339en5B16

16

INFORMATION RETRIEVAL TOOLS: CATALOGUES, INDEXES, SUBJECT HEADING LISTS

16.1 INTRODUCTION

In the previous lesson, you have studied about information retrieval system which is designed to retrieve documents or information required by the users. You have learnt that the IRS should make the right information available to the right user at the right time. Information retrieval tools are basic building blocks for a system that organize recorded information collected by information organizations. This is to establish control of contents for information use as well as for promotion of users' ease of access. The basic information retrieval tools include: bibliographies, catalogues, indexes, finding aids, registers, online databases, etc. In this lesson, you will be introduced to information retrieval tools, viz. catalogues, indexes, subject heading lists and thesauri with examples.

16.2   OBJECTIVES

After studying this lesson, you will be able to:

- explain the use of various information retrieval tools, viz. catalogues, indexes and lists of subject headings;

- define   subject cataloguing;

- describe the controlled indexing language;

- illustrate types of controlled indexing languages like lists of subject headings and thesauri; and

- explain different types of indexes, viz. book index, author index, title index, and subject index.

16.3 INFORMATION RETRIEVAL TOOLS

In order to organize knowledge, librarians and information professionals have to create a variety of tools. Traditionally, the tools of information retrieval have been catalogues, bibliographies and printed indexes. Presently, computerized databases and their indexes are also important in the organization of knowledge. These are gradually replacing the traditional tools in a number of applications. At this point, the traditional tools and the computer based tools provide a unified approach to the organization of knowledge which are discussed in the following sections.

16.3.1 Catalogues

Catalogues are the windows to the library collection. A catalogue is the record of the collection in the library. It is also a systematic arrangement of items in an alphabetical or other logical order including brief description. A library catalogue is a list of books and other reading material available in a particular library. The card catalogue has been a familiar sight to library users for generations. But it has been effectively replaced by the online public access catalog (OPAC). The library catalogue of one particular library alone may be available in different physical forms to cover different periods of time. For example, an early catalogue may be in card form, and later superseded for recent documents by an online catalogue.
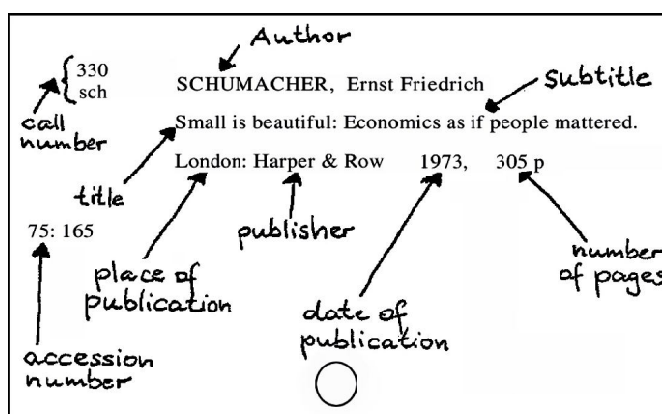
Fig 16.1: Snapshot of Library Catalogue Card

A library catalogue comprises of a number of entries, each entry representing or acting as a surrogate for a document as shown in Fig16.1 There may be several entries per document, or merely one. The types of catalogues found in libraries include the author, title, author/title and subject catalogues as given below.

- **Author catalogues** contain entries with authors' names as the heading. Authors may be persons or corporate bodies and the term author is normally extended to include writers, illustrators, performers, producers, translators, and others with some intellectual or artistic responsibility for a work.

- **Title catalogue** has entries with title as the heading. Some libraries and information centers make title entries for all items being indexed, but in other situations title entries are made selectively for only one type of material.

- **Author/Title catalogues** contain both title and author entries. As both titles and authors' names are in alphabetical order, it is easy to file together authors' names and titles as headings.

- **Subject catalogues** have an indication of the subject of the documents being indexed as their headings. The entries are arranged in an appropriate systematic order. There are two significantly different types of subject catalogues, (i) *Alphabetical subject catalogues* have headings which are words or index terms designed to summarize the subject content of the document. For example, car, lawyers. These entries are arranged alphabetically, according to the subject heading, and (ii) *Classified subject catalogues* have headings on entries which are classification symbols, for example 020 (Library science), 200 (religion), which have been drawn from a classification scheme (Dewey Decimal Classification). In a classification scheme, each subject is allocated a piece of notation, and that notation is used to represent the subject. The headings will be arranged according to the filing sequence of the notation.

The catalogues that have been described above are in single sequence. A catalogue for a complete library collection will normally combine a number of these single sequence catalogues. As you have already studied in Lesson 10, there are two types of catalogue, namely a classified catalogue and a dictionary catalogue.

- **Classified catalogue** is a catalogue with three or four separate sequences, (i) an author/title catalogue or index (or separate author and title catalogue), (ii) a classified subject catalogue, and (iii) a subject index to the classified catalogue.

- **Dictionary catalogue** is a catalogue with only one sequence which has author, title and alphabetical subject entries interfiled. As all of the headings are in alphabetical order, it is possible to file together entries regardless of the nature of their heading.

Notes

16.3.2 Indexes

An index is essentially a filter or pointer, or indicator, or a systematic guide to the items contained in, or concepts derived from a collection. Another dictionary definition is that an index is an alphabetical list of terms usually at the end of a book along with the page number where the term appears in the book. Subject indexing is a classification process. On the basis of their subject matter, we group together (perhaps physically, as in the classified arrangement on library shelves) similar subject documents and separate them from other documents. Then we label these classes of related documents in order to maintain the established groups, and to refer to them. In other words, we name the classes and the names we give to them are our index terms.

16.3.3 List of Subject Headings

In traditional library practice, when a controlled vocabulary is set up in the form of an alphabetical listing of index terms, the individual terms are known as *subject heading* and the controlled vocabulary as a *list of subject headings*. Subject headings lists are useful to understand the relationship among concepts to a certain degree, besides their application in indexing.

Subject headings lists are highly valuable for indexing. Subject headings are provided in the catalogue entries to provide subject access to information. Cataloguers depend on Lists of Subject Headings from which they can assign subject headings to the catalogued documents. The conceptual relationships are indicated in the list and the choice of terms and preference are given. In recent decades, these lists have also introduced many thesaurus features. The *Library of Congress Subject Headings (LCSH) List* is one of the best tools for indexing and retrieval. The *Sears List of Subject Headings (SLSH)* is a shorter version useful for smaller libraries. *Medical Subject Headings (MESH)* and *Subject Headings in Engineering (SHE)* are some of the other specialized ones highly used. In the Social sciences, the *Public Affairs Information Service (PAIS) Subject Headings List* is extensively used.

INTEXT QUESTIONS 16.1

1. Define a library catalogue.

2. Which are the different types of catalogues commonly found in libraries?

3. Explain an Index.

4. What do you understand by the term 'Subject headings lists'?

5. Mention any three subject heading lists.

16.4 SUBJECT CATALOGUING

Subject cataloguing deals with what a book or other library item is about. The purpose of subject cataloguing 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 that uniform word or phrase used in the library catalogue to express a topic. The use of authorized words or phrases only, with cross-reference from unauthorized synonyms, is the essence of bibliographic control in subject cataloging.

As mentioned above, subject cataloguing is the process of preparing subject entries for documents and organizing them for subsequent retrieval. It involves assignment of subject headings and references like *see* and *see also,* that helps in retrieval of documents from the catalogue in cases where users request for documents on specific subjects.

16.4.1 Objectives

Subject cataloguing has a lot of advantages with regard to retrieval of information from documents, when the user's approach is only to find information if a subject is known. It enables users seeking information to identify and provide access to all the documents on a subject. It also brings together all the related materials on a subject at one place, thus making things easier for the users. Subject cataloguing also enables access of materials if the users are using various kinds of vocabulary, i.e., synonyms, homonyms, etc. Besides this, user can always retrieve the latest information as all related information on a subject is at one place.

When we index using a classification schedule as controlled vocabulary, we assign class labels to a document to represent its subject matter. For convenience, however, we use class number in place of the natural language terms. The process has thus become known as classification, whereas the assignment of subject headings is generally referred to as *subject indexing* or *subject cataloguing.*

16.5 INDEXING LANGUAGES

Indexing language is the language used to describe a subject or other aspects of information in a library catalogue or an index. Indexing language is defined as- 'a list of terms or notations that might be used as access points in an index.' An indexing language may also be referred to as a retrieval language. As discussed in lesson 15, there are three main types of indexing languages:

- Controlled indexing language - Only approved terms can be used by the indexer to describe the document

- Natural language indexing language - Any term from the document in question can be used to describe the document.

- Free indexing language - Any term (not only from the document) can be used to describe the document.

In this lesson, controlled indexing language is discussed in detail.

16.5.1 Controlled Indexing Language

Control is necessary in respect of terms used as subject identifiers in a catalogue or index, because of the variety of natural languages. Such control may involve barring of certain terms from use as headings or access points in a library catalogue or an index. The terms which are to be used are specified and the synonyms recognized and as far as possible are eliminated. Preferred word forms are noted. The list of terms, thus, prepared constitutes what is called *controlled indexing language*.

Controlled indexing languages are indexing languages in which both the terms that are used to represent subjects, and the process whereby terms are assigned to a particular document are controlled or executed by a person. Normally there is a list of terms which acts as the authority list in identifying the terms that may be assigned to documents, and indexing involves a person assigning terms from this list to specific documents. There are two types of controlled indexing languages, namely, alphabetical indexing languages and classification schemes. In alphabetical indexing languages, terms are embodied in thesauri and subject headings lists. The subject terms are the alphabetical names of the subjects. Control is exercised over terms that are used, otherwise the terms are ordinary words. In classification schemes each subject is assigned a piece of notation. The usual objective of assigning notation is to place a subject within a context with respect to other subjects. Both types of device can be found applied in catalogues, indexes to books and periodicals, bibliographies, current awareness bulletins, selective dissemination of information, computerized databases and data banks, abstracting and indexing services, encyclopedias, dictionaries and directories. Classification is also prominent in the physical arrangement of documents. There are different examples of controlled Indexing languages, viz. Lists of Subject Headings, Classification schemes, Thesauri, Thesaurofacet and Classaurus.

Lists of subject headings and thesauri are explained below with examples.

16.5.1.1 Lists of Subject Headings

The purpose of subject headings is to give the cataloguer a way to describe the content of materials in the library. Many times this can be done with a single word, such as ARCHITECTURE. Sometimes, the concepts that are covered require subject headings that can provide further definition, or a more specific definition of the content. Subject headings can also have subdivisions that help to clarify the meaning or focus of the subject term. Some examples of lists of subject headings are:

(a) Library of Congress List of Subject Headings (LCSH)

The *Library of Congress Subject Headings (LCSH)* comprises a thesaurus of subject headings, maintained by the *United States Library of Congress* for use in bibliographic records. LCSHs are applied to every item within the library's collection, and facilitate access to items in the catalogue that pertain to similar subject matter.

LCSH was first published as Subject Headings used in the Dictionary Catalogues of the *Library of Congress (LC)* between 1909 and 1914. Later on supplements were published followed by the second edition in 1919. The List is in its 33rd edition at present (Fig16.2), which was published in 2011.

LCSH system was originally designed as a controlled vocabulary for representing the subject and form of the books and serials in the *Library of Congress* collection. The main purpose of LCSH is to provide subject access points to the bibliographic records contained in the Library of Congress catalogues. LCSH is now used as a tool for subject indexing of library catalogues in general. It has also been used as a tool in a number of online bibliographic databases.
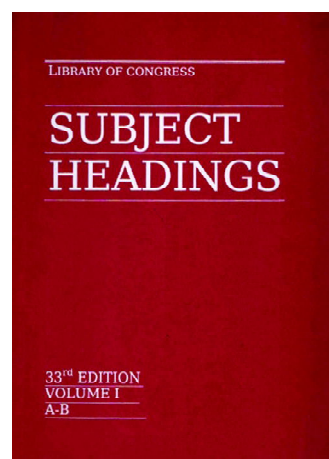

Fig 16.2: Snapshot of Library of Congress Subject Headings

LCSH has become a tool for subject indexing of library catalogues in general. There are four types of term relationships in LCSH.

i)   Equivalence relationships

ii) Hierarchical relationships

iii) Associative relationships, and

iv) General and Specific references.

LCSH provides useful reference guide for the user and also to the library reference staff. LCSH is the only subject headings list accepted as the worldwide standard. LCSH is the most comprehensive list of subject headings in print in the world and provides an alphabetical list of all subject headings, and cross-references.

(b) Sears List of Subject Headings (SLSH)

The *Sears List of Subject Headings (SLSH)* owes its name to its originator Minnie Earl Sears. The first edition in 1923 was published as List of Subject Headings based on *Library of Congress List of Subject Headings* (LCSH) designed for small and medium sized libraries. She edited the List till its 3rd edition in 1933.

SLSH is a thesaurus like database delivering a core list of headings, together with patterns and examples to guide the cataloguer in creating further headings as needed. Since the first edition in 1923, the *Sears List* has served the unique needs of small and medium sized libraries, suggesting headings appropriate for use in their catalogues and providing patterns and instructions for adding new headings as they are required. The successive editors of the List have faced the need to accommodate change while maintaining a sound continuity. The new and revised headings in each edition reflect developments in the material catalogued, in the use of the English language, and in cataloguing theory and practice. The aim is always to make library collection as easily available as possible to library users.

SCSH has got a new face since the 15th edition, published in 1994. Since then the editions have been coming quite regularly viz., 16th in 1997, 17th in 2000, 18th in 2004, 19th in 2007 and 16th in 2010 (Fig. 16.3). and the new face is due to the change in format that follows the NISO standards for thesauri. SCSH labels the references BT, NT, RT, SA, Use and UF for broader terms, narrower terms, related terms, see also, use and used for.

a) **NT (Narrower Term)** refers to more specific headings than the boldface heading

For Example : Birds

NT Pigeon

b) **BT (Broader)** refers to more general subject headings. These headings would be useful if you need to broaden your topic.

For Example : Applied Psychology

BT-Psychology

c) **RT (Related Term)**. These terms can provide ideas of other terms related to the topic.

For Example : Applied Psychology

RT-Educational Psychology

RT-Social Psychology

d) **USE and UF (Used For)** refer to preferred and non-preferred subject headings.

For Example : Chennai
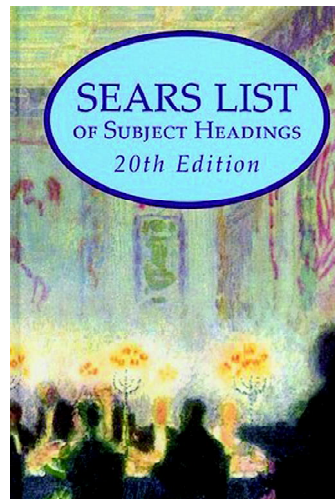
UF Madras

For Example : Madras

USE Chennai

Fig.16.3: Snapshot of Sears List of Subject Headings

16.5.1.2 Thesaurus

The dictionary meaning of the word 'thesaurus' is 'a collection of words put in groups together according to likenesses in their meaning rather than an

alphabetical list.' However, in library and information science parlance the word thesaurus means an authoritative list showing terms which may or may not be used in a catalogue or index to describe concepts. Technically, a 'thesaurus' is defined as 'a compilation of words and phrases showing synonymous, hierarchical, and other relationships and dependencies, the function of which is to provide a standard vocabulary for information storage and retrieval systems'. In a thesaurus, each term is usually given together with terms which are related to it in a number of ways.

Information retrieval thesauri are based on three types of relationship among concepts, viz. equivalence, hierarchical and associative relationships.

The thesauri basically serves as a controlled vocabulary used for indexing and searching of information in an information storage and retrieval system. There are several examples of thesauri. The *Engineering Joint Council (EJC) Thesaurus* and the *Thesaurus of Engineering and Scientific Terms (TEST)* are among the thesauri used for scientific and engineering subjects. In the social sciences, the *Educational Resources Information Center (ERIC) Thesaurus* is extensively used. Few others subject thesauri are *INIS Thesaurus, INSPEC Thesaurus* and *AGROVOC Thesaurus.*

INTEXT QUESTIONS 16.2

1. What is subject cataloguing.

2. List the types of Indexing languages.

3. What do you understand by Controlled indexing languages?

4. How many types of term relationships are present in 'Library of Congress Subject Headings'?

5. Who is the originator of 'Sears List of Subject Headings'?

6. Define a 'thesaurus'.

16.6 TYPES OF INDEXES

Indexing is broadly of two kinds, namely, derived and assigned indexing. Both are different so far as the representation of contents of documents is concerned. In derived indexing, index terms are derived directly from the document. No attempt is made to use an indexer's own knowledge of the subject or other guides. On the other hand, in assigned indexing, an indexing language is used for both input and query formulation. In the former, the index terms are often

derived from the title or the text, whereas in the latter, terms may be assigned to indicate a subject. Types of indexes are explained below.

16.6.1 Book Index

Book indexes are usually found at the back of the book. They list, in alphabetical order, specific referenced materials which is included in the book, giving page numbers or other information to assist readers in locating the material. Many book indexes are predominantly lists of subject terms.

16.6.2 Author Index

The index which has entries under authors' names in alphabetical sequence is called author index. It is the index where the works of a number of authors are listed in one index. An author index provides access to a document via authors' names.

16.6.3 Title Index

In the index which has entries under titles in alphabetical sequence is called title index. It is also an alphabetically arranged list of the titles of the works covered in a serial or non-serial publication. Title and author indexes are sometimes combined.

16.6.4 Subject Index

A subject index has alphabetical terms or words as headings. These terms represent concepts or subjects. Entries are arranged in alphabetical order according to the letters in the heading.

INTEXT QUESTION 16.3

1.   What is an assigned index?

2.   Differentiate between an author and a subject index.

WHAT YOU HAVE LEARNT

* Information retrieval tools are basic building blocks for a system that will organize recorded information that is collected by libraries, archives, museums, etc.

- The basic information retrieval tools include: bibliographies, catalogues, indexes, finding aids, registers, online databases, etc.

- Library catalogue is a list of books and other reading materials available in a particular library.

- The types of catalogues commonly found in libraries are: author, title, author/title, and subject catalogues.

- There are two types of catalogues, namely, classified catalogue and dictionary catalogue.

- An index is a filter or pointer, or indicator, or more fully, a systematic guide to the items contained in, or concepts derived from a collection.

- Subject indexing is a classification process.

- Subject headings lists are useful to understand the relationship among concepts to a certain degree, besides their application in indexing. Subject headings are provided in the catalogue entries to provide subject access to information.

- Indexing language is defined as a list of terms or notations that might be used as access points in an index. An indexing language may also be referred to as a retrieval language.

- Indexing languages may be of three distinct types: (i) Controlled indexing language, (ii) Natural indexing language, and (iii) Free indexing language.

- Controlled indexing languages are indexing languages in which both the terms that are used to represent subjects, and the process whereby terms are assigned to a particular document, are controlled by an indexer.

- There are two types of controlled indexing languages: alphabetical indexing languages and classification schemes. Classification is also prominent in the physical arrangement of documents.

- The Library of Congress Subject Headings (LCSH) comprises a thesaurus of subject headings, maintained by the United States Library of Congress for use in bibliographic records.

- The Sears List of Subject Headings is a thesaurus-like database delivering a core list of headings.

- Thesaurus means an authoritative list showing terms which may and sometimes may not be used in a catalogue or index to describe concepts. In a thesaurus, each term is usually given together with terms; which are related to it in a number of ways.

Notes

- Different types of indexes are book index, author index, title index and subject index.

TERMINAL QUESTIONS

1. Mention different tools of information retrieval?

2. What is the difference between 'Classified catalogue' and 'Dictionary catalogue'?

3. What do you understand by the concepts 'alphabetical indexing languages' and 'classification schemes?

4. Explain in brief about LCSH and SLSH.

5. Briefly explain different types of index.

ANSWER TO INTEXT QUESTIONS

16.1

1. Library catalogue is a list of books and other reading materials available in a particular library.

2. There are four types of catalogues commonly found in libraries, such as: author, title, author/title, and subject catalogues.

3. Index is an alphabetical list of terms, usually at the end of a book along with the page number where the term appears in the book.

4. Subject headings lists are useful to understand the relationship among concepts to a certain degree, besides their application in indexing.

5. The *Library of Congress Subject Headings (LCSH) List* is one of the best tools for indexing and retrieval. The *Sears List of Subject Headings (SLSH)* is a shorter version useful for smaller libraries. *Medical Subject Headings (MESH)* and *Subject Headings in Engineering (SHE)* are some of the other specialized ones highly used.

16.2

1. Subject cataloguing deals with what a book or other library material is about. The assignment of subject headings is generally referred to subject indexing or subject cataloguing. Purpose of subject cataloguing is to list

under one uniform word or phrase, all the material on a given topic that a library has in its collection

2. Indexing languages are of three distinct types: (i) Controlled indexing language, (ii) Natural indexing language, and (iii) Free indexing languages.

3. Controlled indexing languages are indexing languages in which both the terms that are used to represent subjects, and the process whereby terms are assigned to a particular document, are controlled or executed by an indexer.

4. There are four type of term relationships in LCSH, (i) Equivalence relationships; (ii) Hierarchical relationships; (iii) Associative relationships; and (iv) General and Specific references.

5. Minnie Earl Sears

6. Thesaurus is a compilation of words and phrases showing synonymous, hierarchical, and other relationships and dependencies, the function of which is to provide a standard vocabulary for information storage and retrieval systems.

16.3

1. In an assigned index the index terms are assigned by an indexer by using a guide or based on his own knowledge of the subject.

2. An index which has entries under authors' names in alphabetical sequence is called author index. A subject index has alphabetical terms or words as headings. These terms represent concepts or subjects.

GLOSSARY

Assigned indexing   : The system of indexing in which terms are assigned by the indexer on the basis of the conceptual relations existing between them.

Classaurus   : An elementary category-based faceted systematic scheme of hierarchical classification in the verbal plane having all the necessary and sufficient features of a conventional alphabetical, thesaurus excluding the enumeration of the Related terms (RT).

Derived indexing   : The system of indexing in which terms are derived from the terms used in the text.

		Information Retrieval Tools: Catalogues, Indexes, Subject Heading Lists
Index	:	A tool that exhibits the analyzed contents of a bibliographic entity or group of such entries, arranged in a systematic way, so that retrieval of specific items of information could be done easily.
Indexing	:	Indexing is the process whereby indexes and associated tools for the organization of knowledge are generated. Indexing can be done both manually as well as with the help of a computer. Subject indexing comprises three stages: familiarization with subject, analysis and assigning of terms to represent concepts using an indexing language.
Indexing Language	:	It is a set of items (vocabulary) and devices for handling the relationships between them in a system for providing index descriptions. Indexing language is also referred to as a retrieval language.
Subject Headings List	:	A list of subject headings or terms, including references, to be used as a standard cataloguing or indexing.
Thesaurofacet	:	Thesaurus which has a classified and alphabetical part.
Thesaurus	:	It is a compilation of words and phrases showing synonyms, hierarchical and other relationships and dependencies, the function of which is to provide a standard vocabulary of information storage and retrieval systems.
Vocabulary Control	:	A controlled set of terms used to represent the subjects of documents in indexing, and to search for these documents in a particular system.

WEBSITES

http://en.wikipedia.org/wiki/Controlled_vocabulary

http://www.slideshare.net/ImeAmorMortel/indexing-10954481

http://www.loc.gov/aba/cataloging/subject/weeklylists/

http://www.ebscohost.com/public/sears-list-of-subject-headings