, the online library catalogues with new interfaces call for immediate attention. An unprecedented amount of research effort is now being put by the library automation vendors, and open-source champions in creating new library interfaces to match with users' expectations.

The Net Gen library users of today are Web-savvy, forward-looking and little intolerant. To attract them toward the library web sites and retain their attention, we must recreate our web interfaces competing the commercial Web. If the interface doesn't respond within seconds, a large percentage of users will click away to other sources.

Library catalogues are lagging behind the commercial and social interfaces we find on the Web, both in terms of looks and functionality. The web users prefer to follow a search model comprising faceted navigation and result clustering. It starts with entering a general term, receiving abundant results, and drilling down to incrementally narrow those results. This contrasts against the structured method still prevailing in library environment, where the search process starts with a pre-formulated Boolean-logic, contrary to direct method of keyword searching. It has been observed in 1997 that the library users, following the general trend in web environment, search online catalogue more often by keyword than not.

The gap between the online library catalogue search experience and search experience in commercial and open-source arena will continue to grow further until we reconstruct OPAC - the user-interface of our online catalogue.

13.5.4 OPAC in Next-Generation Web Catalogue

There are quite a few next-generation library interfaces now available on Internet. Those are still in their early stages of development but already have demonstrated the revolutionised ways libraries can interact with users. These interfaces are easy to use, look impressive, and take sophisticated approach to find information, and put the details on view helpfully. Some of the most innovative projects, like Endeca, Encore, eXtensible Catalogue, Aqua Browser Library, Koha, are now working in a number of forward-looking institutions.

13.5.5 Search Tools

Basic expectations from a web-based interface that it should provide the user with search results in order of relevancy ranking, facets (automatically generated terms that can be clicked to narrow results), corrections of misspellings, and suggestions to alternative search terms. To meet these expectations, the new generation library interfaces offer: relevance ranking, faceted navigation, search result clustering, breadcrumb trails, and a faster, more friendly search environment. Each has its strengths. By critically examining the key concepts, features, and techniques associated with these tools, librarians should be able to identify the essential elements of successful search interface that libraries can adapt to their systems.

13.5.6 Relevance Ranking

Whoever searches Internet often becomes used to the idea of relevance ranking where the best and most interesting items rise up to the top of their hit list. This is because all the giant search engines on web, like Google and Yahoo!, work this way.

Relevance ranking works satisfactorily when comes through careful refinement of the formulas that determine the ordering of results. One must develop skill for achieving good relevance ranking.

With relevance ranking, users may expect to get a large result set. That does not make them upset. They know, by looking up the top of the first page the most relevant information should be found.

In fact, most users do not notice how big the result set is.

In spite of the great advantage of relevance ranking, users may often require to sort the search hits by title, author, date, or by some other elements. Provision of sorting by elements of users' choice is another attractive feature of new OPAC.

13.5.7 Faceted Navigation

A well-constructed faceted navigation scheme will allow users to quickly drill down from a broad set of results to a manageable group of results. The search interface selects the facets from metadata that's associated with the items in the body of information. Names, subject headings, publication dates, and other fields from a structured record provide good opportunities for generating facets.

There has been considerable interest in Faceted Application of Subject Terminology, an alternative way of applying the Library of Congress Subject Headings. This system has been designed to be more amenable to interfaces that employ faceted navigation.

Collections that already have rich metadata, such as library catalogues, lend themselves to faceted navigation. Those with sparse metadata or none at all call for result clustering.

13.5.8 Clustering

Clustering of results, by subject, author, genre format or date, can allow users to easily refine a search, with one click of the mouse. New methods of browsing—via peer recommendations, or through subject taxonomies, or related websites—have emerged.

Result clustering, like faceted navigation, aims to give the user a fast and easy way to narrow results. While faceted navigation usually relies on metadata terms, clustering operates by analysing the raw text of the items in the result set to create labels that represent each group or cluster. From the user's perspective, faceted navigation and clustering look much the same. With both, you click through a succession to progressively narrow the results down to a manageable number.

Using new search technologies, we can meet this expectation. Since it was first popularised by Google, relevance ranking has transformed the way that people search.

13.5.9 Breadcrumbs

Breadcrumb links, also known as contextual links, are a type of navigation aid for Web pages. They provide a textual representation of a site's structure, usually a vertical hierarchy of a site. For example, on e-commerce websites, breadcrumb links often have the following format:

Home > Category > Subcategory > Product

Breadcrumb links can help user take some steps back toward home when needed. A new link appears with each step through the site and persists until the user safely exits. Breadcrumbs are especially helpful in search environments that use faceted navigation. New breadcrumbs show each facet that's been selected to narrow the search.

Since this drill-down approach is often an iterative process, a good presentation of breadcrumbs allows the user to easily back out of an existing facet selection and choose another.

13.5.10 Search Result Data

Key challenges for these new interfaces include the need to extract data from the library service environment and provide up-to-the-minute status information on each item in the library's collection, for example 'on circulation'.

Librarians need to construct search interfaces that include the catalogue of hybrid collections that may include a large portions of the electronic content to which they subscribe. The nature of information requirements and information seeking behaviour both have changed in 21st century, as the figure-1 shows. Some of the new generation OPAC are paving the way toward providing more immediate and seamless access to diverse content collections, giving equal footing to both the digital and print collections. As we are experiencing currently, the medium and form of document is irrelevant to information user. What they need is the information. They do not care much if the source document format is a print or electronic, local or remote, or whether it is a text, video, or audio. Our regenerated OPAC must take care of their need and their information seeking behaviour.

Self Check Exercise

Note: 1) Write your answer in the space given below.

2) Check your answer with the answers given at the end of this Unit.

5) What is Relevance Ranking? How does it help researchers? What else they may need to arrange their search hits helpfully?

....

....

....

....

13.6 ONLINE CATALOGUING UTILITY SERVICES

Libraries invest a substantial proportion of funds to generate and update machine-readable catalogue– a labour-intensive time-consuming process. The job demands, besides cataloguing knowledge and skill, a command on metadata for networked resources.

Online catalogues are generated in two ways: by *original cataloguing* or by *copy cataloguing*. In networking environment, copy cataloguing work includes (1) finding sources of MARC records of target documents, (2) downloading records, (3) eliminating unwanted fields, and (4) adding fields of local relevance. Since cataloguing procedures are carried out programmatically and often supported by utility services, it takes less time, involves less cost and less intellectual effort. On the other hand, original*cataloguing* stands for complete cataloguing of a document from scratch. It is undertaken when downloadable records of corresponding documents are not found.

The objectives of the online cataloguing utilities are to provide ways and means to reduce costs enhance productivity and improve cataloguing workflow.

There are websites offering free catalogue search and download options for records retrieved. LC database is one of them. This gesture helps researchers. However, for library cataloguing, the facility proves disobliging. The utility networks operate at regional and international levels, and extend variety of bibliographic support services and products. Distribution of cataloguing data in MARC format against subscriptions is one common agenda. Top organisations, like Library of Congress and OCLC are among them. Besides the Library of Congress and OCLC, there are hundreds of commercial and no-profit agencies who offer variety of tools and support services for original cataloguing, copy cataloguing, data editing, record validation, MARC conversion, and even for spell checking, link checking, display tools. A selected few are mentioned here group wise with their service profiles.

13.6.1 Search View Print

Book Where

WebClarity Software Inc.'s BookWhere product allows users to simultaneously search and retrieve metadata from over 1900 pre-defined databases located around the globe. Bibliographic and media records can be found in seconds and exported in a wide variety of formats including MARC 21, FINMARC and UNIMARC. www.webclarity.info

FRBR Display Tool - Free

The FRBR Display Tool sorts the bibliographic data found in a set of MARC records into hierarchical displays by grouping the bibliographic data using the "Works," "Expressions," and "Manifestations" FRBR concepts. Possible uses for the FRBR Display Tool include experimenting with the collocation and sorting of search result sets into the FRBR categories to test concepts; and applying FRBR to local data to evaluate its consistency for FRBR-type development. www.loc.gov/marc/frbr/tool.html

MARCView ™

MARCView™ is an easy-to-use program to view, search, and print any MARC 21, USMARC, CanMARC, UNIMARC, or MARCXML bibliographic or authority file. Records are formatted for easy viewing and printing. Navigation to any record in the file is instantaneous. Searches can specify field, subfield, both, or neither. www.systemsplanning.com/marc

MarciveWeb SELECT

Enables librarians to search 10 million records from LC, NLM, NLC, GPO, A/V Access(R), and other sources, and obtain customized MARC 21 bibliographic records, cataloguecards, smart barcode labels, book labels, and MARC 21 authority records. www.marcive.com

13.6.2 Copy Cataloguing

Impact/ONLINE CAT

Impact/ONLINE CAT is a Windows-based cataloguing system which provides users with the ability to search multiple databases of MARC 21 bibliographic, authority and community information records. Features include a MARC Editor with built-in MARC validation, and the ability to download records to the local hard drive. www.auto-graphics.com

Connexion

This is a robust suite of full-service online cataloguing tools and services backed by OCLC's 35+ years of cataloguing experience. Its enhanced features provide unparalleled flexibility for libraries and other allied institutions. Connexion lets you create and edit high-quality bibliographic and authority records, then share them with the entire OCLC cooperative, which benefits libraries around the world.

OCLC-MARC Record Delivery

A variety of WorldCat services generate bibliographic records in OCLC-MARC format. OCLC makes copies of these records available for you to import to your local system to keep your local holdings in synch with your holdings in WorldCat. Downloading records on a regular basis makes it easier for your systems staff to manage and upload them into your OPAC.

Surpass Copycat

Surpass Copycat is a Windows-based Z39.50 copy cataloguing tool that allows users to find and download free MARC records from the Internet. Search multiple libraries simultaneously, such as the Library of Congress, public libraries, medical libraries, state-wide union catalogues and more. Over 100 libraries come pre-configured. Copycat also features "scan and search" that allows the user to simply scan the EAN/ISBN barcode from the back of the book they wish to catalogueto instantly launch a search for that book. www.surpasssoftware.com/copycat.htm

WorldCat Cataloguing Partners

OCLC record delivery in coordination with your vendor orders

Through a collaborative effort with materials vendors, corresponding OCLC MARC records are delivered with the materials you order through participating vendor partners. Additionally, your library's holdings are set automatically in WorldCat.

13.6.3 Converter

MARC RTP - Free

MARC RTP will read files of bibliographic records in MARC format, and convert them to a format that the user designs. The program can also produce a human readable listing or summarize the structure of a file of records. www.loungebythelake.com/marcrtp/

MARConvert™

MARConvert™ handles special problems or unusual requirements in converting records into or out of MARC 21, USMARC, CanMARC, UNIMARC, or MARCXML. It will also convert MARC records to another character set, such as ANSEL, Latin-1, Unicode, or UTF-8. Operates in both interactive mode, and batch mode for converting multiple files. For Windows 95/98/NT/2000/XP. www.systemsplanning.com/marc/mvd.asp

MARCMaker - Free

MARCMaker is developed by the Library of Congress (LC) that generates the MARC record structure from preformatted text. It runs under DOS or Windows 95/98/ME/2000. (www.loc.gov/marc/makrbrkr.h tml#download)

(www.loc.gov/marc/makrbrkr.html

USEMARCON Plus - The Universal MARC Record Converter - Free

Like the original UseMARCON program, the USEMARCON Plus program enables libraries to create rules-based systems to convert records between national MARC formats. It also allows users to create and modify rules files, used to achieve MARC conversions, in order to meet specific local requirements. www.bl.uk/services/bibliographic/usemarcon.html

13.6.4 Validator

MARC Report

MARC Report validates MARC records according to the latest LC and OCLC standards. The validation that is applied is customizable by the user. The program runs either in interactive mode (record-by-record) or in batch mode (producing a report of all problems found). Unique to MARC Report are hundreds of cataloguing cross-checks which check the internal logic of each record, making sure that data elements present in one field do not conflict with those present in another. www.marcofquality.com

Validator ™ Subjects and Names Authority Database

Validator ™ Subjects and Names Authority database includes the complete Library of Congress Subjects and Names Authority Files on CD-ROM. Included are 248,000 subject headings and 4.7 million names records specifying personal and corporate names, series and uniform titles. Validator is versatile and useful to both cataloguers and reference staff. www.att.com/gov/library/

13.6.4 Validator

CILLA

Co-operative of Indic Language Library Authorities

The CILLA service provides quarterly book lists for Bengali, Gujarati, Hindi, Panjabi, Tamil and Urdu materials suitable for public library audiences. This co-operative service enables library authorities to purchase materials more easily, and in confidence, for languages where they do not have a local specialist. The supply of MARC records also enables library authorities to gain efficiencies in cataloguing.

Cataloguing Calculator - Free

The Cataloguing Calculator finds variable and fixed MARC fields (bibliographic and authority data), language codes, geographic area codes, publication country codes, AACR2 abbreviations, LC main entry and geographic Cutter numbers. http://home.earthlink.net/~banerjek/calculate/

InfoWorks Link Checker

InfoWorks Link Checker is designed for librarians to quickly check URL links in MARC 21 records and works with flat MARC files containing URLs. InfoWorks Link Checker can also check multiple links simultaneously by applying multiple threading technology. www.itcompany.com/linkcheck.htm

InfoWorks Spelling Checker for Database Maintenance

Locates spelling errors and cleans up bibliographic databases. It works with flat MARC files in any integrated library systems. InfoWorks Spelling Checker for Database Maintenance allows users to define which languages and fields to check and provides both batch and interactive checking. It includes a special dictionary for library use and allows users to build custom dictionaries.

www.itcompany.com/checkerd.htm

Self Check Exercise

Note: 1) Write your answer in the space given below.

2) Check your answer with the answers given at the end of this Unit.

6) How machine-readable catalogues are generated in networking environment? How do the utility services help catalogue generation? Name one 'Copy cataloguing' utility service.

....

....

....

13.7 SUMMARY

The theme of this Unit is 'online cataloguing: design and service'. We discussed six different topics touching upon different aspects of online cataloguing. Before going into the details of online cataloguing, we reviewed the development of library catalogue from its early stage up to its latest manifestation in networking environment. This helps students to gain an insight into the objectives of a catalogue, and importance of cataloguing standards. It was then; we take up the functional aspects of online catalogue, as distinguished from electronic catalogue used for library automation, in context of MARC database. Subsequently, the internal structure of a MARC record was examined taking the requirements of bibliographic fields in view. Next, we observed the functionality of few metadata tools in building online catalogues, including AACR2/RDA. Then, OPAC as user interface was discussed in the context of advancing web 2.0 technologies. Last, we observed the requirements of 'copy cataloguing' and 'original cataloguing' and the features of various utilities, offered by network services for generation, conversion, validation of MARC records, and other supports. This gives an exposure to the different tasks associated with cataloguing of networked resources, apart from awareness of the utility services as such.

13.8 ANSWERS TO SELF CHECK EXERCISES

1) The technology of Web 2.0 has opened up alternative ways to develop many different search models for bibliographic database. Quite a few successful catalogue systems already visible online, though still in experimental stage. These products of Web 2.0 technology are commonly referred to as next-generation catalogue. They:

- give the user a simple search interface that allows the user to enter vague, broad, and simple searches.

- allow the user to drill down through the large result list, narrowing it down by whatever criteria they choose, until it is as precise as they want.

- sort the results list so that the most relevant items are at the top of the list
- tolerate misspellings and unusual word choices in the user's search.

The list shows what users desire to get, and what the next-generation catalogues can meet. Next-generation catalogues give users the same tools they already enjoy on websites. Since library databases are generally built in compliance with MARC, they look beyond ILS and will have little problem in working with a variety of systems.

2) MARC is an acronym stands for MAchine Readable Catalogue. MARC standards consist of the MARC formats, which are standards for the representation and communication of bibliographic and related information in machine-readable form. The key to the machine-readability is this common record format. MARC follows a physical structure and a set of control mechanism for enabling the machine to identify and process the data elements, whenever needed. This physical data structuring scheme, originally developed as a carrier of MARC data, is now being used as a standard, nationally (ANSI Z39.2) and internationally (ISO 2709), by other communication formats for bibliographic information interchange, like UNIMARC, CCF, national MARCs, etc. They all are implementations of ANSI Z39.2 / ISO 2709 standard.

The LC MARC found its way to US MARC, as an acknowledged national format for bibliographic communications, and after that evolved into MARC 21. MARC 21 is not a new format but a harmonized edition of USMARC, CAN/MARC, UKMARC and AUSMARC brought out as a consolidated scheme. Although, one finds there little change content wise, MARC 21 differs significantly in its vision. It tends to take issues beyond national preferences, and to negotiate with the up-coming events as well. Being its focus shifted from geographic to temporal zone, MARC 21 appears as a MARC version for the 21st century.

3) **Field Tag**

MARC reserves a number of 3-digit codes, or Field Tags, each represents a particular type of data. The 3-digit codes are, in fact, abbreviated form of the field names used for describing bibliographic units. Field Tag provides bibliographic format with flexibility. This is an indispensable feature for recording bibliographic elements since many data fields, e.g. Title, Author, etc. *vary in length*. Field Tag, being potentially a repeatable device, supports *multiple occurrences* of any field specified as 'Repeatable' in MARC 21 documentation. When a field, e.g. Tag 700, repeats twice it will produce two Alternate Authors fields. The nature of the data content of a field type determines repeatability (R) /non-repeatability (NR) of fields. For example, a bibliographic record supports only one non-repeatable (NR) main entry field 100 for Personal Author.

100 1_ $a Spilsbury, Richard, $d 1963-	NR	
020 __ $a 0516242946 (lib. bdg.)	R	
020 __ $a 0516278819 (pbk.)		
650 _1 $a Tigers.	R	
650 _1 $a Endangered species.		

Examples of Repeatable Field Tags

4) **Metadata**

In 2004, NISO defined MEADATA as 'Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage" information objects'. It is machine-understandable and support wide range of operations, and may be categorised as Descriptive metadata, Structural metadata, and Administrative metadata. It is believed that 'if librarians are involved at all, their role with respect to metadata will be vastly different from their old cataloguing role'.

Cataloguing Metadata Tools

The Metadata tools for cataloguing networked resources are too many and many of them are dependent on the nature and type of resources. Typology of Metadata can be mapped in the following way:

- Data Structure: Dublin Core, MODS, CDWA, VRA, LOM
- Data Content: AACR2, RDA, FRBR, CCO, CDPDCMBP
- Data Value: LCSH, AAT, TGN, LCTGM, ULAN, W3CDTF
- Data Format: XML, SGML, MARC
- Data Presentation: ISBD, CSS and/or XSLT

Here we briefly introduce the core sets of Metadata developed particularly for online cataloguing, namely AACR/RDA, and MARC, and two more related standards, namely, ISBD and FRBR.

5) **Relevance Ranking**

Whoever searches Internet often becomes used to the idea of relevance ranking where the best and most interesting items rise up to the top of their hit list. This is because all the giant search engines on web, like Google and Yahoo!, work this way.

Relevance ranking works satisfactorily when comes through careful refinement of the formulas that determine the ordering of results. One must develop skill for achieving good relevance ranking. With relevance ranking, users may expect to get a large result set. That does not make them upset. They know, by looking up the top of the first page the most relevant information should be found. In fact, most users do not notice how big the result set is.

In spite of the great advantage of relevance ranking, users may often require to sort the search hits by title, author, date, or by some other elements. Provision of sorting by elements of users' choice is another attractive feature of new OPAC.

6) Online catalogues are generated in two ways: by *original cataloguing* or by *copy cataloguing*. In networking environment, copy cataloguing work includes (1) finding sources of MARC records of target documents, (2) downloading records, (3) eliminating unwanted fields, and (4) adding fields of local relevance. Since cataloguing procedures are carried out programmatically and often supported by utility services, it takes less time, involves less cost and less intellectual effort. On the other hand, original *cataloguing* stands for complete cataloguing of a document from scratch. It is undertaken when downloadable records of corresponding documents are not found.

The objectives of the online cataloguing utilities are to provide ways and means to reduce costs enhance productivity and improve cataloguing workflow.

Connexion Utility Service

This is a robust suite of full-service online cataloguing tools and services backed by OCLC's 35+ years of cataloguing experience. Its enhanced features provide unparalleled flexibility for libraries and other allied institutions. Connexion lets you create and edit high-quality bibliographic and authority records, then share them with the entire OCLC cooperative, which benefits libraries around the world.

13.9 KEYWORDS

AACR2	:	Anglo-American Cataloguing Rules [AACR2 - 2nd Edition (1978), AACR2R - Revised 2nd edition (1988)]
ALA	:	American Library Association
BL	:	British Library
DUBLIN CORE	:	Dublin Core is a 15-element metadata element set intended to facilitate discovery of electronic resources
FRBR	:	Functional Requirements for Bibliographic Records—or FRBR is a conceptual entity-relationship model developed by IFLA that relates user tasks of retrieval and access in online library catalogues and bibliographic databases from a user's perspective.
IFLA	:	International Federation of Library Associations and Institutions
ISBD	:	International Standard Bibliographic Description
LA	:	Library Association
LIBRARY 2.0	:	Application of interactive, collaborative, and multi-media web-based technologies to web-based library services and collections
OCLC	:	Online Computer Library Centre Inc.
OPAC	:	Online Public Access Catalogue
MARC	:	An acronym stands for Machine Readable Catalogue. MARC standards consist of the MARC formats, which are standards for the representation and communication of bibliographic and related information in machine-readable form
MARC21	:	The "harmonization" of USMARC and CAN/MARC; maintained by the Network Development and MARC Standards Office of the Library of Congress.
METADATA	:	Metadata is information about an informational data about data resource, be that a document (such as a webpage), image, dataset or other resource.

RDA	:	RDA stands for "Resource Description and Access" the new standard that will replace AACR2
UNICODE	:	Unicode is a computing industry standard allowing computers to consistently represent and manipulate text expressed in most of the world's writing systems.
UNIMARC	:	Universal MARC Format
WEB 2.0	:	Refers to a perceived second generation of web development and design, that facilitates communication, secure information sharing, interoperability, and collaboration on the World Wide Web.

13.10 REFERENCES AND FURTHER READING

Kehal, Harbhajan S and Varinder P. Singh. *Digital Economy: Impacts, Influences, and Challenges.* Hershey, PA : Idea Group Publishing, 2005. Print.

Drake, Miriam A. "Language, Arts and Disciplines".*Encyclopedia of Library and Information Science*: Pub-Zoo. New York: Marcel Dekker, 2003. Print.

Kesselman, Martin Alan and Irwin Weintraub. *Global Librarianship.* New York: Marcel Dekker, 2004 . Print.

Mukhopadhyay, Asok. *Guide to MARC 21for Cataloguing of Books and Serials, with Functional Definitions, Examples, Working Resources.* New Delhi: Viva, 2007. Print.

Verheul, Ingeborg. *Networking for Digital Preservation: Current Practice in 15 National Libraries* . Munchen : IFLA/Saur: 2006. Print.

Joachim, Martin D. *Languages of the World: Cataloguing Issues and Problems.* Chicago: University of Chicago Press, 1994. Print.

Casey, Michael E. and Laura C. Medford. *Library 2.0: A Guide to Participatory Library Service.* N.J.: Information Today, 2007. Print.

Grosch, Audrey N. *Library Information Technology and Networks.* New York: Marcell Dekker, 1995. Print.

Fritz, Deborah A. and Richard J. Fritz. *MARC 21 for Everyone: A Practical Guide.* Chicago: American Library Association, 2003. Print.

UNIT14 OVERVIEW OF WEB INDEXING, METADATA, INTEROPERABILITY AND ONTOLOGIES

Structure

14.0 OBJECTIVES

After reading this Unit, you will be able to:

- define the meaning and need of web indexing;
- explain the role, usage and importance of metadata;
- define is ontology and its importance in web parlance;
- explain interoperability and various methods of interoperability; and
- discuss protocols for interoperability.

14.1 INTRODUCTION

Index is a tool that has been in use for a long time to locate information. It is a list of key words or terms that supplement a document at the end of text for fruitful navigation and browsing. An index not only provides a chance to highlight content and provide a bird's-eye-view to the document, it also helps to identify the inconsistencies and improve upon content of the document for the author. The Web has emerged as an enormous source of information with a lot of chaotic information content also. Structurally, it is a collection of websites hosted at different domains round the globe. Websites can be defined as sources of information, consisting of individual webpages. Index to this content is available at individual level (through websites) and at global level (through search engines).

14.2 WEB INDEXING

Web index is a tool used for searching web documents like, individual websites or collections of web sites or collections of webpages and so on. It is a browsable list of terms or sections leading towards further reading/resources to the desired topic or subject. Sitemap is an example of a web index.

Indexing is an intellectual activity where the indexer determines what concepts are worth indexing. The entries and arrangement of these entries are equally important. There is a view that web indexing as well as traditional indexing is best done by individuals, skilled in the art of indexing. It requires imagination and formal knowledge of the subject.

However, there are automated ways of doing web indexing. Search engines use a program called spider or crawler to extract the search terms from the individual webpages. These spiders or crawlers collect the terms and store the terms inside a local database of a search engine and use it as a search index. These terms are either extracted from the 'meta tag' or from the contents of the webpage.

A web index is often a browsable list of entries from which the user makes selections. The index may not be displayed to the user but the user may retrieve information by just typing her/his query into a search box. A website A-Z index is a kind of web index that resembles an alphabetical back-of-the-book style index, where the index entries are hyperlinked directly to the appropriate web page or page section, rather than using page numbers.

According to British Indexing Standard (BS3700:1988) "Web index is a systematic arrangement of entries designed to enable users to locate information in a document."

Web indexing is the process of creating index manually or mechanically for the content of web documents. It includes back-of-book-style indexes for individual websites or an Intranet. It may further include the creation of metadata based keywords to provide a more useful vocabulary for the Internet or onsite search engines. With the increase in the number of e-journals, web indexing has also become important for publishing houses.

14.2.1 Concept

Index is a tool to help users locate information quickly and easily. Often it is understood as list of terms or phrases. But it is something beyond that. It brings like concepts together by grouping and creates a concept map in the mind of user about the document. Similarly, web index is a tool to locate easily and quickly the information on a website. A site map of a website is an index. Normally, indexes are used for web browsing. The terms in the sitemap of a website are directly hyperlinked to the topical web page or to the topic itself within the webpage. It performs the following three important tasks:

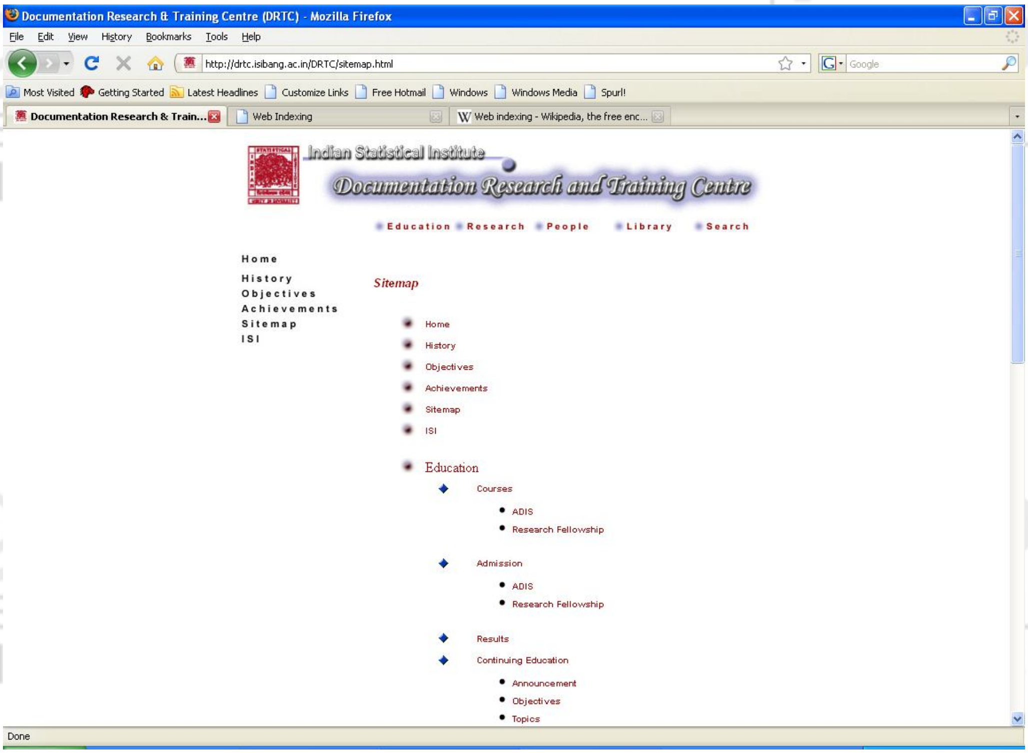


Fig. 14.1: Example of Sitemap Index of DRTC website

Source: http://drtc.isibang.ac.in/DRTC/

1) It describes the relationships among subjects. These include hierarchical and other relationships that exist amongst the subjects.

2) Index once stored in the database can be used as a source of meaningful metadata for searching. The search engines use index terms stored in such a type of database. An index provides visibility to all the available literature. Sometimes a user makes a search but cannot express the exact search phrase. In such a case, a good index brings material related to what s/he is looking for, with the help of related terms or concepts.

3) It is not only useful from the searchers' point of view but also for the point of view of the author. An index focuses on the content of the document and demonstrates the inconsistencies about the treatment of the topics. Hence, it is an aid for authors to review the writing for completeness of content.

Web index speeds up the browsing by presenting a browsable conceptual map about the content of the document. It also facilitates searching if used with proper search algorithm and search tools.

Self Check Exercises

Note: 1) Write your answer in the space given below.

2) Check your answer with the answers given at the end of this Unit.

1) Discuss the need of index in web parlance.

....

....

....

....

14.2.2 Types of Web Indexes

Web indexes are of following types:

Hyperlinked A-Z indexes

Hyperlinked index is a kind of back-of-book-index. It is arranged alphabetically A-Z. Normally, in a back-of-book index the terms or phrases are listed with the appropriate page number or section number. In web environment, A-Z index is a Webpage or a group of pages. Each entry in the Webpage is hyperlinked to a topic or to be more precise to the anchor tag of the resource. The list may also contain synonymous terms linked to the same resource.

If, hyperlinked A-Z Index is prepared manually the rendering of search term rectifies several searching problems like, spelling mistakes, spelling variants, singular plural and so on. There are following visible advantages with the hyper-linked A-Z Index.

- A-Z indexes are most user-friendly.
- The browsable nature of the index can reveal other topics of interest to the user.
- Index entries can link to precise points within a Webpage through the use of named anchor links.
- An A-Z index can enhance the search engine optimisation ranking of the website.

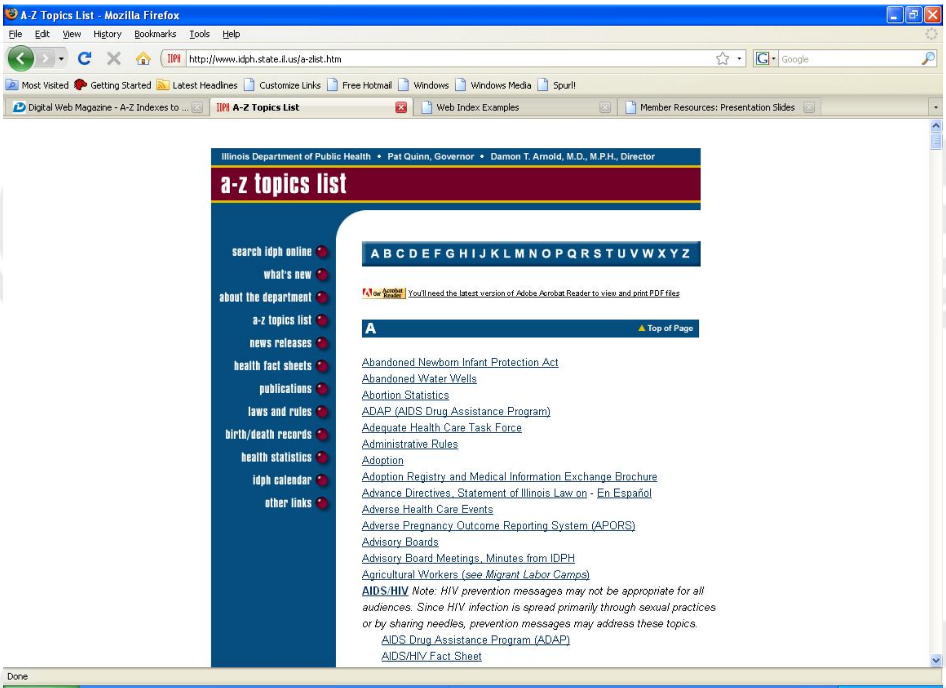


Fig. 14.2: Hyperlinked A-Z Index

Source: http://www.idph.state.il.us/a-zlist.htm

Meta-tag Keyword Indexing

Meta-tag is used in HTML (Hypertext Markup Language) documents for page description, keywords and other metadata. It is used in the header section of the web. The content of the tag is not visible on web browser. Metadata is normally referred as 'data about some object'. The object could be anything. The data about the object reflects the properties of the object. Some examples of objects in a bibliographic database are Title, Author, Place, Publisher etc.. Rendering of meta-tag in an HTML document is done as follows:

```
<html>

      <head>

            <title>Title of the webpage</title>

            <meta name="title" content="Indira Gandhi National
      Open University" />

            <meta name="author" content="Aditya Tripathi" />

      </head>
```

Fig. 14.3: Metatag in HTML Document

Note: Apart from describing the webpage, meta-tag is used for number of other purposes like, redirection from one page to other, handling the robot of search engines etc.

Robot is a program used by search engine in order to extract data from the web pages so that pages can be searched using the search engine's search interface. Robot is also known as Crawler or Spider. Following are the names of robots used by popular search engines,

Table 14.1: Robots used by Search Engines

Search engine	Robot
Google	Googlebot
Yahoo	Slurp
MSN	MSNbot

If the search engine is compliant with a metadata schema for example, Dublin Core then robot of the search engine extracts the metadata easily given on webpage and stores in the database of search engine.

Each metadata entry of webpage is used as an index term or phrase, further broken into keywords. With such an index context of the keywords or phrase is also extracted with the name of meta-tag. This kind of index is known as meta-tag keyword indexing.

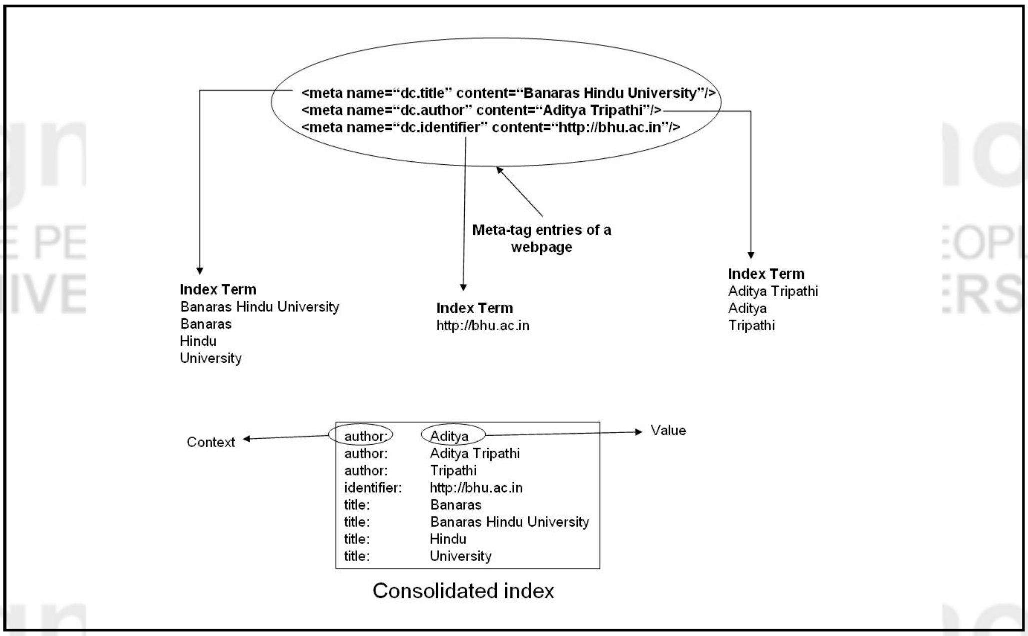


Fig. 14.4: Meta-tag Keywords Indexing

The advantages of metadata index is as follows:

- Context is preserved with the search term, which leads to precision in the search results. For example, documents on Ranganathan and documents by Ranganathan can be easily differentiated.
- Terms are extracted automatically from the web pages through a robot. This kind of index is useful in automatic indexing.

Keyword Creation for Search Engine Optimisation

Search engines look for search term which is queried in its database and fetches the result. If a document appears first in the order of search result then it is said that the page has better visibility. The ordering of results from the search engine is known as ranking. Webmaster who designs and maintains the website attempts to have best visibility in the search result. In order to achieve the visibility it is required to render relevant terms in the meta-tag element of webpages. This is known as Search Engine Optimization (SEO). Search Engine Optimization is of two types,

- White Hat SEO
- Black Hat SEO

White Hat SEO

Often search engines provide guidelines for the webmasters or content developers to have better visibility and ranking of website. For example, Google and Yahoo both provide guidelines for webmasters or content developers. If a webmaster follows these guidelines her/his website will get better visibility. In other words, White Hat SEO is a kind of web development technique that promotes accessibility.

Black Hat SEO

In order to improve upon ranking of webpage in search result many webmasters resort to unfair means of using heavy number of keyword count within the page. This is known as 'Spamdexing'. They often put more number of keywords in page with same colour as the background. Because of this keywords are not visible for human eye where as robots can read them. Similarly, webmasters play trick and present different webpage for search engine and human accesses to the website, deceiving search engines. This is known as cloaking. Search engines attempt to find out such kind of unethical methods of improving ranking and often lead to banning these websites.

Taxonomies/Categories

Taxonomy refers to the abstract structure of a subject. It is also referred as subject-based classification. Taxonomy typically displays the hierarchical structure of various components or sub-disciplines.

In a taxonomy like terms or subjects are grouped together so that finding the correct term becomes easy. It is used to identify the subject of the document. A typical, taxonomy is given as follows:

Library Science

+Classification

-Enumerated classification

-Analytico-synthetic classification

+Cataloguing

-Descriptive Cataloguing

-Simplified Cataloguing

```
Library Science
     +Classification
          -Enumerated classification
          -Analytico-synthetic classification

     +Cataloguing
          -Descriptive Cataloguing
          -Simplified Cataloguing
```

Fig. 14.5: Taxonomy

Use of taxonomy for the purpose of indexing, facilitates grouping the like objects or documents together. It displays all the objects or documents which belong to one category. Taxonomy is a kind of a controlled vocabulary. Hence, it can be also used as authority control.

Thesauri

Thesauri are also a kind of controlled vocabulary. Thesaurus is taxonomy with enhanced functionalities. Thesaurus demonstrates the relation of terms with respect to Broader Terms (BTs), Narrower Term (NTs), Related Terms (RTs), Synonymous Terms (SNs), Usage, Top Term (TT) and so on. The terms in a thesaurus are usually listed alphabetically.

Table 14.2: Terms in a Thesaurus

Broader Term	Broader in scope than the terms that are subordinate to it in a thesaurus hierarchy
Narrower Term	More specific concept than its parent term in the thesaurus hierarchy
Related Terms	A Preferred Term linked to another preferred term conceptually but not hierarchically
Top Term	The most general terms in a thesaurus hierarchy
Synonymous Terms	Term carries same meaning

Following example is taken from the thesaurus on agriculture, AGROVOC.

```
Pollution
     NT: Acid deposition
     NT: Air pollution
     NT: Nonpoint pollution
     NT: Sediment pollution
     NT: Water pollution
     RT: Environmental degradation
     RT: Pollutants
     RT: Pesticides
```

Fig. 14.6: Terms used in AGROVOC

Source: AGROVOC

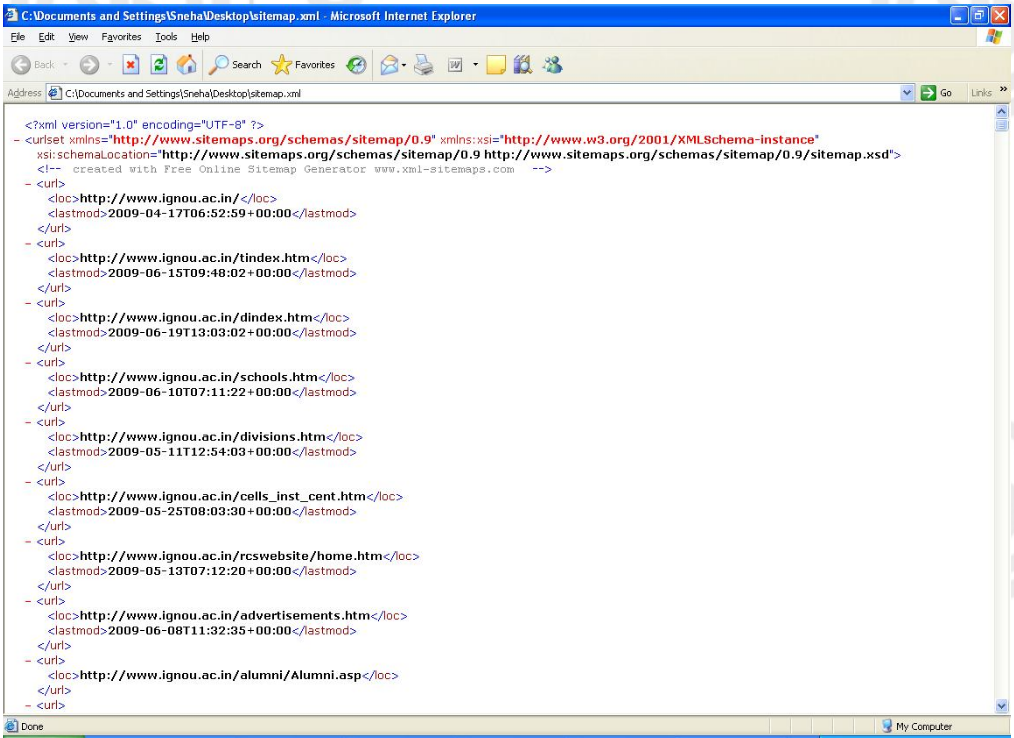
Using the thesaurus has the following benefits:

- ➢ An index with the help of thesaurus brings standardisation in rendering the terms.
- ➢ Use of thesaurus brings lot of relations like BT, NT, RT and so on which further leads in search refinement.
- ➢ If thesaurus is bilingual or multilingual, it can be used for text translation or cross-lingual information retrieval.
- ➢ Thesaurus can be used as authority control.

Sitemaps

A good website should be supplemented by good sitemap then only it is said to be complete. A sitemap displays structure of website and the flow of information in it. Hence, sitemap is a document detailing the various pages on a website and their links to each other. This helps the visitors both in finding and searching the pages. The use of the sitemap is to enhance browsing. Though this is the original idea of preparing site map but in due course of time the use of sitemap has changed a lot. Now it is used for exposing the hidden and dynamic content to the search engines using a 'sitemap index' file.

Sitemap index is an XML file (Extensible Markup File), which is prepared in a particular format and submitted to a search engine. There are programs available over the Internet which generates XML based sitemap index. This file can be downloaded and kept in the root directory, when search engine's crawler visits the website it picks up *sitemap.xml* file. Otherwise, it can be submitted directly to the search engines like Yahoo or Google. There may be some difference among the format of *sitemap.xml* file depending upon the search engines.




```
<?xml version="1.0" encoding="UTF-8" ?>
- <urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9 http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd">
   <!--  created with Free Online Sitemap Generator www.xml-sitemaps.com  -->
 - <url>
     <loc>http://www.ignou.ac.in/</loc>
     <lastmod>2009-04-17T06:52:59+00:00</lastmod>
   </url>
 - <url>
     <loc>http://www.ignou.ac.in/tindex.htm</loc>
     <lastmod>2009-06-15T09:48:02+00:00</lastmod>
   </url>
 - <url>
     <loc>http://www.ignou.ac.in/dindex.htm</loc>
     <lastmod>2009-06-19T13:03:02+00:00</lastmod>
   </url>
 - <url>
     <loc>http://www.ignou.ac.in/schools.htm</loc>
     <lastmod>2009-06-10T07:11:22+00:00</lastmod>
   </url>
 - <url>
     <loc>http://www.ignou.ac.in/divisions.htm</loc>
     <lastmod>2009-05-11T12:54:03+00:00</lastmod>
   </url>
 - <url>
     <loc>http://www.ignou.ac.in/cells_inst_cent.htm</loc>
     <lastmod>2009-05-25T08:03:30+00:00</lastmod>
   </url>
 - <url>
     <loc>http://www.ignou.ac.in/rcswebsite/home.htm</loc>
     <lastmod>2009-05-13T07:12:20+00:00</lastmod>
   </url>
 - <url>
     <loc>http://www.ignou.ac.in/advertisements.htm</loc>
     <lastmod>2009-06-08T11:32:35+00:00</lastmod>
   </url>
 - <url>
     <loc>http://www.ignou.ac.in/alumni/Alumni.asp</loc>
   </url>
 - <url>
```

Fig. 14.6: Sitemap.xml of IGNOU website

Sitemaps are important and beneficial at places where:

- ➢ some part of website is not visible due to use of dynamic scripts like Java pages or PHP pages, or

- pages where rich Ajax or Flash content is used.

Self Check Exercises

Note: 1) Write your answer in the space given below.

2) Check your answer with the answers given at the end of this Unit.

2) Discuss the different types of web indexes.

....

....

....

....

14.3 METADATA

14.3.1 Concept

Metadata is information about an informational resource, the document can be a webpage, image, dataset or other resource. Metadata is valuable towards storage and retrieval of documentary resources. Structured metadata make objects easily discoverable. In library parlance catalogue is known as metadata which provides descriptive information about an object or resource available in library. The resource can be physical or electronic. Metadata is a tool used to locate the object or document. There are various metadata schemas some are as follows:

- Anglo American Cataloguing Rule 2 (AACR2)
- MARC21
- Government Information Locator Service (GILS)
- Encoded Archive Description (EAD)
- Dublin Core

Hence, metadata is "data about data". It is structured set of data which describes the various characteristics of an object or document.

The most commonly used metadata schema is Dublin Core Metadata Initiative (DCMI) over the Internet. The standard is developed and maintained by DCMI and DCMI Task Groups. There are 15 elements given in Dublin Core. Apart from these 15 elements there are other metadata set vocabularies which should be used with 15 elements. The following example shows Dublin Core Metadata Record.

Table 14.3: Dublin Core Metadata Record

Creator	Aditya Tripathi
Publisher	Documentation Research & Training Centre
Identifier	http://drtc.isibang.ac.in
Subject	Library and Information Science
Format	txt/html
Language	English
Rights	Indian Statistical Institute

Purpose

Metadata is used for various purposes. These are to:

- retrieve a document;
- define the structure of document and its future maintenance;
- store the preservation conditions; and
- preserve the additional information regarding handling and usage of a document.

14.3.2 Types

As discussed above, it is common practice to use metadata for easy retrieval. But application of metadata has much more role to play in an electronic environment. Based on their roles, metadata are classified in the following types:

- Administrative metadata;
- Technical metadata;
- Structural metadata;
- Descriptive metadata; and
- Preservation metadata.

Administrative Metadata

When a document is created there are several kinds of information also generated with it. These information are valid and useful during the whole life span of the document. These data are stored in as administrative metadata. Administrative metadata is related with the life cycle of the document. It includes information regarding serials in the digital environment concerning:

- Ordering
- acquisition
- maintenance
- licensing
- rights
- ownership and
- provenance

Out of the above information 'rights' and 'digital provenance' are very important.

Technical Metadata

The technical metadata stores information regarding the file type and associated content type and how it should be rendered. It stores information regarding how the bytes should be read or in other words how the file should be read. Apart from this it also stores information regarding size or the extent of the file.

This information is very useful for playing the file. Further it is also useful for digital preservation particularly for migration and refreshing of the document. Technical metadata is helpful in checking the intactness of the object.

Structural Metadata

Structural metadata or structural map of an object explains different components and their role. This handles various sections and sub-sections of the documents and their corresponding relations and roles.

For example, structure of a book is defined as follows:

```
<mets:structMap TYPE="physical">
  <mets:div TYPE="book" LABEL="Martial Epigrams II">
      <mets:div TYPE="page" LABEL="Blank page">
      </mets:div>
      <mets:div TYPE="page" LABEL="Page i: Half title page">
      </mets:div>
      <mets:div TYPE="page" LABEL="Page ii: Blank page">
      </mets:div>
      <mets:div TYPE="page" LABEL="Page iii: Title page">
      </mets:div>
      <mets:div TYPE="page" LABEL="Page iv: Publication info">
      </mets:div>
      <mets:div TYPE="page" LABEL="Page v: Table of contents">
      </mets:div>
```

Fig. 14.7: Example of Structural Metadata

Descriptive Metadata

The metadata used for describing the documents in library is descriptive metadata. AACR2 or MARC21 are good example of descriptive metadata. In library parlance we call it descriptive cataloguing. Descriptive metadata stores information regarding title, author, place, and publisher and so on. This metadata set is important for identifying and locating the documents. For document location over the web, Uniform Resource Identifier (URI) is used. In case of traditional documents in library it is call number of the document where as for web document MARC21 defines field 856 for document location. Dublin core metadata elements have a field called *Identifier* used for document location.

Preservation Metadata

One of the most important metadata set used for digital longevity is preservation metadata. Digital preservation is process of increasing the longevity of documents from physical deterioration. The deterioration of digital objects is against time, technology, media and transfer. In order to secure the document and its original features, libraries, archives and museums need some kind of documentation in the form of metadata. Preservation metadata stores the preservation conditions of a document and its original features at the time of its digital provenance.

PREMIS (Preservation Metadata: Implementation Strategies) is the standard metadata set used for digital preservation purpose. It is joint venture of OCLC and RLG (Research Library Group). The working group comprised of experts of international repute working in digital preservation and metadata usage. The working group developed a core set of implementable preservation metadata and implementation guidelines in terms of creation, management and use of metadata. The working group came out with a set of Data Dictionary for Preservation Metadata. Current version of PREMIS is 2.0. However, PREMIS does not concentrate on descriptive metadata set because it is domain specific and second there are many descriptive metadata schemas available for use.

The preservation metadata is useful for following purposes:

- Supporting the viability, renderability, understandability, authenticity, and identity of digital objects in a preservation context;
- Representing the information most preservation repositories need to know to preserve digital materials over long-term;
- Emphasising "implementable metadata": rigorously defined, supported by guidelines for creation, management, and use, and oriented toward automated workflows; and
- Embodying technical neutrality: no assumptions made about preservation technologies, strategies, metadata storage and management, etc.

Self Check Exercises

Note: 1) Write your answer in the space given below.

2) Check your answer with the answers given at the end of this Unit.

2) Discuss the different types of metadata.

....

....

....

....

14.4 ONTOLOGY

14.4.1 Concept

Ontology is derived from the Greek *Onto* (being) and *Logia* (written or spoken discourse). It is part of metaphysics, branch of philosophy. Ontology studies existence of entities and their relationships. The relationship is derived due to grouping the entities based on formed groups. These groups are formed due to likeness or similarities of characteristics or attributes of individual entities. The relationship is depicted in the form of hierarchy and subdivisions. In other words, it is conceptualisation of world, based on entities and their mutual existence as it is studied in Philosophy.

However, ontology is also studied in computer science and information science. In computer science, ontology is the formal representation of a concept or a set of concepts within a specific domain of knowledge and their relationships. For example, Organisms are classified in two categories i.e. Plantae and Animalia. Then Animalia is further classified into Chordata and Non-Chordata. Chordata is further classified into Protozoa, Coelenterate and so on. This is taxonomy of animal kingdom.

One of the important parts of ontology is definition of classes and their properties. Class represents a group of concepts or objects having same kind of properties. Properties can be defined as distinguishing features for identification of a class or an individual object. Hence, it can be stated that the use of ontology is the use of classification for web documents. The term classification is also known as taxonomy.

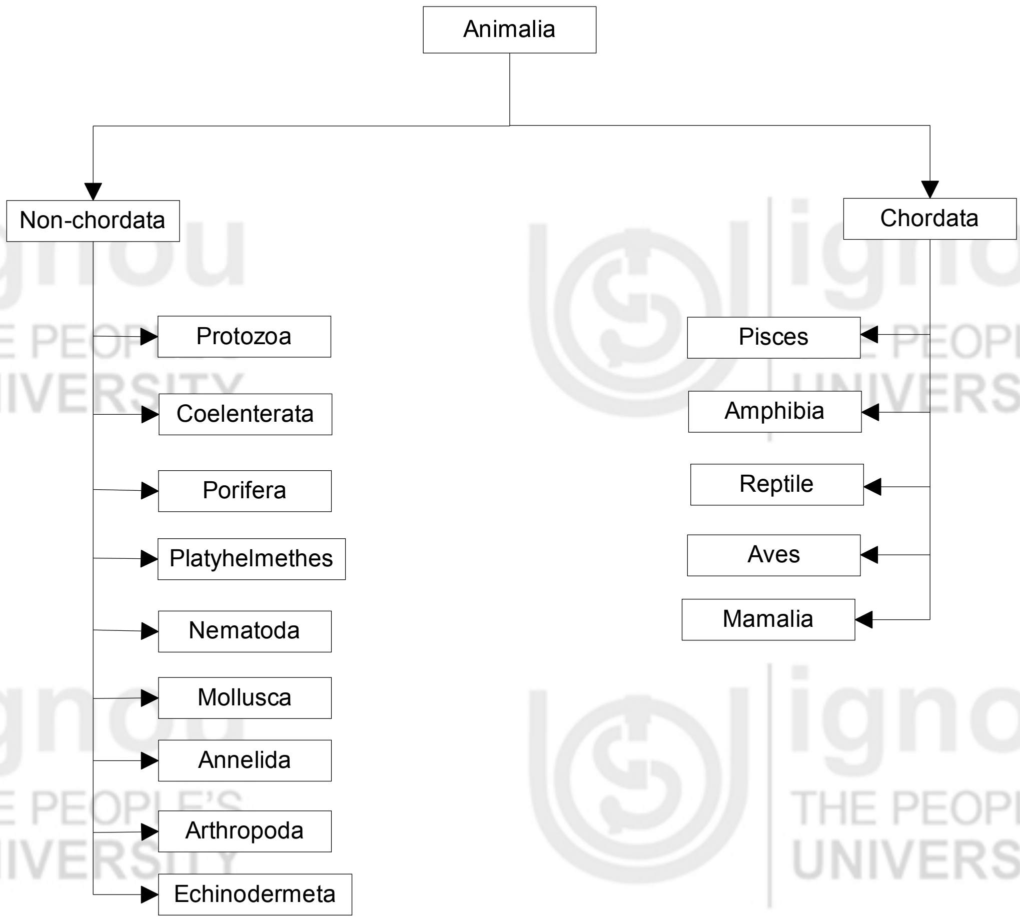


Fig. 14.8: Taxonomy of Animal Kingdom

In Computer Science, Ontologies are expressed in the languages that allow abstraction of concepts. Hence, ontology can be defined as "as a level of abstraction of data models, analogous to hierarchical and relational models, but intended for modeling knowledge about individuals, their attributes, and their relationships to other individuals". (Ref. 10)

14.4.2 Web Ontology

Presently, search engines perform searching over stored indexes in their databases with pattern match algorithm. This search lacks representation of concept with search term. This inherent problem is not due to any difficulty with search engines rather it is due to representation of data in webpage using Hyper Text Markup Language (HTML), the language of the Web. Hence, a mechanism is visualized to represent the data of web pages using another language i.e. Extensible Markup Language (XML) with a standard data description framework called as Resource Description Framework (RDF). It is understood that each individual web page can be considered as an entity and will have its attributes or characteristics. Based on this property the pages can be grouped and further they can form relation with other web page(s) or group of web pages. This develops a kind of web based ontology also known as web ontology for web documents but the original idea of ontology remains same. This framework uses standard vocabularies like Resource Description Framework Schema (RDFS) and Web Ontology Language (OWL) for describing the concepts and their relations with other concepts. The search engines extract the data from the web page and preserve the relation with the data, so that meaningful results can be generated.

14.4.3 Types

Generic Ontologies

Generic ontologies cover large spectrum of knowledge domains. They defines concepts at very broad level. Generic ontology represents broad concepts and their relationships. These ontologies are easy to reuse. Generic ontologies represent class of libraries which can be used with different problem domains and environment. These ontologies are like an umbrella ontology which can be further used for more specific purpose in conjunction with more specific ontology. These ontologies provide a mechanism for interoperability among different related ontologies. However, generic ontologies have following key features:

1) Generic ontologies are created from thesaurus, term dictionary or classification schemes and so on.

2) It provides logical concreteness and suitability for information interchange.

3) These ontologies don't provide any informational explanation for content used.

4) It is suitable to be used with more than one discipline or domain of knowledge.

5) Normally, this kind of ontology lacks sound principles of development or in other words, they follow popular approach.

Core Ontologies

With regard to ontological content there are two schools of thoughts. One claims that content depends highly on the context and hence any ontology prepared can work and only work with the same content or concept. However, the other school suggests that there are ontologies which follow minimal standard vocabulary. The vocabulary used is from philosophy or cognitive science. Hence, the used vocabulary is domain independent or in other words it is only dependent on philosophy and cognitive science. But the content it represents belongs to specific domain of knowledge. This kind of ontology is known as core ontologies.

The core ontology has been used to reach an agreement on the types of entities (and their relationships) needed in a community of practice. It is being used to dynamically negotiate the intended meaning across a distributed community. It has been used to align, integrate and merge several sources of metadata or ontologies. Hence, it can used to build more than one application or service. It can be adopted as a template for specifying the content in some domain.

Hence, the key features of core ontology can be given as:

- the core ontology specialises a foundational or top-level ontology
- the core ontology has been built through a well-motivated methodology that nonetheless avoids the reuse of a foundational ontology
- the core ontology has "built-in" (but explicit) criteria for well-foundedness.

Domain Specific Ontologies

Ontologies are developed keeping specific objectives in mind like, defining various components, describing specific functionality and so on. Often specific ontologies are required based on a specific knowledge sphere, area, field, region or realm. Such ontologies are known as domain specific ontolgoies. For example, ontology of organisational chart can be considered as domain specific ontology. The following example also demonstrates an ontology of still camera.

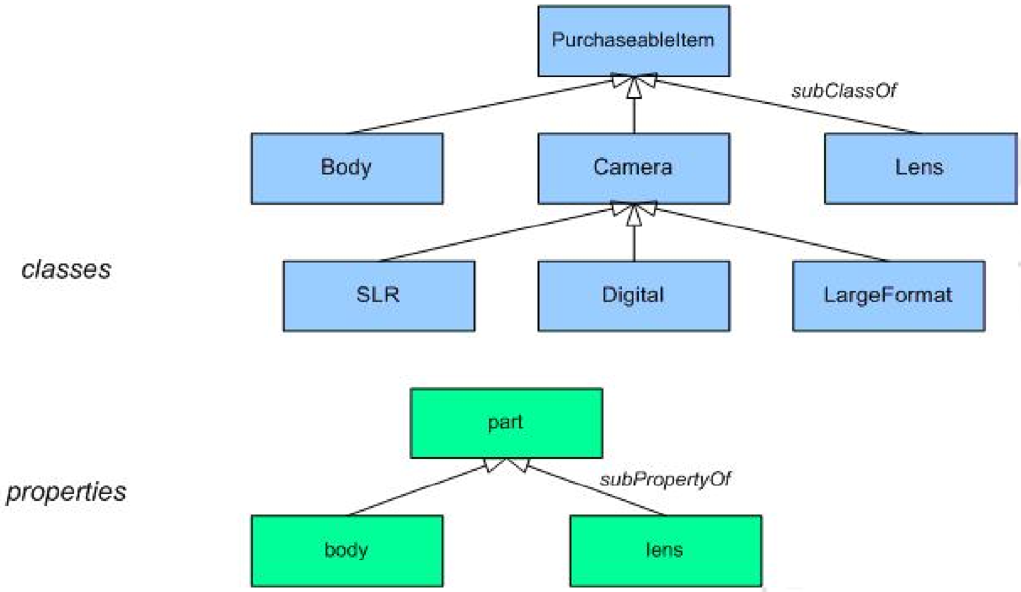


Fig. 14.9: Ontology of Still Camera

Source: http://www.seasr.org/wp-content/plugins/meandre/rdfapi-php/doc/tutorial/img/Camera-classes.png

Key features of Domain Specific Ontologies are as follows:

- It is restricted to a specific domain or area
- It highlights all the components and there interrelations of a domain
- It not only highlights components but also highlights their properties

Task Oriented Ontologies

In a more complex system, the operations are broken into different levels like top level, middle level and inner most level. Each level may have its own objectives as defined by the system analyst. Therefore, each level performs its individual task and transfers the output to next level. A conceptual framework of described system is known as Task Oriented Ontology. Like other ontologies it also consists of taxonomy and axioms. Axioms are rules for reasoning, principles, or constraints among the concepts. Hence, a task oriented ontology has three parts,

- Lexical level
- Conceptual level
- Symbol level

Lexical level

At lexical level task ontology provides human-friendly understanding in terms of which users can easily describe their own task. It provides comprehension for human readability and descriptiveness.

Conceptual level

At conceptual level task ontology simulates the various problem solving processes at the conceptual level and demonstrates the possible solutions through the rules or reasoning. It provides operationality only at conceptual level rather implementation or execution level.

Symbol level

This level provides operationality at implementation or execution level. The ontology makes system run the task description by translating it into instructions.

Self Check Exercises

Note: 1) Write your answers in the space given below.

2) Check your answers with the answers given at the end of this Unit.

4) What is ontology?

....

....

....

....

5) Discuss different parts of task oriented ontology.

....

....

....

....

14.5 INTEROPERABILITY

The term interoperability means working in collaboration. In a distributed service environment, different resources work together to produce a common service or goal. It is very common to understand different modules or services of an object/product. It is not desired to know how each module or service is functioning. But all together should able to work in collaboration to produce one service or product. The individual module or service must have enough common ground so that exchange of individual output can be shared without any error and misunderstanding. This requires standardisation of individual output according to some specifications. The standardisation provides common platform for exchange of communication or services. There are several definitions given for interoperability,

According to National Information Standard Organisation (NISO), "Interoperability is the ability of multiple systems with different hardware and software platforms, data structures, and interfaces to exchange data with minimal loss of content and functionality"

14.5.1 Need

In the library parlance, concept of interoperability is used since long. The use of MARC21 bibliographic standard or any other bibliographic standard, in conjunction with ISO 2709 or MARC format or MARC XML format provides facility to exchange bibliographic data among libraries. This exchange can be used in various ways like, generating a single platform based search facility for a number of libraries, reusability of library catalogue and so on.

The most important use of interoperability is seen in telephone industry. Irrespective of operators one can make phone calls or send messages. This is because of adherence to one kind of standard. Similarly, emails can be sent across different service providers because of Simple Mail Transfer Protocol (SMTP).

Another example of interoperability is seen in the field of industry. In medicine it is very important to record the case histories, hence an efficient Electronic Health Record (EHR) systems is required. If different hospitals record case histories under their individual specifications then communication and exchange of case studies among hospitals will become impossible. Hence, it is the need of the hour that medicine comes out with a standard for maintaining and managing EHR systems so that such systems can communicate among themselves.

14.5.2 Interoperability and Web Search

Interoperability has a major role to play in the Web parlance. Web is unorganised and a distributed environment. Information over the Internet is presented in HTML format. In order, to improve upon the search results of search engines, use of metadata schemas is thought of. But soon it was realised that there is a plethora of metadata standards. Though, Dublin Core metadata elements given by World Wide Web Consortium (W3C) evolved as a de facto standard for describing web documents but other standards have also made a mark over the scenario like, MARC21, Government Information Locator Service (GILS), e-Government Metadata Standard (e-GMS), Encoded Archival Description (EAD), Geospatial Metadata (GEO) and so on. It has been continuously observed and felt that search engines should come up with some kind of interoperability model so that cross standard search is possible. The concept is known as federated searching.

Currently, search engines read data from metatag <meta> of an HTML document or an XML file which describes the resource in Resource Description Format (RDF). RDF is used for defining onotologies in order to describe web resources and it has broader scope than the former ones.

14.5.3 Methods for Achieving Interoperability

Mapping/Matching

Mapping is one of the methods used for achieving interoperability. Mapping means relating or corresponding one to one between the entities of two sets. Mappings between two ontologies means establishing correspondence between each entity of ontology A against entities of ontology B with respect to their meanings. The mapping does not lead to creating new set of entities rather it only produces correspondence.

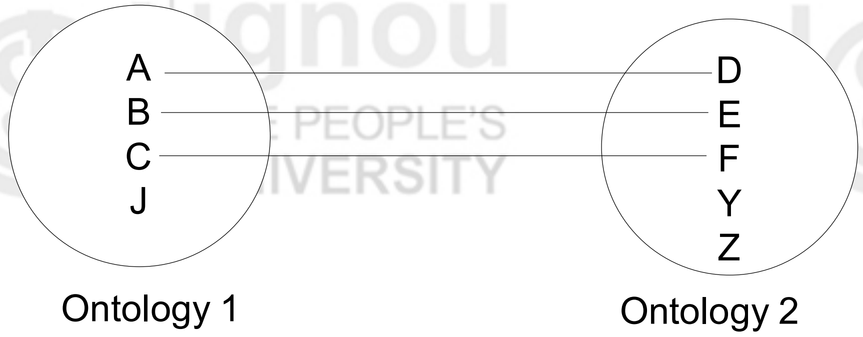


Fig. 14.10: Mapping of Ontologies

Alignment

Ontology alignment is a process of bringing different ontologies into mutual agreement. This process involves bringing ontologies together such that redundancies are removed and logical elements are kept. Hence, the process requires transformation of the involved ontologies. However, any element which is expected in the mutual ontology may also be included. Therefore, alignment might bring a complete new picture of ontology.

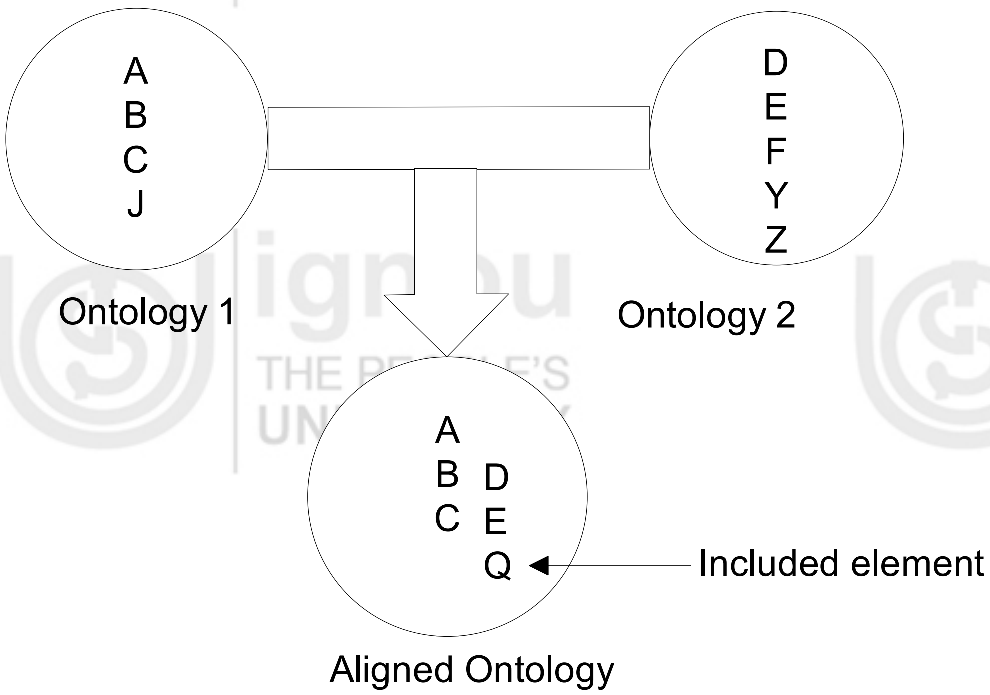


Fig. 14.11: Alignment of Ontology

Transformation

Transformation leads to complete change in the original ontology. The change may occur in terms of elements, attributes or concepts. Hence, the resultant ontology would be a completely new ontology based on the previous one. However, the degree of change in the structure or semantics may vary depending on situation. The process of transformation may be additive or subtractive depending on the needs of the original.

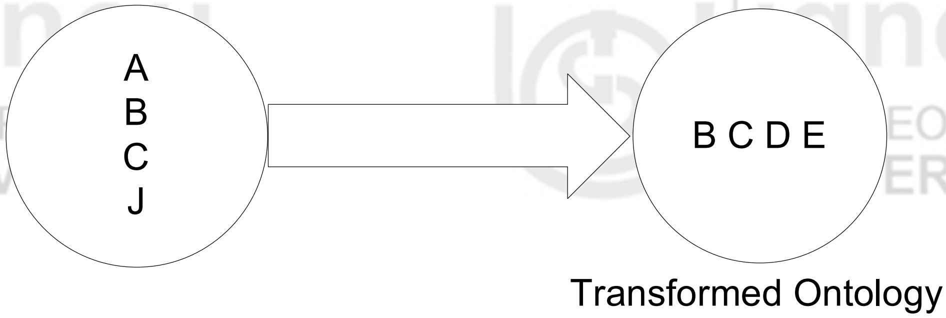


Fig. 14.12: Transformation of Ontology

Translation

Often, it so happens that ontology is to be used in different environments. The change of environment can be subject domain, software or language. In such a situation it is required that the original ontology is to be changed according to the new environment. However, it is expected that the conceptual meaning or semantics of the original will not change and remain as close as possible to the original.

Merging/Integrating

When two or more ontologies are merged together and form a new ontology it is known as merging or integrating of ontologies. This process leads to a formation of completely new ontology based on the previous once.

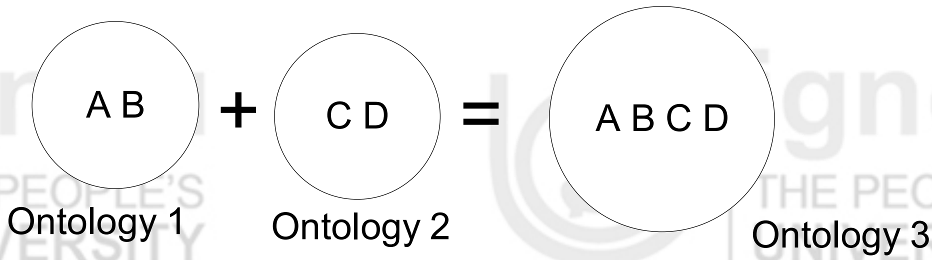


Fig. 14.13: Merging of Ontology

Self Check Exercises

Note: 1) Write your answer in the space given below.

2) Check your answer with the answers given at the end of this Unit.

6) Name different methods of interoperability of metadata.

....

....

....

....

14.5.4 Protocols for Interoperability

The Internet is a source of many online resources containing documents in different formats like, text, graphics audio and video. Individual resources hosting these documents may follow different metadata standards. These documents are to be searched using a search engine. Hence, a cross platform mechanism has been established in the form of protocols to perform searching different resources in one stroke. The whole such system is a distributed system and completely untamed. Developing search agents for such a system is a big challenge. The use of protocols allows users to search several data sources with single effort irrespective of the metadata standard used. Z39.50 (ZEE Thirty Nine point Five Zero), OAI-PMH (Open Archive Initiative and Protocol for Metadata Harvesting) and SRW/U (Search/Retrieve via the Web or URL) are developed for this purpose. Interoperability techniques are still being improved and becoming further sophisticated in order to provide more power and features in the hands of searchers.

There are two protocols which are widely used over the Internet for cross domain search:

- Z39.50 and ZING
- OAI-PMH

Z39.50 AND ZING

The core of interoperable searching is use of protocols. The use of Z39.50 is well accepted and oldest in library services. The protocol was developed to search Online Public Access Catalogues (OPACs) of different libraries. In due course of time, the protocol evolved with several applications like searching deep web (databases over Internet), publishers' catalogue, digital repositories and so on. The protocol performs real-time information retrieval from the source. A Z39.50 server (for example, Zebra server from Index Data www.indexdata.dk) is queried by a Z39.50 client (for example, Yaz client from Index Data www.indexdata.dk). The client searches various Z39.50 servers individually and presents the results of all the servers collectively (refer Fig.14) The server hosting the data must be available in case of using Z39.50 protocol at the time of searching. The only tricky issue in using Z39.50 is mapping of different standards.

The next generation of the Z39.50 protocol is ZiNG (Z39.50-International: Next Generation) maintained by Library of Congress. The protocol is an encapsulation of three protocols

- Z39.50
- Search and Retrieve Web Service (SRW)
- Search and Retrieve URL Service (SRU)

The protocol exploits features of the Z39.50 and web technology. The need of searching multiple domains using WWW created the scope of expansion of Z39.50. The SRU is simple method of searching the Web using GET method and HTTP. The request is carried in name and value pair through URL. The SRW carries a request in the form of a packet known as SOAP (Simple Object Application Protocol). In both the cases result is thrown back in an XML format. The difference between SRW and SRU is that SRW returns XML stream encapsulated in SOAP envelop. The databases are queried with a standard Common Query Language (CQL), a language for database searching. The protocol is supported by several of the search agents specific to libraries. The most important is Library of Congress.

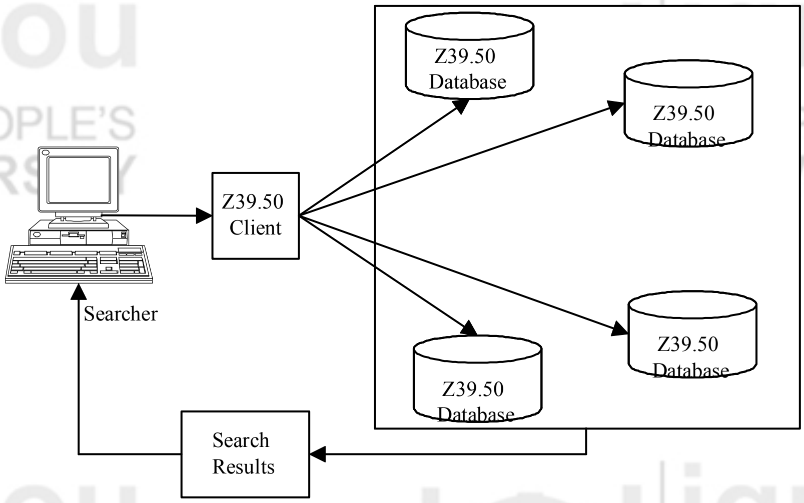


Fig. 14.14 : Model of Basic Z39.50 Protocol

Open Archive Initiative and Protocol for Metadata Harvesting (OAI – PMH)

This is also an HTTP embedded protocol used extensively for interoperable searching and retrieval. The protocol is simple and developed for searching across digital repositories. OAI-PMH intermediary (service provider or search agent) harvests metadata in anticipation from the distributed resources and offers search to the clients. Harvesting means extracting metadata from different resources and storing inside own database. Searching is done by intermediary service provider using its own harvested metadata. However any request for the searched document is directed to the resource or repositories. In case of using OAI-PMH search can also be made to a resource even if that is not available. The result is returned with an XML data format. OAI-PMH supports Dublin core elements.

This protocol has been supported by many digital libraries around the world. OAIster harvester of University of Michigan provides 15,601,208 records from 944 repositories. In India, Search Digital Libraries (SDL) at Documentation Research & Training Centre, Bangalore provides 21220 records from 9 different repositories. This service is first of its kind in India. DP9 at Old Dominion Universtiy is a harvester to enable search engines harvests records from OAI-PMH repositories.

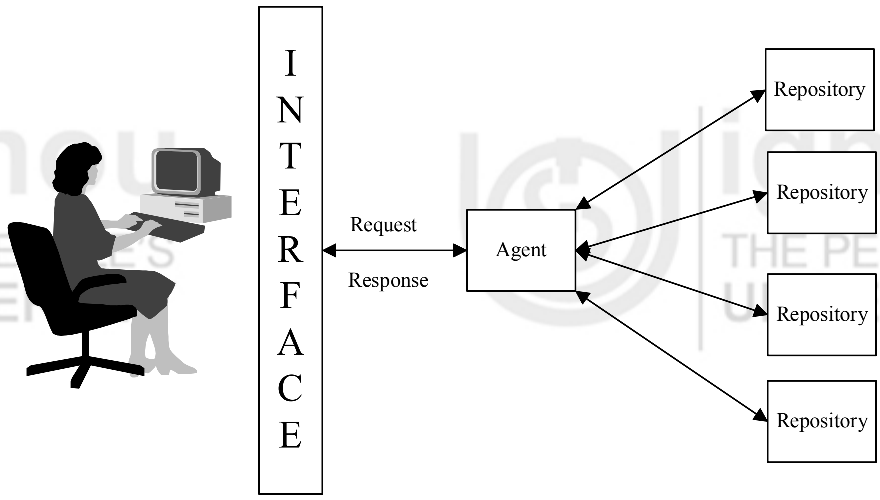


Fig. 14.15: Model of Metadata Harvesting with OAI-PMH

14.6 SUMMARY

The objective of web indexing is to make a website searchable and browsable for the intended users. Web index helps search engines to store more meaningful keywords about the website. Search engines are automated tools for searching the web pages and metadata is one approach towards rendering more semantic based keyword to the search engines. Initially, meta tag <meta> was used for rendering the keyword within a webpage. But it is realised that this approach fails to store the context of search terms. Henceforth, use of ontologies came into being and Resource Description Framework (RDF) is used for representation of intended knowledge with context for automated data extraction by search engines. Sooner it is realized that there is going to be flood of ontologies and metadata schemas which gave rise to the concept of interoperability among the standards.

14.7 ANSWERS TO SELF CHECK EXERCISES

1) Index has a great role to play in the web parlance:

   - Index demonstrates the relationship of topics.
   - Index is used as a source for searching. The search engines use index terms stored inside their database.
   - A good index provides visibility of all the available literature.
   - Index brings material related to what user is looking for, with the help of related terms or concepts.
   - Index focuses on the content of the document and demonstrates the inconsistencies about the treatment of the topics. Helping authors to review the writing for completeness of content.

2) There are different types of Web Indexes:

   a) **Hyperlinked A-Z indexes**

   In web environment, A-Z index is a web page or a group of pages. Each entry in the web page is hyperlinked to a topic or to be more precise to the anchor tag of the resource. The list may also contain synonymous terms linked to the same resource. In an alphabetical index, terms may be written in normal order or in displaying other suitable order.

   b) **Meta-tag keywords indexing**

   Metadata is normally referred as 'data about some object'. The object could be anything. The data about the object reflects the properties of object for example, Title, Author, Place, Publisher etc. are used to describe a book. Each metadata entry of webpage is used as an index term or if phrase, is further broken in keywords. With such index context of the keywords or phrase is also extracted with the name of meta-tag. This kind of index is known as Meta-tag keywords indexing

   c) **Keyword creation for search engine optimization**

   Search engines look for search term which is queried in its database and fetch the result. If a document appears first in the order of search result then it is said that the page has better visibility. An attempt to get better visibility is known as search engine optimization.

d) **Taxonomies/categories**

Taxonomy refers to abstract structure of subject. It is also referred as subject-based classification. Taxonomy typically displays the hierarchical structure of various components or sub-disciplines. Use of taxonomy for the purpose of indexing, facilitates grouping the like objects or documents together. It displays all the objects or documents which belong to one category. Taxonomy is a kind of Controlled vocabulary. Hence, it can be also used as authority control.

e) **Thesauri**

Thesauri are also a kind of controlled vocabulary. Thesaurus is taxonomy with enhanced functionalities. Thesaurus demonstrates the relation of terms with respect to Broader Terms (BTs), Narrower Term (NTs), Related Terms (RTs), Synonymous Terms (SNs), Usage, Top Term (TT) and so on. The terms in a thesaurus are usually listed alphabetically. (Ref. 2)

f) **Site maps**

Sitemap index is an XML file (Extensible Markup File), which is prepared in a particular format and submitted to search engine. There are programs available over Internet which generates XML based sitemap index. This file can be downloaded and kept in the root directory, when search engine's crawler visits the site it picks up *sitemap.xml* file.

3) There are different kinds of metadata.

There are different types of metadata:

Administrative metadata

When a document is created there are several kinds of information also generated with it. These information are valid and useful during the whole life span of document. These data are stored in as Administrative metadata

Technical Metadata

The technical metadata stores information regarding the file type and associated content type and how it should be rendered. It stores information regarding the how the bytes should be read or in other words how the file should be read. Apart from this it also stores information regarding size or the extent of the file.

Structural Metadata

Structural metadata or structural map of an object explains different components and their role. This handles various sections and sub-sections of the documents and their corresponding relations and roles.

Descriptive Metadata

The metadata used for describing the documents is descriptive metadata. AACR2 and MARC21 are good example of descriptive metadata. Descriptive metadata stores information regarding title, author, place, publisher and so on. This metadata set is important for identification for locating the documents.

Preservation Metadata

One of the most important metadata set used for digital preservation is Preservation metadata. Digital preservation is process of increasing longitivity of documents from physical deterioration. The deterioration of digital objects is against time, technology, media and transfer. In order to secure document and its original features libraries, archive and museum need some kind of documentation in a form of metadata.

4) Ontology studies of existence of entities and their relationships. The relationship is derived due to grouping the entities based on formed groups. These groups are formed due to likeness or similarities of characteristics or attributes of individual entities. The relationship is depicted in form of hierarchy and subdivisions.

5) There are three parts of Task Oriented Ontology:

Lexical level

At lexical level task ontology provides human-friendly understanding in terms of which users can easily describe their own task. It provides comprehension for human readability and descriptiveness.

Conceptual level

At conceptual level task ontology simulates the various problem solving processes at the conceptual level and demonstrates the possible solutions through the rules or reasoning. It provides operationality only at conceptual level rather implementation or execution level.

Symbol level

This level provides operationality at implementation or execution level. The ontology makes system run the task description by translating it into instructions.

6) Different methods of Interoperability of metadata are:

- Mapping/matching
- Alignment
- Transformation
- Translation
- Merging/integrating

14.8 KEYWORDS

Bandwidth : In computer networks, bandwidth is often used as a synonym for data transfer rate - the amount of data that can be carried from one point to another in a given time period (usually a second). This kind of bandwidth is usually expressed in bits (of data) per second (bps).

Browser : A Client program (software) that is used to look at various kinds of Internet resources.

Client : A software program that is used to contact and obtain data from a Server software program on another computer, often from a distance.

Domain Name : The unique name that identifies an Internet site. Domain Names always have 2 or more parts, separated by dots. The part on the left is the most specific, and the part on the right is the most general. For example *ignou.ac.in*

Download	:	Transferring data (usually a file) from one computer to another computer.
Email	:	Also known as Electronic Mail, is messages, usually text, sent from one person to another via computer. E-mail can also be sent automatically to a large number of addresses.
Home Page (or Homepage)	:	Originally, the web page that your browser is set to use when it starts up. The more common meaning refers to the main web page for a business, organisation, person or simply the main page out of a collection of web pages i.e. index page.
Host	:	Any computer on a network that is a repository for services available to other computers on the network.
HTML (HyperText Markup Language)	:	The coding language used to create hypertext documents for use on the World Wide Web.
HTTP (HyperText Transfer Protocol)	:	The protocol for moving hypertext files across the Internet.
Hypertext	:	Generally, any text that contains links to other documents - words or phrases in the document that can be chosen by a reader and which cause another document to be retrieved and displayed.
Internet	:	The vast collection of inter-connected networks that are connected using the TCP/IP protocols and that evolved from the ARPANET of the late 60's and early 70's. Also known as 'network of networks'.
Meta Tag	:	A specific kind of HTML tag that contains information not normally displayed to the user. Meta tags contain information about the page itself, hence the name ("meta" means "about this subject") Typical uses of Meta tags are to include information for search engines to help them better categorize a page.
Network	:	Any time connecting more than one computer together so that they can share resources, is known as computer network.
Protocol	:	Protocols are rules define an exact format for communication between computers. For example HTTP protocol defines the format for communication between web browsers and web servers.

RDF (Resource Definition Framework)	:	The Resource Description Framework (RDF) is a general framework for how to describe any Internet resource such as a Web site and its content. An RDF description (such descriptions are often referred to as metadata, or "data about data") can include the authors of the resource, date of creation or updating, the organisation of the pages on a site (the sitemap), information that describes content in terms of audience or content rating, key words for search engine data collection, subject categories, and so forth.
Search Engine	:	A (usually web-based) system for searching the information available on the Web.
SEO (Search Engine Optimization)	:	The practice of designing web pages so that they rank as high as possible in search results from search engines.
Server	:	A computer, or a software package, that provides a specific kind of service to client software running on other computers.
SMTP (Simple Mail Transfer Protocol)	:	The main protocol used to send electronic mail from server to server on the Internet.
SOAP (Simple Object Access Protocol)	:	A protocol for client-server communication that sends and receives information "on top of" HTTP. The data sent and received is in a particular XML format specifically designed for use with SOAP.
SQL (Structured Query Language)	:	A specialized language for sending queries to databases.
Terminal	:	A device that allows you to send commands to a computer somewhere else.
URI — (Uniform Resource Identifier)	:	An address for s resource available on the Internet.
URL — (Uniform Resource Locator)	:	The term URL is basically synonymous with URI. URI has replaced URL in technical specifications.
URN — (Uniform Resource Name)	:	A URI that is supposed to be available for a long time. For an address to be a URN some institution is supposed to make a commitment to keep the resource available at that address.
Web page	:	A document designed for viewing in a web browser. Typically written in HTML. A web site is made of one or more web pages.
Website	:	The entire collection of web pages and other information (such as images, sound, and video files, etc.) that are made available through what appears to users as a single web server.

XML (eXtensible Markup Language) : A widely used system for defining data formats. XML provides a very rich system to define complex documents and data structures such as invoices, molecular data, news feeds, glossaries, inventory descriptions, real estate properties, etc. As long as a programmer has the XML definition for a collection of data (often called a "schema") then they can create a program to reliably process any data formatted according to those rules. Libraries use XML for bibliographic data exchange.

Z39.50 A NISO and ISO standard protocol for cross-system searchand retrieval. Officially, international standard, ISO 23950, Information Retrieval (Z39.50): Application Service Definition and Protocol Specification, and ANSI/ NISOstandard Z39.50.

14.9 REFERENCES AND FURTHER READING

A-Z Indexes to Enhance Site Searching Web. 24 September 2012. <http://www.digital-web.com/articles/a_z_indexes_site_searching/>

AFS Ethnographic Thesaurus. Web. 24 September 2012. <http://www.afsnet.org/?page=AFSET >

Berners-Lee, T., Hendler, J., and Lassila, O. "The Semantic Web". Scientific American, 284.5 (May 2001) 34-43. Print.

Bray, T. "Measuring the Web". Computer Networks and ISDN Systems. 28(7-11-May), 992-1005, 1996. Print.

Browne, Glenda Michelle. "Indexing Web Sites: A Practical Guide". Internet Reference Quarterly. 5.3 (Sep 2001): 27-41. Print.

Diego Calvanese, Giuseppe De Giacomo, and Maurizio Lenzerini. "Description Logics for Information Integration". Ed. Kakas, A.C and F. Sadri. Computational Logic: Logic Programming and Beyond: Kowalski Festschrift, LNAI 2407. Berlin: Springer Verlag , 2002. 41–60. Print.

Genesereth, M. R. and Nilsson, N. J. Logical Foundations of Artificial Intelligence. San Mateo, California: Morgan Kaufman Publishers, 1987. Print.

Gruber, T., "A Translation Approach to Portable Ontology Specifications". Knowledge Acquisition 5.2 (1993), 199-220. Print.

Gruber, T., "Towards Principles for the Design of Ontologies Used for Knowledge Sharing". International Journal of Human and Computer Studies 43. 5/6 (1995), 907-928. Print.

Gruber, Tom "Ontology". In the Encyclopedia of Database Systems, Ed. Ling Liu and M. Tamer Özsu . Springer-Verlag, 2009. Web. 24 September 2012. <http://tomgruber.org/writing/ontology-definition-2007.htm>

Guarino, N., "Formal Ontology in Information Systems", Proceedings of FOIS'98, Formal Ontology in Information Systems, Trento, 3-15, 1998. Print.

Halpin, T.A. Information Modeling and Relational Databases: From Conceptual Analysis to Logical Design. San Francisco, California: Morgan Kaufman Publishers, 1999.

Hruby, Pavel. Ontology-based Domain-driven Design. Web 24 September 2012. <www.softmetaware.com/oopsla2005/hruby.pdf>

Kohler, Jacob. "Ontology Based Text Indexing and Querying for the Semantic Web". Knowledge-Based Systems 19 (2006) 744–754. Web. 24 September 2012. <http://www.researchgate.net/publication/222566181_Ontology_based_text_indexing_and_querying_for_the_semantic_web>

Pretorius, A. Johannes. Ontologies: Introduction and Overview. Web. 24 September 24 2012.

http://www.starlab.vub.ac.be/teaching/Ontologies_Intr_Overv.pdf

Maedche, A.D. Ontology Learning for the Semantic Web. Norwell, Massachusetts: Kluwer Academic Publishers, 2003. Print.

Meersman, R., "Can Ontology Theory Learn from Database Semantics?" Proceedings of the Dagstuhl Seminar 0121, Semantics for the Web, Dagstuhl, 2000. Print.

Meersman, R., "Ontologies and Databases: More than a Fleeting Resemblance" Ed. D'Atri,A. and Missikoff, M. OES/SEO 2001 Workshop, Rome, 2001. Print.

Mitchell, W.J. City of Bits: Space, Place and the Infobahn. Cambridge, Massachusetts: MIT Press, 1999. Print.

Noy, N.F. and Hafner, C.D. "The State of the Art in Ontology Design – A Survey and Comparative Review". AI Magazine. 36.(1997), 53-74. Print.

Ontologies and Knowledge Base. http://briefs.cs.hut.fi/phase8/Ontologies_and_KB/noframes.html

Ontology (information science) http://en.wikipedia.org/wiki/Ontology_(computer_science)

Ontology Interoperability. Draft version 0.3.2. Web. 24 September 2012. <www.dsi.uniroma1.it/~estrinfo/3.%20Ontology%20merging.pdf>

Pisanelli, D.M., Gangemi, A. and Steve, G., "Ontologies and Information Systems: The Marriage of the Century". Proceedings of the LYEE Workshop, Paris, 2002. Print.

Maislin, Seth A. Indexing Online: The New Face of an Old Art. Web. 24 September 2012. < http://taxonomist.tripod.com/websmarts/onlineindexing.html>

Shahar, Yuval, Miksch, Silvia and Johnson, Peter. A Task-Specific Ontology for the Application and Critiquing of Time-Oriented Clinical Guidelines. Web 24 September. <www.springerlink.com/index/d4483r305758786r.pdf>

Summit on Serials in the Digital Environment. Web. http://www.loc.gov/acq/conser/glossary.html

Understanding Metadata. ISBN 1-880124-62-9 www.niso.org/standards/resources/UnderstandingMetadata.pdf

Web Indexing SIG (Special Interest Group). www.webindexing.org

Wikipedia. http://en.wikipedia.org/wiki/Web_indexing

Taylor, Chris. Introduction to Metadata. Web. 24 September 2012. <http://www.itb.hu/fejlesztesek/meta/hgls/core/Background/An_Introduction_to_Metadata.htm>

Zhao, G. and Meersman, R. "Ontology Modelling for Collaborative Business Processes". StarLab Technical Report. Brussels: Vrije Universiteit Brussel, 2003. Print.