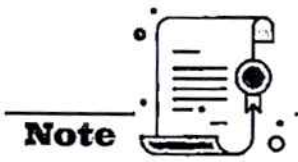


Note

7. **Methods of Compilation & Revision:** Following are the methods of compilation and revision.
   (i) The information should be collected and compiled personally by visiting the included libraries.
   (ii) The lists of included libraries should be taken and assembled.
   (iii) The largest library among the included libraries should circulate the list of its collections to other included libraries where each library can make use of it.

3.4 Layout and Rules for the Union Catalogues of Books

In order to try to bring uniformity in the cataloging practice, it is necessary that in the list codes of different levels, the normative principles at the international, national, regional and local level, the rules related to the writing style and the general opinion should be followed in relation to the layout of the library catalogue code.

This layout is based on the Practice for Layout of Libraries catalog by the Indian Standards Institute. The indexing and abstracting magazine is being laid here:

3.4.1 Parts of the Catalogue Code

A catalogue code consists of the following parts, which are as follows:

1. Preliminary material
2. Accessories
3. General Entry
4. Specific Entry
5. Lists other than library lists
6. Documents other than traditional books and periodicals.

1. **Preliminary Material:** Information regarding Guiding Principles, terminology and personal name should be given under the introductory material for indexing and abstracting journals.

   Here the sub-formula rules and principles come under the directive principle, which are as follows:

- Canons of Cataloguing:
  1. Canons of Ascertainability
  2. Canon of Prepotence
  3. Canon of Individualization
  4. Canon of Sought Heading

Note

5. Canon of Context
6. Canon of Currency
7. Canon of Permanence
8. Conon of consistence
9. Canon of Recall Value

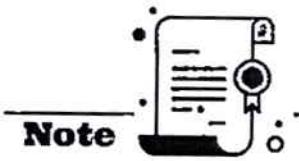
- General Normative Principle & Laws :
  1. Five Laws of Library Science
  2. Law of Interpretation
  3. Law of Impartiality
  4. Law of Symmetry
  5. Law of Parsimony
  6. Principle of Local Variation
  7. Principle of Osmosis.

Similarly, under the terminology, the definition of technical words and the artistic side of the words used are given and the structure and functions of simple names and compound names are kept under personal names and their information is given as follows:

Entry elements, Secondary elements and Indivalation elements.

Example: RANGANATHAN (SR) (1892-1972).

2. **Auxiliaries :** It includes the following elements:
   - (i) **Determination of Authorship :** Under this, rules should be kept to clarify the difficulties faced in the problems of determination.
   - (ii) **Rendering :** Under the hypothesis of various names, the rules related to the names of individuals, population rules, books and books, and the rules related to the hypothesis of artificial names should be given.
   - (ii) **Recording :** For the listing of entries under Recording, rules related to the language and script related system, writing style, printing, abbreviation and the aspects of arrangement or arrangement of entries should also be given.
   - (iii) **General Entry :** General entry includes general subject entry and cross reference entry and the list of subject headings or the use of chain process should be given in detail.
   - (iv) **Specific Entry :** Under the special entries, entries of a volume,



general book, multi-section book and indexing and abstracting journals are included and in this the main entry the rules for CIE and BIE entries should also be given.

(v) **List of Documents other than Library Catalogue :** Apart from the library list, rules should be given in these catalogue codes for the creation of other types of lists. In this way, the rules for the preparation, arrangement and layout of joint lists of books and periodicals, national lists should be given.

(vi) **Documents other than traditional books and periodicals :** In addition to traditional books and periodicals, subjects should also be given in the catalogue code for other specific types of text materials. It includes the following study material.

- Handwritten text
- Reprograp
- Tape record
- Cinema reel
- Pictures
- Instrumental recording of accompanying documents and visual objects.

3.5 Periodicals

A periodical is defined as "a publication issued in successive points usually at regular intervals and as a rule intended to be continued indefinitely" by A.L.A. Glossary of Library Terms. On the other hand, Ranganathan defines a periodical publication as a Conventional Document of kind I with the following attributes :

1. A volume, or a small group of volumes of it, is intended to be published or completed normally once in a year, though irregularity in interval is not reeded out.
2. Each successive volume or periodical group of volumes, is usually distinguished by the year of publication and/or by a number belonging to system of simple or complex ordinal number. Such a number is usually called the volume number.
3. The intention may be to continue the publication of the volumes for ever, though not actually carried out.
4. They intention may be to continue the same title in all the volumes, though this may not actually be carried out.

Note

3.5.1 Indexing

Presently, as a result of the discovery of the latest technologies in paper making and printing art, the publication of text materials has started increasing at a rapid pace. Nowadays, text materials are being published in many languages in the world every day, as a result of which readers are taking interest in the text material related to their specific subject instead of taking interest in any one specific book or writing. People engaged in research have also started turning away from books like articles published in current journals. If we look at the history of the library, in the past, on demand by the readers in the library, overall marks of the text materials were given, but due to this the readers had to waste a lot of time in searching for their useful articles. Therefore, in order to solve this problem at present, the articles published in the periodical magazines are now listed in alphabetical order according to the subject. Due to which it has been facilitated in the selection of useful topics without wasting our time in turning the pages of the magazine. The method of listing published articles in this way is called indexing and list is called Index.

3.5.1.1 Definition

According to the Oxford English Dictionary – *"An alphabetical list, placed (usually) at the end of a book of names, subjects etc. occuring in the with indication of the places in which they occur".* —Oxford English Dictionary

NS. According to C. Vickery – *"An index is a working tool designed to help the user to find way about the mass of documented information in a given subject field"* —B. C.Vickery

3.5.1.2 Functions of Index

The index plays an important role in making the user familiar with its text material due to the influx of literature, some of its functions are as follows:

1. It works to overcome the problem of language barrier due to special headings given in the user's language.
2. Provides information about the progress / development of subjects happening in all fields globally.
3. Through the index, the literature published in any field is collected and kept in one place.
4. An information acts as a source or tool between the source and the user.
5. Text material is easily available in this, which is very important for a searcher.

Note

6. It contains complete details of the text material up to the date of publication of the index.
7. It is a process through which it provides service to the readers.

3.5.1.3 Indexing Language

Method of obtaining documents on the basis of specific headings and submission of entries under subject headings has been prevalent among catalogers and indexers during the time of the C.A. Cutter. In presenting specific subject indexes, many difficulties have to be faced due to multifaceted development of subjects, inter-disciplinary research and specialization and abundance of fine documents. Therefore, more importance is being given to the need of coordinating or combining many terms in indexing. Coordination of index terms is considered an important aspect of indexing language. In the language of indexing, coordination actually means coordination of concepts.

Indexing language is the method of naming a subject for indexing. Like any language, it also has two parts. Vocabulary and syntax is a list of vocabulary terms that are included in the method. Syntax index shows the relationship between terms in a phrase.

There are two types of indexing languages. 1. Derived term form and 2. Assigned - term. In the derived term system all the terms are obtained from the document itself, whereas in the assigned -word system, the indexer creates the index term. This method is better by which the correct and proper subject title is assigned after getting the appropriate specific subject of the document. Thus author indexes, text indexes, citation indexes and indexes in natural language are derived term systems, whereas all indexing languages in which terminology is controlled, such as subject title lists, theories and classification systems, are assigned term systems. Derived-word systems are almost clerical and can be mechanized very easily. On the other hand, the assigned term systems introduced are intellectual, so they take more time and money. The structure in indexing generates and makes the index unusable

1. **Natural Language :** Natural language has its own importance. Its terminology can be updated and new concepts can be incorporated. It has its own syntax and rules of grammar. Which are able to show the true meaning of the specific subject. But natural language suffers from problems due to its antonyms and synonyms.
2. **Artificial language :** Themes are representative of title classification methods and theoretic indexing languages. Due to the controlled vocabulary, they are well organized and because there is a provision

Note

of reference entries see and see more in these, there is no problem of synonyms and antonyms in them. Nowadays, several components are needed to represent the specific themes of micro-documents. So these components are sorted according to the syntax rules of the indexing method. For example, in systems such as chain process, precision popsy, etc., the subject of the subject document is formulated on the basis of their own syntax rules.

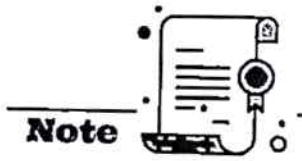
When a topic is formatted in a selected index language, it provides only one access to the index file. In order to provide access to other constituent terms, the index language provides for mechanization of permutation of constituent terms. In permutation, each constituent term is used one after the other as a leading term. Since these forward terms are related to other terms, so in each condition the forward terms are also added with the referenced terms.

3.5.1.4 Features of Indexing Language

1. Such language is developed and used for a specific purpose. The exchange of ideas, along with being a means of service purpose, has to fulfill some specific purposes as well.
2. Generally the word used in indexing language has a very precise and clear meaning.
3. Words having the same meaning and same pronunciation are removed.
4. As far as possible, an effort should be made to establish a unitary relationship between a language concept and terms. When terms have that kind of relationship, those terms can serve as a better directory of information in an index file.
5. Although it is not possible that we should completely eliminate words having the same meaning. The indexing language controls the occurrence of scattering or explosions in the indexing subjects when a term is selected to represent a particular concept from among the many words of similar meaning available in a natural language.

3.5.1.5 Recall and Precision Devices in Indexing Languages

Due to the availability of literature in abundance, an ordinary researcher has to face great difficulty in finding information. Many times a situation also arises that the user does not even understand what the related information can be about him and where it can be available.



Indexing language is the only medium that tells the users where the information related to a particular topic is located. In what form? Keeping all these important facts in mind, the importance of indexing language has been understood.

The following are the advantages of creating or using an indexing language:

1. By this it is possible to list the literature scattered in books and magazines on a specified subject at one place and it is very useful for researchers.
2. Indexing language serves as a communication between the source of information and the recipient of information.
3. It is estimated that progress is being made in various fields on the world level.
4. The topic title is in the language of the users, so the language barrier is also removed.
5. Previously published literature is easily surveyed.
6. Saves time.

3.5.1.6 Recall and Precision devices in Indexing Language

The effectiveness and quality of any indexing system depends on a few factors. Two important variables known as 'Recall and Precision' play an important role in measuring the effectiveness of information retrieval.

They are also related to each other. If a user wants 'High Recall' then he accepts 'Low Precision' and similarly the user who wants 'High Precision' will have to get 'Low Recall'. In this way they are highly related to each other. That is to say that the recovery effectiveness in the indexing system can be measured by 'Recall Ratio' and 'Precision Ratio'. With respect to 'Query' we can say that not all related documents can be retrieved but only some parts of them are retrieved. Generally not all retrieved documents are related. Similarly, the same situation prevails for unrelated documents. This situation can be understood as follows:

Documents	Retrieved	Non-Retrieved	Total
Relevant	a	b	a + b
Non-relevant	a + c	b + d	a + b + c + d

In this way, based on the above discussion, the recall ratio and precision ratio can be obtained.

1. **Recall Ratio :** The retrieval of related documents by a system is known as Recall. The number of documents retrieved from all the related documents through the indexing system is called Recall

Note

Ratio. In mathematical interpretation it is described as a/a+b. Let's calculate the Recall Ratio in the following way:

$$\frac{\text{Number of relevant items retrieved}}{\text{Total number of relevant items}} \times 100\%$$

or

$$\frac{a}{a+b} \times 100\%$$

For example, we assume that the total number of related documents in a file is 100 and the index is able to retrieve only 75 documents out of them and the remaining 25 documents are not able to be retrieved, then we can get its Recall like this

$$75/75 + 25 \times 100 = 1\%$$

or

$$75/100 \times 100/175\%$$

2. **Precision Ratio :** The ability to restrict unrelated documents in the system is called Precision Ratio. In mathematical form, finding out the number of documents retrieved out of all the related documents out of all the related documents i.e. out of the total number of all related documents is called Precision Ratio. In mathematical interpretation it is described as a/a + c.

Let us calculate the Precision Ratio in this way.

$$\frac{\text{Number of relevant items retrieved}}{\text{Total number of relevant items}} \times 100\%$$

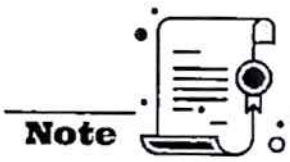
or

$$\frac{a}{a+c} \times 100\%$$

For example, we assume that the number of documents retrieved in a system is 150, out of which 75 are related, then its Precision Ratio will be as follows.

$$75/100 \times 100\,/1\% \quad \text{or} \quad 50\%$$

But it is very difficult to find the total number of all the documents from the collection. For this reason it is very difficult to calculate the recall value in an indexing system. For this reason Recall and Precision Value are very important in any indexing method. In this way we can say what should be found in the indexing system for the effectiveness of retrieval and what documents should be restricted? Recall and Precision prove to be very helpful in telling this.



Books for any reader who comes to the library to get their information, the material related to it is always available in the collection of library. Precision Ratio plays an important role in making the desired information available to the reader in a broad way.

3. **The Recall :** Precision Curve: In this we can divide the classes of four types of

   (i) A U B (ii) A - B (iii) B - A (iv) (A∩B)'

   All these are in the form of a matrix.

	Retrieved	Non-retrieved	Total
Relevant	AB	A – B	A
Non-relevant	B – A / B	(A ∩ B)'/B	A/L

On the basis of the above format, we can say that some writers understand this format better than other formats.

When we try to develop Recall by putting this form into practice then we get that the importance of Precision also goes on increasing and so on. If we unfortunately describe group A to a very small extent. It is probably used only under experimental conditions.

In this way we see that the quality and success of any indexing system has an important role. They represent the average performance of any indexing system and not the totality and their results also depend on the suitability of the conditions.

3.6 Major Indexing Services

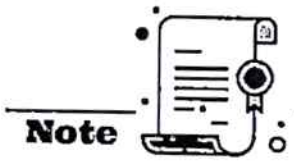
1. **Biological and Agricultural Index :** This is a monthly indexing journal, which has been published since 1964 by the New York publishing body H.W. Wilson Company. This is a proper subject index published in the English language. In which about 190 periodicals are indexed. The compositions of biology, agricultural science and allied sciences are indexed in the same chronological order by author and subject. The index of book reviews is given in separate sections according to the authors. Each entry is complete from the point of view of bibliographic information.
2. **British Technology Index :** Now published as 'Current Technology Index'. It has been organized and published by the Library Association, London since 1962. In a year, 11 issues and an annual volume are published. This magazine is a great tool for exploration and recovery of specialized and innovative information available in the fields of business knowledge, engineering and manufacturing technology.

Note

Published in around 300 British magazines, it helps in establishing contact with technical creations. Its indexing methodology is based on the Word Index and Trade Name Indexing (CATNI) method which is accessible separately.

3. **Index Medics :** Index Medicus has been compiled and published by the National Library of Medicine, Washington since 1960. It indexes 150,000 works annually. It contains more than 2000 medical journals. An index is prepared under author and subject. Monthly scores are published as cumulative volumes under the heading 'Cumulated Index Medicus'. Reports published in foreign languages are translated into English and the respective language is indicated in parentheses. Subject headings are given for each entry as required.
4. **Chemical Titles :** This fortnightly index journal has been published by the American Society since 1960. In this, articles of about 700 periodical journals of chemistry are indexed. Each article, author and keynote is indexed using the Indexing Method (KWIC). Its main objective is to keep chemists, researchers and chemical engineers fully aware of the specific and current information related to their subject area.
5. **Index India :** This is a quarterly documentation index which was started by the University of Rajasthan in 1967. It contains material from about 1000 selected publications. Efforts are being made to include in it the published literature related to India in the field of social sciences and humanities. It does not include legal reports and material related to the firm and the bank.

   In its main categorical part the entries are arranged in chronological-categorical order, while in the second part the three indexes alphabetically, author and review indexes are arranged alphabetically letter by letter. A list of current periodicals included is given at the beginning of each issue.
6. **Library Literature :** This bi-monthly index journal is being published by the H. W. Wilson Company of New York since 1934. Its annual volumes and bi-annual cumulative volumes are also published, about 225 journals published from different parts of the world available in the field of library and information science are included in it. In addition, on a selective basis, journals related to other subjects included in Wilson's other indexes are also included in it. Indexing of booklets, films, film plates, microcards, microfilms, dissertations, research articles, etc., related to library and information



science.is also done. Microcard and microfilm related material that has been published in print is omitted. In this, the author, title and subject entries are arranged chronologically in the same sequence.

7. **Citation Indexing :** The basis of this new method lies in the failure of classification systems. who have been unable to cope with the abundant emergence of knowledge. In which subject the original research works related to these indexes are cited from the author and the report. All contributions are grouped under each citation. The cited works of each author are placed in chronological order. The credit for indexing the citation goes to Eugene Garfield of the Philadelphia-based Institute for Scientific Information. This organization has been publishing 'Science Citation Index' since 1963 and 'Social Sciences Citation Index' since 1973'. Now the publication of the Humanities Citation Index has also started.

The following entry points have been provided to search for content in the Citation Index

1. Population Index Division
2. Quote index
3. Source index, and
4. Permutation Words, Subject Index -

All those items which become its specialty under the name of an organization and all those publications which are published in the said period are included in the population index.

The citation authors are arranged in chronological order in the citation index and the year cited is arranged chronologically.

The source index is an alphabetically arranged author index of 5,40,000 articles and editorial material, each entry containing the author/s name, language sign, article count,

Name of the magazine, volume number, issue number, number of pages, year of publication, number of references and Full address of the first author. This index is helpful in the use of citation and permutation word subject index.

Permutation word subject index is created by computer. The important words of the story of each article are changed in order and added to each other word of the story with which it is included in the belief of any article. The exploratory technique is used to follow the instructions quoted in the related works.

3.7 National Bibliographies

The term 'bibliographic record' is a relatively new term. The use of this term is associated with library computerization. In general terms, a bibliographic record is the sum of the elements and fields used for the description, identification or retrieval of the physical fields (such as publications, documents, etc.) of the subject matter of any information.

In the words of Gradle and Hopkinson, "A collection of data elements that are logically arranged and represent a bibliographic subject is called a bibliographic record." Gradle and Hopkins have defined the textual subject as:

"Any document of human communication, book publication or any other transcript or group of documents or part of any document etc. may be taken as an entry.

Any bibliographic record format that is valid for data exchange must have the following three basic components:

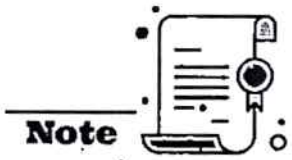
1. **Physical Structure :** Rules for the adjustment of data on the computer storage medium.
2. **Content Designators :** Codes to identify the various data elements of a record (i.e. author, title, date of opening of journals, etc.)
3. **Content :** The content of any record that is bound by the rules for the creation of various data elements, they are associated with the content designator. Data components from any exchange format are identified by a code that is not only in the form of content, but that the record is suitable for use by any other agency.

Exchange of any bibliographic data between different agencies is possible only if their data set includes the appropriate three components. In any information system the records that make up the database are mainly available in various formats. Some of the main formats are

1. A format to be incorporated into the record system.
2. Such a form which is suitable for storage for a long period.
3. A format that helps in easy information retrieval.
4. The form in which the record may be displayed.

Along with this, it is also necessary that if two or more organizations want to exchange data among themselves, then they should use a common standard. Using a uniform standard is a must for any networked world. Every organization should be able to convert its records to a standard that allows for smooth and smooth exchange of data. Although at present there

Note



are many national and international standard exchange formats available and there is a lot of uniformity in them, due to their not being uniform, data exchange from one format to another is not easy. Keeping this fact in mind, ISBD was developed by IFLA.

3.8 ISBD-International Standard for Bibliographic Description

IFLA held a meeting of the ICCP in 1967 to standardize bibliographic description and as a result step towards the preparation of a standard for various bibliographic descriptions. In this sequence, different I.S.B.D was developed. In 1974 the first I.S. B.D was published which was related to the monograph. After this other I.S. B D. was formed as follows. These are discussed further.

Type Publication	Year	Detail
ISBD (a)	1980, 1991 (Second Edition)	For publication of old monographs (Anti Quarian)
ISBD (cf)	1990	
ISBD (cm)	1997	For details about the map/atlas
	1987 (Revised Edition) (Cartographic)	
ISBD (CR)	2002	For serial and other serial resources
ISBD (ER)	1997	For electronic resources

3.9 Abstract: Definition and Concept

When the details given in a document are presented in a concise form, then that brief description is called Abstract. Through this summary, information is provided to the reader about what has been highlighted in the related document. In this way the process of preparing a summary of documents, articles, etc. is called abstracting.

Various definitions of essence have been given by various scientists, some of the major definitions are as follows:

According to Kallison, "A summary is a summary of all the important things that are given in the original document according to the order of the information given in an original primary article. The author presents all those points in a systematic and concise form in his own language.'

According to Dr. Ranganathan, a summary is a summary of the essentials of the subject described in the article of a general journal, which is done by

Note

a professional other than the author and the details of the availability of documents are indicated in it.

According to Mazel, "The abstract presents the purpose, scope, and conclusion of the document. It is mainly produced for users who do not have access to the original document.

According to Ashworth, "A summary is a collection of information." It is an abbreviation with which the context of the search is included. Often nowadays this short summary form is indicative of information contained in an article in a magazine.

In the context of these definitions, it can be said that a brief statement of the characteristics of the essential and important subject matter of an original document, articles of a particular type of magazine is called a summary. It provides information to the user to determine whether the said document can meet their requirement. In the summary, the related document or the bibliographical details of the publication are also mentioned which helps in getting the document. Therefore, in a nutshell, it can be said that the summary is that it is a concise representation of the content of a document in a style similar to that document.

3.9.1 Abstract: Scope And Importance

The details of the scope and importance of the abstract can be presented as follows:

1. Literature is published in more than 80 languages of the subject. Normally the user is fluent in two or three languages. Therefore, he cannot use documents that have been published in other languages. All readers can use the document to summarize such documents from the respective language.
2. The reader's time can be saved by the abstraction service.
3. The efficiency of the indexing service can be increased by this service.
4. It plays an important role in literary search.
5. Casual Consciousness Service (CAS) can be expanded.
6. Repetition of research and research work can be prevented.

3.9.2 Types Of Abstract

A concise, accurate and succinct presentation of any document is called a Abstract. Abstracts are of many types due to their purposes and forms. These can be described as follows:

1. Indicative Abstract
2. Informative Abstract

Note

3. Informative-indicative Abstract
4. Critical Abstract
5. Slauted Abstract
6. Auto Abstract
7. Telegraphic Abstract
8. Graphic Abstract

1. **Indicative Abstract :** In this type of abstract, the content of the document is presented briefly. It mentions the area of the document. These abstracts do not provide much information about the document but point to them. Such abstract documents are helpful for the purpose of selection, but they cannot be used in place of the original document. These abstracts are considered very useful for providing Current Information Services (CAS) i.e. indicative abstracts are important in guiding readers.
2. **Informative Abstract :** Under this abstract, quantitative and informative information of the document is provided. This abstract can also be used in place of the original document. That is why it is also called Abstract Synopsis. The idea of the original document, principles, methods used, results and techniques are presented in this way in an informative summary. This is done so that the user does not even need the original document. An average of 200 to 300 words are used in this abstract.
3. **Informative :** Indicative Abstract: This type of abstract includes the characteristics of both informative and indicative abstracts.
4. **Critical Abstract :** In the critical abstract, the original document is presented in a critical way. Through this, not only the summary of the subject given in the document is presented, but the evaluation of that work is also done. The depth of the work is clarified by this summary.
5. **Slauted Abstract :** In the present era, some documents are useful and informative for experts and scientists of different fields due to the development and impact of interdisciplinary subjects.

   In this summary, more importance is given to any one part of the document.
6. **Auto Abstract :** This abstract is created with the help of a computer. Under the automatic summary, the important words that come in the document are selected through the computer and when these

Note

words come again and again, the sentences are sequenced and the summary is made.

7. **Telegraphic Abstract :** In this abstract, the key words are selected based on the original document. In this abstract these words are not arranged in any definite order.
8. **Graphic Abstract :** Graphic abstract is used for communication in chemistry. These abstracts are created in the language of Molecular and Structure Formula.

3.9.3 Abstracting Procedure

After selecting the documents for preparation of abstract, the following procedure is followed while summarizing them:

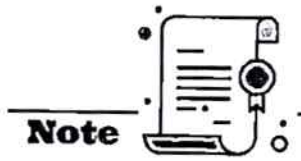
1. After this, a final copy of this rough abstract is made. For this, an effort should be made to end the abstract in an isolated paragraph.
2. In this abstract , it is checked to rectify the mistakes, in which the correct use of words like spelling, comma, punctuation mark etc. is done under the rules.
3. A rough abstract is prepared from the selected key words.
4. After selecting the key words of the document, they are noted down.
5. First of all, the subject-matter, area, conclusion etc. mentioned in the abstract document are carefully studied. For this work, the introduction, conclusion, suggestions, ideas etc. given in the document prove to be helpful.

3.9.4 Canons of Abstracting

Dr. S. R. Like the principles of classification and cataloging by Ranganathan, the canons of abstracting have also been presented. All these canons give a theoretical form to the abstraction work. These canons are as follows -

1. Positive Canon,
2. Negative Canon

1. **Positive Canon :** What elements should be included while abstracting a document? They are selected by this canon.
   (i) The content described in the original article should be summarized.
   (ii) In the abstract, the basic and elementary progress in the field of knowledge should be clearly mentioned.



(iii) Secondary progress in the field of knowledge should also be mentioned.

(iv) The methods, technical findings, etc., described in the document should be included along with their characteristics.

(v) The important data given in the document should be briefly summarized.

2. **Negative Canon :** Which elements should not be included while abstracting. It is determined by the negative canon.

(i) Abbreviations of subjects should not be used in the abstract.

(ii) Information should not be provided which reveals the report of the document.

(iii) The examples mentioned in the document should not be elaborated.

(iv) The sentences of the document should not be used in the abstract, hence telegraphic language should be used.

3.9.5 Principles of Abstracting

To make an abstract answer and useful, the following principles should be kept in mind:

1. **Size :** The size of any summary makes it useful and important. Various scholars have rendered their views regarding its size. According to the summary,

(a) According to Borko and Vernier – 1 to 1,000 words

(b) Informational Abstract (as per DDC) – 150 to 200 words

(c) For Serials – 50 to 100 words

(d) Best Summary – 200 to 250 words

2. **Writing Style :** According to this principle, the abstract should be written in the simplest style. The ideas, concepts, activities etc. presented in the document should be written in the past, but the conclusion should be presented in the present.

3. **Word Selection :** The sequence of the words selected for the summary should be arranged. One should observe frugality in the use of words.

4. **Sentences :** The title of the document should not be repeated under the summary and whatever sentences are used should be completely clear and concise.

Note

5. **Language :** The language used in the abstract should be simple and include abbreviations, punctuation marks. And trade names should not be used. In addition, presented in the original document. The subject matter should be presented completely.

3.9.6 Qualities of a Good Abstract

The characteristic of an answer summary is to present maximum information about the subject material contained in the original document in the least amount of words. An abstract should have the following characteristics to save the reader significant time

1. Brevity,
2. Accuracy,
3. Simplicity,
4. Reliability
5. Authenticity
6. Clarity

1. **Brevity :** Conciseness is the main feature of the abstract. The abstract should be concise as compared to the original document but it is necessary that it should have full capacity to express the purpose, method, conclusion etc. of the original document. Adopting brevity in abstracts saves time for the reader and staff as well as space if the summary is concise. So the cost of presentation, printing and publication is reduced. The sizes of the bars vary. But in how many words it is to be presented, it depends on the document. A good summary should be 200 to 250 words.
2. **Accuracy :** While preparing the abstract, the abstractor should try that the essence should be impeccable, that is, the purity of the essence is very necessary so that any kind of confusion cannot arise. To avoid any errors in the abstract, first a rough summary of the original document should be made and this type of re-examination should be done. Most errors in the abstract are related to mathematical base material, proper names and punctuation marks, which must be verified.
3. **Simplicity :** The abstract made from the original document should be simple and the abbreviation and trade name should not be used in the abstract. Therefore, the language and writing style of the abstract should be simple, which can be easily understood by the users.
4. **Reliability :** The credibility of the essence is very important. Therefore, its rules should be followed in summarizing and presenting the summary. For this The abstractor should have thorough knowledge of all the techniques of abstracting.

Note

5. **Authenticity :** Authenticity is the main quality among all the characteristics of essence. The authenticity of the essence depends on the person doing the abstract. The conclusions presented in the abstract should be according to the authentic and original document of principles, techniques and methods, for this work the abstract should be expert and trained.
6. **Clarity :** The language, style, words etc. used in the abstract should be completely clear and understandable. The words selected while preparing the abstract should be taken from the original document itself. Because the words used by the abstractor can be ambiguous.

Therefore, the essence of an original document should be clear, concise, accurate, reliable and authentic so as to provide maximum information to the user.

Conclusion

In this chapter, we studied the contribution of scholars like Cutter, Kotsa and Ranganathan in cataloging and the concept of cataloging, to know their ideas, abstraction, indexing and bibliographic list. Also the layout of the Cataloguing of the books.

Important Terms

- Cutter introduced the concept of a specific subject in the 'Rules for Dictionary Catalouge' in 1876.
- Union List refers to the combined list of various libraries of a region.
- An indexing list is a list that is usually at the end of a book with respect to a name, subject, etc., in which the place to find them is indicated.
- Any document of human communication, book publication or any other set of transcripts etc. can be taken as one entry.
- Abstraction is a concise presentation of the content of an original document using the same style that was available in the original document.

Note

Exercises

Very Short Answer Type Questions

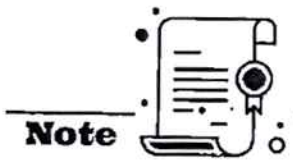
1. Write the meaning of indexing?
2. State the meaning of abstraction.
3. What do you understand by bibliographic record?

Short Answer Type Questions

1. Write a note on the layout of the Union Catalogue Code of Books.
2. Describe the contribution of Cutter in the field of Cataloging.
3. Describe the contribution of Ranganathan in the field of cataloguing.

Long Answer Type Questions

1. Defining indexing, describe its functions and characteristics.
2. Defining abstraction, describe different types of abstracts and briefly explain the process of abstraction.



UNIT 4

Online Cataloguing

4.1 Introduction

In this chapter, we will learn about the online indexing system OPAC and Web OPAC. You will also learn about the subject listing and what its purpose is and will also study in depth on this topic. We will study why LCSH is used in the library and what kind of facilities are available in running the library by using it.

4.2 OPAC

The short form of On-line Public Access Catalog is called OPAC. This catalog proves to be more user-friendly than the catalogs traditionally provided in libraries and helps the users in their information discovery process at every step.

Due to the explosion of information in the modern science era, various networks are providing cataloging work in the form of magnetic tapes in the spirit of cooperation. The OCLC system is providing the Library of Congress (LC) cataloging data via the Network. Users of OCLC. ask for bibliographical description of their desired information in magnetic tapes. For this reason, many libraries from the 1970s onwards moved rapidly towards computerized catalogs in place of their card catalogs and made their catalogues available in the form of Micro Fiche. In the modern era, most libraries are making their catalogs available through OPAC.

Most libraries in the era are making their catalogs available through OPAC. OPAC is a computer-generated catalog available in the form of magnetic tapes. This type of catalog can be used by any user through a computer. Thus the desired document is searched through All Line. The search for any document and information can be done by OPAC through the subject approach.

Note

Today more and more libraries and information centers are using OPAC. Because the electronic format of this catalog proves to be more user-friendly for the users than the card catalogue.

4.2.1 Merits of OPAC

Following are the advantages of using OPAC in libraries and information centers :

1. When the extension service (SDI) selected by OPAC is over, Smt. Reminders or requests for extension of date can be received through e-mail.
2. Through OPAC, the user can request for the acquisition of his desired book or document. Apart from this, reservation of desired documents can also be done online.
3. OPAC can also conveniently search for information through Boolean operators.
4. OPEC, being an on-line facility, provides the facility to search information through various indexes, such as publisher index, seminar location index, Quick and Quoc (KWIC/KWOC) etc
5. OPAC allows users the freedom to search for content from different approaches. Through this, the user provides the facility of searching documents and information by author, title, subject, publisher etc.
6. Through OPAC, the user himself can search for his desired material on-line and he does not need any reference staff for this work.
7. Most of the users are not aware of the content of the available content. In this regard, it proves to be User Friendly and also facilitates the user to select alternative material in addition to the desired material.
8. With the help of this, the searcher can be searched for his desired material from remote areas online at one place.

4.3 Web Opac

The concept of Web OPACs is the recent origin and it is serving as a gateway to the resources not only held by the respective library but also to the holdings of other participating libraries without to local collection but going, beyond further to regional, national, international levels. It allows users to interact with documents stored on computers all over the world and makes easier access to catalogue data in the form of bibliographic records. It becomes another search engine-referred as 'web cat' and as an 'information gateways'. It can support such as Telnet HTTP, FTP and Gopher. According to Online Dictionary for Library and Information Science defined as : "An

online Public Access Public catalogue that uses a World Wide Web, as opposed to a text-based interface accessible via Telnet."

4.3.1 Features of Web OPAC

The features of Web OPAC are :

1. It is accessible through the internet.
2. It is possible to search independently by Author, Keyword, Title or Year.
3. Displays complete bibliographic information as appeared on reprints.
4. Features of traditional OPACs such as storing bibliographic and sometimes full-text databases, providing direct access to a library's bibliographic database by means of terminal of PC, search result in readily understandable form, reference help etc.
5. It has the ability to use hypertext links to facilities navigation through bibliographic records.

4.4 OCLC

OCLC is the abbreviation for On-line Computer Library Center. It is the library and information network of America and the largest information network in the world. This network is an important center for resource sharing and information acquisition between libraries and information centers. This network was named "On-line Computer Library Centre" by "Ohio College Knowledge Organization" in 1971.

This information network was established in 1967 in Dublin, Ohio. This network was the most popular regional network at that time. On-line services were also started by this center in 1971. As a result its nature was also changed. In this regard, the Marc format proved to be a good and important step. At present, this network is a huge network, which is not only the network of information and libraries in America but has been established as the largest information network in the whole world.

4.4.1 Aims of OCLC

Following are the main objectives of the system:

1. To make available the database of documents in the form of magnetic tapes to the users.
2. Mechanized all the activities of the library.
3. To provide inter-library loan service for the purpose of information communication between libraries and information centers.
4. To support and promote research and development (Rand D) activities.

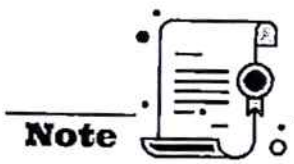
Note

5. To promote resource sharing among all libraries and information centres.
6. Automation of the process of acquisition and cataloging in libraries and information centers.
7. To make available to the users their desired information and documents on-line.
8. Establish an international level network of libraries and information centers.

4.4.2 Activities of OCLC

The main activities of OCLC are:

1. **Union Catalog :** The facility of on-line cooperative list is provided through this network; these union lists are made available by member libraries and information centers of newly Local Services -received documents. This information is very important and incomparable in relation to newly received documents in any library. The main feature of the program is to provide union lists . This feature saves the time of the users.
2. **Local Services :** It is not only an international network but also provides local services to on-line users. Many library units have been created through these services, in which centralized cataloging service, on-line and information service is being provided.
3. **MARC Tapes :** Through this center, services of all the library members are made available through marc tapes. This service can be done off-line sub. is being made available. This information can also be purchased. Material published before 1968 is not included in the mark tapes. The tapes contain about 6 million documents which are available in 468 languages. These can be searched online through Marc from anywhere in the world.
4. **Database Services :** The collection of this center is in the form of a database which is a collective record of all the member libraries, such as marc tapes of Library of Congress, National Medical Library, GPO . Monthly list etc. Through this network, databases of documents published up to and before 1968 can be obtained.
5. **Inter-Library Loan Service :** Inter-Library Loan Service (ILL) service is being provided by this network under Reference and Documentation Services. A sub-system of this system is providing Service on-line between member libraries and information centers within the state of New York.



6. **Printing of Catalogue Cards :** In 1972, this center merged the catalogs of all the member libraries and information centers and the Library of Congress (LC) together so that their titles could be identified and catalog cards can be created. For this task, a Catalogue Production System was started through this network. This system is being served through MARC Tapes and is being distributed to the Member Libraries and Information Centres.
7. **Full Cataloguing of Periodicals :** The entire cataloging process of periodicals is done by this center. The facility of inquiry, demand and union catalog is also provided by this network. Through this system, the facility is available in the list on-line.
8. **Acquisition Sub-system :** Through this network the facility of acquisition sub-system is provided to the user. This system is called SUBY / OCLC. Through this system, the concerned librarian can be ordered to acquire the desired documents. After this, this center receives and prints that order and sends the order to the concerned book publisher to send the book.

4.5 Subject Catalouging

This list was first published in 1923 by Minnie Earl Sears. This list was compiled by Sears to meet all the subject heading needs of medium-sized public libraries and schools. It is an abbreviation for LCSH. positions in this list.

The arrangement is a dictionary. The purpose of this list is that all the documents available in the library on a particular subject should be kept under one form of subject headings. Apart from this, in this list, after x, xx, for mutual references, those terms have been given from which the subject "See" and 'See also" can be formed.

The arrangement of words in this list is lexical. Words used are printed in thick and dark ink and synonyms/titles which cannot be used as titles have been printed in light ink.

The titles in this list are printed in a variety of letters:

1. Headings where the user is directed to select the appropriate title by the See Also (SA) directive.
2. See references as required from those headings before which x is marked
3. Those headings which are preceded by xx signs, see it as per the need (See also references) instructions are made. From the 15th edition, these symbols have been replaced by BT, RT, NT, SA and UF, which are used as related terms.

Note

4.5.1 Subject Headings Format

Subject headings in SLSH are designed in a variety of ways by expanding descriptive phrases containing two or more words into isolates word headings and isolates concepts of compound and compound subjects.

1. Single word heading
   Agriculture
   Religion

Individualizing element is given with headings to identify all words with different meanings. for example

Seals (Animals)

Seals (Law)

2. Compound heading
   Bow and Arrow
   God and Evil
3. Adjective phrase heading
   American Literature
   Civil Engineering
4. Prefix phrase heading
   Electricity in Agriculture
   Freedom of information
5. Complex subject
   Education-History
   Mathematics-Study and teaching
   United States-History-1945-1953

4.5.2 Entry Format

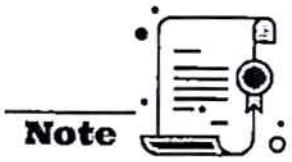
Sears List also uses different symbols with headings, as in 'LCSH'. Subject headings are printed in bold letters and referential terms in lowercase letters. From the fifteenth edition the 'see' and 'see also' are interchangeable references and the ' and 'xx' symbols have been replaced by UF/USE, BT, NT, RT and SA.

4.5.3 Adding Headings by Lister

In the Sears list, listers are allowed to add headings as needed, as follows:

(i) Names of persons

(ii) Family names



(iii) Names of places

(iv) Name of Nationality

(v) Name of the National Language and Literature

(vi) Names of wars and battles

(vii) Names of macro bodies

(viii) Common names such as names of animals, fruits, sports, diseases, parts of the body, chemicals and minerals.

4.5.4 Key Headings

The following main posts have been given as a guide for efficient performance by the lister. According to which alternative provision has been made for other main heads.

(i) Person: Shakesphere, William, 1564-1616

(ii) Names of families; Indian of North America

(iii) Place: United States, Ohio, Chicago -

(iv) Language and Literature: English literature

(v) War: World war 1939-1945

Cross references are also given which are very helpful in finding the most suitable topic title. The website of MeSH is

http://www.nlm.nih.gov/mesh

An updated printed version of this controlled glossary is published in January every year.

4.6 Library of Congress Subject Headings (LCSH)

The Library of Congress Content Title Catalogue is a comprehensive English-language lexical control method that is used by library readers to access lexical lists of college, university and major public libraries in the United States, Canada, and many other countries. This subject title list, which originated in the Library of Congress in 1897, has been continuously revised and updated to this day. It is currently in its twentieth edition. This table of contents is now available in print form, on microfilm and CD ROM as well. It can also be obtained on the Internet.

This list was created by the Library of Congress of America. This list is mainly used as a tool for indexing books, periodicals, and information materials acquired in the Library of Congress. Its main objectives are as follows:

(i) To prepare such a map of knowledge in different subjects in which

Note

we want to retrieve information. Both the parts of the software working as a co-operative do the application together. Clients in the absence of a server, or servers without a client, may not be able to do any useful work. The software tools for the client/server system always work as a couple and share the workload of computer related tasks.

The server program is responsible for holding the data that may be available and for receiving and returning the data when requested by the client. In other words, it is responsible for the creation of indexes, searches and sometimes for data collection and management. Most importantly, the server provides a means to allow universal access to the data it holds. It waits for the client software to send a request to perform some action and in response to those requests it transmits the replies received by its efforts to it. Thus, the client program is responsible for dealing with the users. It is used locally and acts as an interface between the user and the system, providing information as per the needs of the user. Collects them in a packet and transmits them to the relevant computer server, represented in an agreed language for communication between the client and the server. When the server, with some data | responds, the client then opens the coded content and converts it to be stored as a file for suitable display on the user's computer.

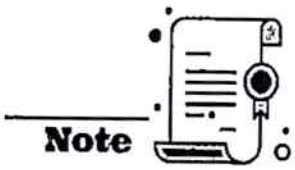
The most important utility of the client/server architecture is that there is no need for continuous communication between the client and the server. They can communicate one by one. There is no need to establish a connection between the client and the server. Using a client server architecture, it is easier for different systems to communicate. In other words, for each task the clients/servers use a common language to communicate with each other the tasks they want to do, as well as to communicate requests and send responses back and forth use iternet protocols.

4.6.7 Internet Connection

It is important to know that connecting to the Internet can be done in a variety of ways. These practices determine how we will be able to act on it. In the second, the manner in which a user is connected to the Internet greatly affects its usefulness.

TCP/I on a LAN server. P. Connection of local area network with internet can be achieved by installing networking software. The Internet can be accessed by a variety of computers connected to a LAN. This includes personal computers based on DOS and Windows, and Unix work stations, etc.

There are basically two modes of connection—full connection or terminal connection. In the case of full connection there is permanent



telecommunication connectivity and the computer concerned is given a registered Internet name and address. The second type of combination is known as a dial-up combination. In this, the related computer is connected with the original computer with full access to the Internet through a temporary telecommunications link. The third type of connection is called a gateway connection, in which the connection is provided by another network or service provider, such as Compu Serve.

Connections can also be established directly, as done by universities or community bodies, where a computer or network of computers is permanently connected to the Internet by dedicated lines and has its own address or by serial line Internet Protocol (Serial Line) Internet Protocol. Connected directly to a remote computer via (SLIP: Serial Line Internet Protocol) or Point-to-Point Protocol (PPP: Point to Point Protocol). The utility of full connectivity is that you can install any client software on your computer as per your requirement and use all the facilities of the Internet.

On the other hand, if your local computer is connected to the original computer with full connectivity, then in that case your computer only acts as a terminal for the original computer. You only have to log on to the original computer, after that you can access the Internet through the original computer. In this situation, you can take advantage of limited amount of Internet facilities depending on the type of communication between your computer and the original computer. It is a standard today for providing Serial Line Internet Protocol/Point-to-Point Protocol (SLIP/PPP) for commercial basic computers. But in those areas where telecommunication facilities are not sufficiently available and the Internet is still developing, the situation is not like this.

4.6.8 Resources and Services Available by Internet

It can be said that the Internet is a vast, searchable, dynamic, expansive, multi-platform information system filled with a wealth of potential resources. The resources available on the Internet are constantly changing. Therefore, any list without being updated can be undesirable. However, in this section, a brief overview of the types of databases and services is given so that you can become familiar with the resources and services of the Internet. There are different types of uses of the Internet. Some of these are given below :

E-mail : Allowing users to send messages to each other.

News : Informing users about available information :

Remote Login : To allow remote login users to login to remote locations;

FTP (FTP: File Transfer Protocol) : It allows users to access and retrieve files from remote locations.

Note

Services provided at remote locations includes :

1. List serves and discussion communities on a wide variety of topics. Participants get the opportunity to exchange current information and conduct conversations. Some list servers are also of special interest to information professionals;
2. Subjective databases of educational institutions specially specializing in various disciplines;
3. **Community Information :** Various communities are often providing access to information, such as library catalogs, demographic and tourism information, etc., through their public libraries;
4. **Government Resources :** All national and local governments are providing information through their websites;
5. **Library Catalogs :** Majority of the libraries are making their catalogs available through the Internet;
6. Commercial resource commercial information database;
7. **Bulletin Board :** This is the electronic equivalent of newspapers;
8. **Shopping and other commercial transactions :** Many important shopping centers on the Internet.
9. **Document Distribution :** Many large libraries and document distribution services are providing the services of document search and distribution through the Internet.

Many such services provide access to database records or documents.

On the basis of the following characteristics, these types of documents are divided into different categories or collections.

- Collection location
- Format
- Size or length
- Depth of Subjectivity
- Language
- Audience or consumer
- Data type
- Transfer Method
- Subject matter or topic
- Frequency and innovation of updates
- Issuer or exponent

It may be mentioned here that there are many directories of Internet resources available. One of the most valuable directories is the Bulletin Board for Libraries or BUBL. Bubbles is an information service designed to empower library and information science professionals. A wide range of services include directory of resources, user and contemporary content lists and texts of library and information science journals.

4.6.9 Search On Internet

A wide range of databases and services are available on the Internet. It is essential that interfaces are designed to help users find information resources and services available on the Internet, as the process of retrieval on the Internet is a complex and significant problem. This is because the databases created on the net are in a variety of different formats and various search retrieval software packages are installed on different computers to provide access to a subset of databases through different interfaces.

Devices used to search the Internet are often operated in Client / Server Mode - Server Software, which enables the user to search the database installed on multiple computers on the Internet using an intuition arrow be able to obtain. The user's local computer runs the equivalent client software that communicates with the server software and provides concurrent interfaces to the data. In other words, the user does not need to know where the data is stored or how the file storage structure of the server system is organized.

4.7 WWW.

(WWW or W3: World Wide Web) is the multimedia part of the Internet that displays a hypertext type of structure and search facilities. It was first introduced in 1989 at the European Particle Physics Laboratory (CERN) in Switzerland. European Particle Physics Laboratory was developed for document sharing between molecular physicists at the CERN (European Particle Physics Laboratory) laboratory, but the first commercial web software was developed in 1991, which Popularized this form of access to the Internet. The main features of the organization and structure of WWW are :

The set of rules that give tags and formats to documents is called Hyper Text Markup Language (HTML). These are used to organize documents into information blocks (pages).

Each individual document or page is assigned a unique address called a Uniform Resource Locator (URL) (URL: Uniform Resource Locator).

Each URL is connected to other types of hypertext. URL and information sections can also be combined with other URLs with each document.

These documents are searchable by interactive interface programs that allow users to view and navigate through documents. These programs are called web browsers.

Communication between web browsers and web servers is handled by a common language that contains a standardized set of rules called Hyper Text Transfer Protocol (HTTP). is used. H. T. T. P. H. T. M. L. Under

Note

each web page. symbols to be interpreted to enable proper page display and transfer of files. The client program or web browser provides controls for the user to enable the recovery process and connections.

It should be kept in mind that individuals and organizations present their information or services on the Internet by creating a home page. A collection of home pages located on a single server is called a website. These pages are accessed through the Uniform Resource Locator (URL) (URL: Uniform Resource Locator), using the browser. Some examples of browsers are Linux, Netscape, Explorer etc. These addresses link users to the original computer and their personal files. These files are then displayed on the user's terminal (workspace). with the help of appropriate software, users can read documents, view pictures, listen to sound and retrieve information.

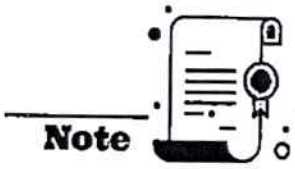
The hypertext structure of the web refers to the retrieval between different web pages through browsing and navigation. In index positions, W.W. W. But the hyperlinks that form the basis for browsing in the network are uncontrolled. Each hyperlink is individually assigned to the H.T.M.L. Coded by the creators of It may be mentioned here that although browsing is not a suitable approach to identify any specific information, the invention of various types of search tools has helped people to search the information from the vast collection of documents in the world done for. These search tools can be divided into two main categories - topic directories and search engines.

4.7.1 Subject Directories

Subject directories are also known as subject guides. Through these, the information of the people under the subject is allowed to be observed. These are hierarchically arranged indexes of different topics which are linked with different websites on each topic. The searcher can browse through the index in search of related topics and click on the specified places representing the related websites. You can get information by entering it. Subject directories are created by indexers. In addition to general topic directories, there are two other directories - specific subject directories and information distribution centers. Storage, which includes providing links to websites in their search directories or links to Web sites for these specific directories. This feature facilitates the placement of specific information on the Internet.

4.7.2 Search Engine

Search engines are designed to help users find relevant terms of information on the Internet. These are based on allowing users to enter a database by selecting key words. These differ from topic directories, in which indexers are used to create an index. Search engines use software programs that



automatically create their own databases that include lists of web pages. Search engines are made up of three different components – (i) a program or robot or crawler called a spider, (ii) a database with indexes, and (iii) a search software spider, moving around the web. travel, slowly crawling from site to site following links between pages. Different types of search engines use different types of spiders while some inspect each possible site, some based on more selective principles. And only inspect popular sites. The first type of spider can search through a large store of information in a short time, while the second type of spider presents only a few pages as a product, possibly with more relevant information. Each page retrieved by Spyder is collected in a database and its contents are indexed automatically from web pages and alphabetically using some of the same principles used as Inverted File, The rows are arranged in descending order from the top. The index, therefore, is a list of each word (except the discarded words) containing a pointer to the location of these words in the database.

Different search engines follow different principles. For example, some index every word on every web page, while other search engines index only the title and top-level phrase on the website. The third component or component of a search engine is. - Search software. It is a program that compares the user's search query (which is typed by keyboard) with the index, finds and finds similar words and sorts them in the order of correspondence. The criteria for determining conformance varies according to the search engines.

The different approaches used by search engines result in completely different results (products) as a result of crawling the web, searching for new pages, and indexing them. Therefore, when a similar topic is searched by different search engines, there are variations in the results obtained. In the way search engines operate, search engines are more oriented to find large volumes of specific information than subject directories. This is because their search is based on web pages and they automatically index editors from sites (instead of a pre-defined index) as subject directories do.

4.7.3 Examples of search engines include

(Alta Vista (http:\ \altavista.digital.com \)]

[Excite (http: \ \ www.excite.com\)]

[Hot Bot (http:\\www.hotbot.com\)]

[Infoseek (http:\\www.infoseek.com\)]

Lycos (http: \\ www.lycos.com\)]

Note

Since different search engines have different capabilities, new search tools have been invented a while back, which enable people to search different databases of search engines simultaneously using the same interface. These tools are called Multi Threaded Search Engine. Although, like individual search engines, not all searches have flexibility, but they are very fast and can search even from a large amount of information. The following are some examples of multi-thread or meta search engines.

(a) Dogpile (http:\\ www.dogpile.com)]

(b) Metacrawler (http:\\ www.metacrawler.com)]

4.7.4 Valuation of Internet Resources

Conducting business (which is providing library service or any other job description) efficiently and effectively through the use of the Internet should be considered as one of its main objectives. What is Internet Information? What facilities and services are required by the employees for their work within an organization and how the Internet can be helpful in their scientific, technical, as well as commercial communication and cooperation, it is very important to evaluate all these.

It has been observed that libraries use Internet resources as users' representative to answer questions, or provide users with exactly the same type of services that direct them to use specific Internet resources. Therefore, it is necessary to evaluate the resources of the Internet.

Although no specific criteria are available for this task, the traditional criteria used to evaluate print work may be useful, to some extent, for evaluating Internet resources. Answers to the following questions may be useful in the process of resource evaluation :

(i) Who is the concerned user? Does he belong to the academic community or to the general public?

(ii) What is the frequency of updates? Is there information about the update?

(iii) To which institution is he affiliated?

(iv) What is the area of expertise of the resource developer? Is there a passage that describes the author/creator of the source?

(v) What is the relationship between resources and other resources on a similar issue? Is there any such link or reference for these related resources?

(vi) Is there any review or evaluation of the site available? What do they tell?

Note

(vii) Whether any permission is required for access, and whether there is any fee for access is taken?

The basic information obtained from the answers to the above questions shows the accessibility of these resources in libraries and their suitability.

4.7.5 Use of Internet Approach in Libraries

Libraries with Internet access can provide the following services to their users :

The users of the library can directly send their messages by e-mail to the basic information related to the functions of the library through the Internet.

New information services, such as linking a home page with electronic text, databases, and other Internet resources.

Access to library facilities in new ways like remote access of catalogs. Interactive home page that provides various features, such as reservation of library materials and inter-library loan and reference queries.

Connecting remote information and establishing links with information sources globally.

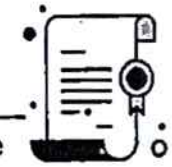
Setting up websites on the World Wide Web does not require expensive or sophisticated equipment. The basic components for its configuration may be a computer operating on any operating system used as a server. Computer systems powered by DOS, Windows, and UNIX can be used as servers for the creation of World Wide Web documents. It is not necessary to mention that different types of software are required for this purpose. Some of these are given below :

1. Server software to operate the Server to allow the Client to interact with your Server information;
2. H. T. M. L. Word Processing Software for the Job;
3. Gopher Software for creating images and icons; and
4. Scripting and programming software that allows the extension of the interaction between the client and the server.

The best approach for selecting a library would be to make certain high quality sources easily available to the users through their high potential interest and interface. Exploring Internet resources and conducting workshops for users to develop their skills to identify, evaluate and select resources relevant to their interest.

4.8 Consortia Meaning, Definition & Development

Information has been communicated from one person to another from the earliest stages of human civilization. Considered an important resource



Note

for dissemination. It is the basis of education and an important element for new ideas. Information exchange or information sharing is the way of cooperation. The mode of collaboration has changed from print based environment to digital environment with the spread of new information technology. Collaboration between resources to share their library resources has been practiced for decades.

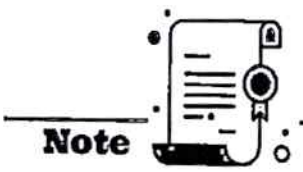
4.8.1 Concept of Consortia

The concept of federation probably evolved from the concept of co-operation, co-ordination and co-operation. The word consortia is derived from the Latin word 'concor' where congress together denotes the meaning and 'dissociation denotes association or society'. Published literature suggests that the concept is not new and refers to collaboration, coordination and collaboration between libraries for the purpose of sharing information resources. A federation can be described as those organizations that come together to fulfill a common objective, which requires co-operation or sharing of resources. The consortia consists of a wide, limited structure, full time staff and executive director web pages, policies, etc. They began to include other types of organizations as well as libraries, museums, hospitals, research groups, and historical societies. Huh. Working in consortia and subscribing to multiple resources also implies the ability to upgrade and adapt to new equipment and technology.

The exact date of the introduction of the word consortia is unclear but the meaning of consortia as being a federation or partnership has long been a doctrine of librarianship. A library consortium is a grouping of two or more libraries that have agreed together to collaborate with each other to meet some common needs. It is not about sharing resources, but also improving access to information. A consortium is a community that has formally agreed to co-ordinate or consolidate certain functions of a cooperative to achieve its objectives.

4.8.2 Functions of Consortia

1. Acts as the nodal agency for promoting collaboration among the participating institutions.
2. It coordinates all activities related to subscription to e-resources.
3. To review the progress of the Consortium at various stages and undertakes to review the progress on other issues.
4. The Union organizes annual meetings of the members.
5. Provides access to electronic resources and integrates them into the library program.



6. Bridging the gap between resource-rich and information resource-deficient libraries.

Conclusion

In this chapter, we have learned the meaning of database and also know how many types of databases are there. In addition, we also studied consortia in depth. Also know about the Internet and also studied the services provided by the Internet in depth.

Important Terms

- In the context of library and information science, a collection of documents which are referred for bibliographic details is called a database.
- Internet is originated from a network called EarthNet developed by the Advanced Research Projects Agency in 1970.
- Internet e-mail, news, remote-login f. It also provides services like TP etc.
- Consortia consists of broad limited structure, full time staff, executive director Deb Page, policies etc.

Exercises

Very Short Answer Type Questions

1. What is a database?
2. Define Internet.
3. Define consortia.

Short Answer Type Questions

1. How many types of databases are there? Briefly describe.
2. Briefly describe the evaluation of Internet resources.

Long Answer Type Questions

1. What is Internet Describing the services of Internet, describe in detail the use of Internet in the library.
2. Defining database, describe different types of databases and their uses.

Activities of Research Institution

In this chapter, we will study the functioning of research institutions as well as the role of library and information research science in India and how it was developing in India.

5.1 Introduction

During recent times, considerable research activities have been carried out in universities and research institutes in different parts of the world. Due to the establishment and active support of University Grants Commission, AICTE and other similar bodies in India many students are taking care of it.

During the pre-independence period, there were only a few doctoral degree holders, but after independence there was a huge increase in research output in every field. There are about 125 universities and research institutes in India offering PhD programs in library and information science. To analyze the research productivity of various universities in India keeping in view the available data.

5.2 LIS Education in India

Library and Information Science (LIS) education in India began in 1911 with the introduction of a training course in the erstwhile state of Baroda. The real beginning of system education at LIS was initiated by Dr. S.R. Ranganathan during the period 1926-1931 in the Library of Madras University with the Madras Library Association. The summer school led to the Certificate in Library Science, which the University of Madras continued until 1937 under the leadership of Dr. SR Ranganathan. Later Andhra University, Banaras Hindu University, Bombay University, Calcutta University and Delhi University started postgraduate diploma courses in library.

Note

During 1947, a total of 27 universities offered diploma courses in library science.

In 1957, for the first time in the country, Aligarh Muslim University started BLSC. Courses offered at different levels like Certificate, Diploma, Bachelor PGs, PG Diploma, Master's and Research degree programs i.e. M Phil and Phd under different mode (regular / on campus or distance / on campus or under some) Is performed. Both timing and plans. (annual or semester). The development of universities was ensured after independent India.

Improving the quality of teaching. It is because of the importance of libraries in various institutions, research centers and government departments, the demand for libraries has also increased. It really gave a boost to library science education in India.

5.3 Research in Library and Information Science

In the past research into LIS was primarily considered to provide a theoretical basis for professional practice. The roots of research in the LIS profession are not very deep. Library science research during the twentieth century can be traced to the Library School of the University of Chicago in the mid-1920s on the grounds of the LIS. The visionary efforts of the Chicago School bore fruit in abundance and provided leadership to the world in library science research (Shera 1976, 145). Due to social pressure, as well as motivation, the pace of library research is increasing everywhere today. Programs in our profession, it is insisted that if librarianship aspires to become a profession, it must rely on research to develop its knowledge base and its theoretical framework.

The credit for formal institution of doctoral degree program in library science in India does not go to Dr. SR Ranganathan (1892–1972). In 1951 he did library science at the University of Delhi. which led to many hardships and personal ridicule. The Delhi University gave the book to D B Krishna Rao in 1957. Awarded the first P.hD degree in Rhythmology, who worked on articulated classifications for agriculture. In 1955 and 1965, when Ranganathan shook the soil of Delhi with his feet, the scope of doctoral research continued. In the 1960s and 1970s, some doctorates on library-related subjects were earned by library professionals under the guidance and supervision of the respective faculties in subjects such as sociology. History, Law, Economics, Management and the like. The mantra of reviving and advancing doctoral research facilities was assumed by JS Sharma (1924–1993), the then University Librarian and Head of the Department of Library Science, Panjab University, Chandigarh. Under his guidance the second lawful Ph.D. Library Science | Awarded in 1977 after a gap of two

full decades. After that, looked back. Many universities followed, mostly with individual effort and enthusiasm. Doctorol Mudkar Nahi Research received a fillip in the 1980s and facilities in India gradually improved to maintain its third world leadership in library research and library literature. PHD. Thereafter programs were organized even after lack of facilities or adherence to standards (Mushroom MP 1999).

Note

5.4 Research Strategies in LIS

Surveys of Research Methods

LIS is a very broad discipline which uses a wide range of contantly evoluing research strategies and techniques. Various classification schems have been developed to analyze methods employed in LIS research. In 1996 kim synthesized previous categories, definitions and introduced a list of research strategies including data collection and analysis methods. The listing four journal strategies are—(i) theoretical/philosophical inequirey (ii) bibliographic research (iii) Development of storage and retrieval systems (iv) action research. Strategies are divided into quantitative and qualitative driven. Quantitative driven strategies are included descriptive studies, predictive studies, bibliometric studies, content analysis and operation research studies. Quantitative driven strategies are considered the following : Case study, bibliographical method, historical method, grounded theory etc.

Systematic studies of research methods in LIS started in the 1980s and several reviews literature have been conducted over the past years to analyze to pics, methodlogies and quality of research. One of the earliest studies was done by Peritz who carried out a bibliometric analysis of the articles publised in 39 core LIS journals between 1950 and 1975. She examined the methodlogies used the type of library and information service activities, a wide spread use of the survey methodology, a considerable increase in theoretical studies after 1965.

So, the literature shows a continued interest in the analysis of published LIS research. Approaches include focusing on particular publication years, geographic areas journal titles, aspects of LIS and specific characteristic such as subjects, authorship and research methods. Despite the abundance of content analysis of LIS literature, the findings are not easily comparable due to differences in the number and titles of journals examined, in the types of papers selected for analysis, in the periods covered and in classification schemes developed by the authors to categarize article topic and research strategies. Despite the differences, some findings are consistent among all studies :

Note

1. Data analysis is usually limited to descriptive statistics, including frequencies, means and standard diviations.
2. Over the years, there has been a considerable increase in the array of research approaches used to explore library issues.
3. Descriptive research methodologies based on surveys and questionaires predominate
4. Information seeking information retrieval, library and information service activities are among the most common subjects studied.

5.5 Data Collection and Analysis

Articles which had been published in the year 2011 were obtained from the following journals : Journal of Academic librarianship, information processing and management journal of Documentation, college and research libraries, library and information science research. These titles had been selected as data sources because they had the highest 5-years impact factor of the journals classified in which's serials. Directary under the "library and information science." subject heading. Only full length articles were collected from the journals. Each article was classified as either research or theoretical. Articles which employed specific research methodology and presented specific findings of original studies performed by the author were considered research articles. Articles reporting research in system design or evoluation in the information systems field were also regarded as research articles. Each article was classified into a topical category according to its main subject. The articles classified as research were then further explored and analyzed to identify (i) research approach (ii) research methodology (iii) method of data analysis.

The final list of the analysis codes was extracted inductively from the data itself. Research approaches are plans and procedures for research. Research approaches grouped as qualitative and mixed method studies.

Research methodologies or strategies of inquiry are types of research models which provide specific direction for procedures in a research design and inform the decisions concering data collection and analysis. The methodology classification included 12 categories. Each article was classified into one category for the variable research methodology. Methods of data analysis refer to the techniques used by the researchers to explore the original data and answer their research problems or questions. The array of data analysis methods included. The array of data analysis methods included the following categories :

1. Experimetnal Evoluation

Note

2. Qualitative Data Analysis
3. Inferential Statistics
4. Descriptive Statistics
5. Other Methods.

The experimental evolution was used for system and software analysis and design studies which assesses the newly developed algorithm tool, method etc. by performing experiments on selected datesheets. Descriptive statistics are used to describe the basic features of data in a study. Infurential statistics investigate questions, models and hypotheses. Qualitative data analysis is the range of processes and procedures used for the exploration of qualitative data from coding and descriptive analysis to identification of patterns and themes.

5.6 Research Approach and Methodology

Research articles were coded as quantitative, qualitative or mixed methods studies. An overwhelming majority of the empirical research articles employed a quantitative research approach.

Table 1 prsents the distribution of research approaches over the five most famous topics. The quantitative approach clearly previals in all topics. However, qualitative design seem to gain acceptance in all topic, while in information behaviour research, quantative and qualitative approaches are most evenly distributed.

Topics	Mixed methods	Quantitative	Quantitative
Information retrieval	0.0%	0.0%	100%
Library Services	3.6%	39.3%	57.1%
Information Behaviour	14.0%	40.4%	45.6%
Organization and Management	4.8%	23.8%	71.4%
Information Behaviour	14.0%	40.4%	45.6%

The most frequently used research strategy was survey, accounting for almsot 37.1 of all research articles followed by system and software analysis and design. This result is influenced by the fact that information processing and management addresses issues at the intersection between LIS and computer science and the majority of its articles presents the development of new tools, algorithms, method and systems and their experimental evaluation.

UDC- The system is kept up to date by the International Federation of Documentation, but due to the complex process of revision, it takes considerable time to incorporate the latest branches of knowledge science

5.4.2 Functions of Various Symbols used in UDC

Symbol	Functions
+ Plus sign	To combine the non-trailing digits
+ Stroke	To combine the trailing digits
: Colon	To combine two or more related concepts of equal importance
= Equal Sign	For Subdivisions of 4 Linguistics Section
( ) brackets	Indicates the geographical extent of Place Notation Subject in Brackets
: used with indo arabic numerals	To show the physical appearance displayed in the document zero in parentheses
(=) equal sign in parentheses	to indicate nationality and race
reverse coma	to show time
.00 Point zero zero	to show perspective
- hyphen	To show analytical representations of concepts to demonstrate perspectives
.0 Point zero	Analytical - but for the analysis of concepts
' Apostrophe	Interposition device to abbreviate notation to prevent repetition of two elements contained in a document title that have the same number
[ ] Square brackets	Investment strategy

5.4.3 Sequence of symbols used in UDC (Filling Order)

Logo	Example
=	= 30 German language
(0.)	(0.035.22) Microfilmed documents
(1/9)	430 Germany
(= ...)	(= 1.410) British nationals
" ="	"19" Nineteenth Century (Nineteenth Century)
+	622.341.1 + 669.1 Iron-ore mining and ferrous metallurgy. (Mineralization and Ore Metallurgy in Iron Mines)
/	622.341.1/2 Mining of iron ores and manganese ores

Note

Simple No.	.341.1 Iron-ore mining
:	622.341.1 : 338.124.4, Economic crisis in metallurgy in iron mines (class number can also be anagrammed as 338.124.4 622.341.1.
::	622.341.1 338.124.4 Economic crises in iron-ore mining (may not be reversed)
=	622.341.1 = 30 Documents in German about iron ore mining = 622.341.1 (0.035.22) Microfilmed documents about iron-ore mining = 622.341.1 (430) Iron-ore mining in Germany = 622.341.1 (= 1.366) Iron-ore mining among the ancient British = 622.341.1 "18" Iron-ore mining in the nineteenth century

5.5 Web Dewey

Edition 22 is the first edition of the Dewey Decimal Classification to be produced in the context of the web environment. Hence, Web Dewey means the database of Dewey Decimal Classification on WWW (World Wide Web). WWW is the network of computers across the world interconnected together on the internet, and using the concept of hypertext to link internet sites and information ion the internet. Like print versions of DDC, Web Dewey is also found in both full and abridged versions; they are Web Dewey and Abridged Web Dewey.

5.5.1 Key features

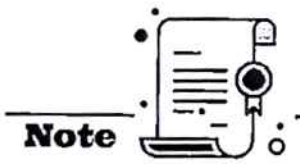
(i) All content from DDC 22 and Abridged Edition 14, including quarterly updates (new developments, new built numbers and additional electronic index terms).

(ii) An easy-to-use, browser-based interface that allows you to search the DDC (and related terminology) efficiently and navigate intuitively.

(iii) Thousands of Relative Index terms and built numbers not available in the print DDC.

(iv) Library of Congress Subject Headings (LCSH) that have been intellectually mapped to Dewey headings by DDC editors.

(v) Selected LCSH mappings from the Forest Press publication, People, Places & Things.



(vi) LCSH that have been statistically mapped to Dewey numbers from records in WorldCat (the OCLC Online Union Catalog).

(vii) Links from mapped LCSH to the LCSH authority records.

(viii) Quarterly updates, incorporating the latest changes to the classification and new LCSH mappings, index terms and built numbers.

(ix) An annotation capability, which allows you to add your own notes into Web Dewey to reflect local classification practices.

5.5.2 Web Dewey Updates

One of the main features of Web Dewey is its updates. OCLC make updates periodically to the data accessible via the Web Dewey or Abridged Web Dewey service, using its best efforts to make these updates available on a quarterly basis. The updates available till date are as follows :

2007 September	2007 May	2007 February	2006 November
2006 August	2006 May	2006 February	2005 November
2005 August	2005 June	2005 February	2004 November
2004 August	2004 April	2004 February	2003 December
2003 November	2003 October	2003 August	2003 July
2003 June	2003 April	2002 November	

5.5.3 Latest Enhancements

This release is scheduled for the September 23, 2007 install. This September release consists of updating the DDC 22 and Abridged 14 databases in Web Dewey and Abridged Web Dewey respectively. These database updates contain the latest changes to the DDC and monthly postings (New and changed entries) for May 2007, June 2007 and July 2007.

This September 2007 release includes :

(i) All updates to Dewey Decimal Classification, Edition 22 and Abridged 14 through August 2007 (corrections, new developments, new built numbers, and additional electronic index terms).

(ii) Thousands of Library of Congress Subject Headings (LCSH) that ave been statistically mapped to Dewey numbers from records in WorldCat (the OCLC Online Union Catalog) and intellectually mapped by DDC editors.

(iii) Thousands of Relative Index terms and built numbers not available in print.

Note

(iv) Links from mapped LCSH to the LCSH authority records.

(v) Selected mappings from Medical Subject Headings (MeSH).

(vi) LCSH that have been intellectually mapped to Dewey headings by DDC editors, including mappings to Abridged Edition 13 numbers from the OCLC publication, Subject Headings for Children.

(vii) Mappings between abridged Dewey numbers and the 2005 update to the 18th edition of H.W. Wilson's Sears List of Subject Headings.

5.5.4 Subscribe to Web Dewey

A Web Dewey subscription can begin at any time and extends for one year from the starting date. Requires an annual fee only; no per-use charges. It is available on a single-user or site-license basis.

There are two ways to order :

(i) Online using OCLC's secure Web store

(ii) By mail or fax using a printed order form

Web Dewey is available on an annual subscription basis, according to the table below.

Product	Type of License	Subscription Price
WebDewey (OCCLC Cataloging Services Edition)	Single user (for use by one staff member)	$250.00/year
	Site License (for use by 2-9 staff members)	$250.00/year
	Site License (for use by 10 + staff members)	$825.00/year
WebDewey	Single User (for use by one staff member)	$300.00/year
	Site License (four use by 2-9 staff members)	$665.00/year
	Site License (for use by 10 + staff members)	$975.00/year
Abridged Web Dewey (OCLC Cataloging Services Edition)	Single user (for use by one staff member)	$72.00/year
	Site License (four use by 2-9 staff members)	$165.00/year
	Site License (for use by 10 + staff members)	$225.00/year

Note

Abridged WebDewey	Single user (for use by one staff member)	$82.00/year
	Site License (four use by 2-9 staff members)	$200.00/year
	Site License (for use by 10 + staff members)	$275.00/year

5.6 D.C.C. Problems & Popularity of DD

The main criticism of DDC is being made for its pure and limited notation. In which there is no proper solution to the problems of classifying new subjects in their proper place and long class numbers. Apart from this, its westward orientation has also been a major cause of criticism. But despite these problems and criticisms, D.D.C. is the most popular general classification system used in more than 2 lakh libraries in 135 countries of the world. In the United States, 95 percent of school libraries, 25 percent of other academic libraries and 20 percent of specialized libraries use it. It is translated in Hindi, Arabic, Chinese, French, German, Greek, Hebrew, Italian, Korean, Spanish, and Vietnamese 30 other languages. DDC has been used in 62 national and many merchant music lists. DDC has the largest number of Indian libraries among Asian countries. Class Number of Mark (MARC) is also given in Tapes and in CIP data.

5.6.1 Important Terms

Mark bibliographic data is a set of different formats for use by computers that provide a set of conversions for their identification and adjustment.

Unimarc is an abbreviation for Universal Marc, the credit for its creation goes to Ifla. Its aim is exchange of data internationally in machine readable form between various national agencies. Web Dewy is available on the Internet in the form of multiple learning points and presentations.

5.7 CRG

Classification Research Group is a group of working librarians and others interested in classification research in London. When Ranganathan's ideas of faceted classification began to make an impression in the Western World, the Royal Society's Scientific Information Conference was held in London in 1948, where classification as a method of subject organization, was discussed as one of the themes. Dissatisfied by the prevailing methods of subject organization, a committee with J.D. Bernal as secretary was set up to examine the existing systems and suggest possible improvements. No progress was however, made until 1951, when B.C. Vickery was invited to form a group to take over the work of the committee. This heralded the formation of CRG in 1952. The group consists of people who are keenly

Note

interested in classification research. A perusal of the reports produced by the CRG indicates that the group has been actively involved in the creation of several classification schemes for such organizations. The theoretical work of the group has involved the study of facet analysis, relational operators and the theory of Integrative levels." The efforts of the group are directed in two directions : Classification and data mining. The classification effort focuses on both methodological research and particularly novel, non-standard application. The work in classification has significant overlap with other areas, including machine learning and pattern recognition, so that the publications appear in a wide literature.

5.8 BSO (Broad System of Ordering)

Unesco in 1971 needed a switching language for its UNISIST Programme. Work for such a language started in 1972 as a part of European Communication Project, after the Budepest meeting in 1972, FID set up a new working group ..... FID/SRC whose scope was set out as "For purpose of interconnection and cooperation between information systems, the working group will design and develop a broad subject ordering scheme for all fields of knowledge and usable in manual and mechanised information systems." The first draft was prepared by Eric J. Coates and G.A. Lloyd. Third edition was published in 1978 by the FID.

Its major subject areas are outlined below :

100	Knowledge in general	500	Humanities, Cultural and Social Science
200	Science and Technology	600/890	Technology
300	Life Sciences	910	Language, Linguistics and literature
460	Education	940	Arts
480	Sports and games	970	Religion

It was designed to serve as a switching language *i.e.* to work as an intermediary or conversion language to transfer information from one indexing language to another. It means that is was designed for classifying information centres rather than documents. But, instead of its intended purpose, it is being used as shelf classification in some libraries.

5.9 Automatic Classification

Automated text classification (ATC) is a relatively new research field spun off from traditional information retrieval research. ATC is defined as the task of automatically classifying documents into a pre-defined set of classes.

Note

Rule-based methods and Machine Learning are two dominant approaches to ATC until the late '80s and since the early '90s, respectively. In rule-based approach, knowledge engineers and domain experts define a set of rules to be applied for identifying the class of unclassified documents. In ML approach, the parameters of the models are statistically determined using a corpus of manually pre-classified documents. The categories of unclassified documents are predicted from the parameterized models.

Automated Classification Projects based on Large Classification Schemes

In recent decades, classification schemes-based ATC projects have been undertaken in the information community as an effort to arrange heterogeneous information resources under one single organizational structure. This paper focuses on the approaches for classification scheme-based ATC and presents an analysis of the conceptual models used in the ATC projects. A total of 16 research projects and studies published across the period from 1992 to 2011, which are outlined at the table 1. The selection of the projects and studies are not randomly made. Instead, its primary target is to collect all studies based on the three popular library classification schemes, Library of Congress Classification (LCC), Dewey Decimal Classification (DDC), and Universal Decimal Classification (UDC). Other large scale classification systems such as web directories are also considered for the collection. A project and study often utilizes more than one model for a comparative study with the adopted methods; the first column of the table below indicates the key classification models only.

Table : ATC research projects based on large classification schemes in chronicle order.

Key Classification Models	Classification Projects/Studies	Classification Schemes/Best Result
Clustering algorithm : creating a classification cluster for an LCC class, with inner product for matching: between classification cluster vectors and documents vectors	Automatic selection of Library of Congress Classification numbers based on the titles and subject headings in MARC records in LCC class Z (Bibliography, Library Science and Information Science) : (Larson, 1992)	Library of Congress Classification (LCC) About 47% in accuracy with the test set of 283 MARC records into major ranges of LCC class Z

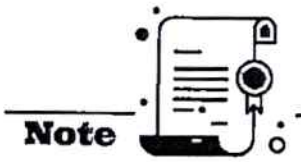
Note

Term weighted algorithm	Automatic classification of a total of 660 databases in a Wide Area Information Server based on keywords, descriptions, and subject terms in the two top levels of UDC : Nordic WAIS/World Wide Web project (Ardö et al., 1994)	Universal Decimal Classification (UDC) A 'fair' number of misclassifications
Vector space model : Latent Semantic Indexing that associates terms with the relevant categories in the LCC Outline	Automatic classification of news articles in Usenet newsgroups based on news article headers in the 4214 classes of the LCC Outline : Alexandria Digital Library Project (Dolin et al., 1997)	LCC No evaluation result has been reported
Vector Space model: with a version of SMART (System for Manipulating And Retrieving Text) system	Classifying web resources into DDC: Scorpion project (Thompson, Shafer, &Vizine-Goetz, 1997)	DDC No evaluative outcome is reported
Clustering algorithm : a cluster corresponding to a DDC class, with cosine normalization for the distance measure between clusters and input documents	Classifying web pages into DDC : Wolverhampton Web Library (Jenkins, Jackson, Burden, & Wallis, 1998)	DDC Report 40 percent accuracy without further details in the experimental setting.
Vector space model: using term weight-based algorithm	Classifying web pages: DESIRE (Koch & Ardö, 2000) European Union-supporting Project	Engineering Information classification system Report about 60% accuracy in a test with about 1,000 web pages.

Vector Space model with the cosine angle to measure the similarity between web pages and subject categories	Classifying web pages: (Chung & Noh, 2003)	DDC 0.77 in precision over 757 subclasses of the DDC main class of 'Economics' (330)
Machine Learning : Support Vector Machine to create a linear classifier combined by pairwise	Automatically assigning a LCC to a work given its set of Library of Congress Subject Headings (LCSH): INFOMINE (Frank & Paynter, 2004)	LCC 55% in accuracy in a test with a collection of 50,000 LCC/LCSH pairs- for the entire LCC Outline
classification for classifying websites		
Probabilistic model: with Naïve Bayesian (NB)	Classifying scientific papers about computer science : (Avancini, Rauber, & Sebastiani, 2004)	ACM classification scheme consisting of 1,474 categories No test evaluation is reported
Probabilistic network model : with ExpNet (Yang, 1994)	Classifying web pages in Google and LookStart web directories (Avesani, Girardi, Polettini, & Sona, 2004)	Google web directories & LookSmart web directories A range from 9.8% to 55.1% in F measure with the subdirectories of 9 top-level Google directory categories and a range from 6.8% to 39.5% in F with the subdirectories of 10 top-level LookSmart directory categories, in a test of classifying about 10K web pages

Note

Probabilistic model: with hidden Markov model (HMM)	Classifying dissertation abstracts : (Yi, 2005)	LCC Higher accuracy classification with HMM over with NB in classifying 625 dissertation abstracts into 25 secondlevel subclasses in Q (Science), S (Agriculture), and N (Fine Arts): More than 45% accuracy with HMM vs. less than 5% with NB
Machine learning technique with Support Vector Machine	Classifying web pages: (Liu et al., 2005)	Yahoo! web directory As low as 0.24 in F measure for about 246,000 categories
Term weight-based method	Classifying web pages into classes of the Ei classification system : (Golub, 2006)	Engineering Information thesaurus and classification scheme No formal evaluation is reported in classifying a sample of 1,000 web pages (preselected and classified by librarians)
Clustering algorithm: with k-nearest neighbors (KNN) and the cosine similarity measure	Classifying Reuters articles : (Pong, Kwok, Lau, Hao, & Wong, 2008)	LCC Report 0.8 in F measure as the best performance in a test of classifying 254 documents into 67 LCC classes or subclasses
Machine learning technique with Support Vector Machine	Classifying bibliographic data : (Wang, 2009)	DDC Report a range from 0.49 to 0.85 in F measure in various tests with 4,861 DDC classes



Citation network-based term weighting method	Classifying scientific literature into DDC : (Joorabchi & Mahdi, 2011)	DDC A range from 0.61 to 0.89 in F measure in classifying 400 documents into 8-level DDC sub-classes

ATC Studies : Approaches and Models

Below I summarize the key models/algorithms that have been adopted and utilized in the classification projects and studies.

- **Vector Space model (5)*** : A set of categories and test documents are represented as vectors in a multi-dimensional vector space. The degree of similarity between categories and documents are generally evaluated by a cosine of angle between category vectors and document vectors.
  - General vector space model (Chung & Noh, 2003; Godby & Stuler, 2001; Koch & Ardö, 2000; Thompson, Shafer, & Vizine-Goetz, 1997).
  - Latent Semantic Indexing (Dolin et al., 1997)
- **Clustering algorithm (4)** : Clustering algorithm group similar objects into a same subgroup called cluster. In classification, a new document is assigned to the majority class of the cluster closest to the document.
  - K-NN (Pong, Kwok, Lau, Hao, & Wong, 2008)
  - Other clustering algorithm (Jenkins, Jackson, Burden, & Wallis, 1998; Larson, 1992)
- **Machine learning (4)** : Machine learning is a statistical learning process that builds a classifier by acquiring the characteristics of the categories from previously existing examples.
  - Support Vector Machine (Frank & Paynter, 2004; Liu et al., 2005; Wang, 2009)
  - Hidden Markov model (Yi, 2005)
- Term weighting algorithm (3) (Ardö et al., 1994; Golub, 2006; Joorabchi & Mahdi, 2011): A metric to measure the degree of similarity (or score) between subject categories and new document is based on term weighting schemes involving indexing terms and subject terms.

*The number inside the parenthesis next to the classification model/algorithm refers to the number of classification projects and studies in which the classification model/algorithm is used as its key model/algorithm.

Note

- Probabilistic model (2): Given a sample input, probabilistic model-based classifier is able to predict a probability distribution over a set of classes.
  - General networks (Avesani, Girardi, Polettini, & Sona, 2004)
  - Naïve Bayesian (Avancini, Rauber, & Sebastiani, 2004)

Automated classification models can be classified into the following two groups, depending on types of learning :

- Supervised model (Pong, Kwok, Lau, Hao, & Wong, 2008; Wang, 2009; Yi, 2005): The category learning process is "supervised by the knowledge of the categories and of the training instances that belong to them." (Sebastiani, 2001, p. 9) Training data include both the training documents and the correct categories.
- Unsupervised model (Chung & Noh, 2003; Golub, 2006; Jenkins, Jackson, Burden, & Wallis, 1998): The category learning process is made "in a way that reflects the statistical structure of the overall collection of input patterns." (Dayan, 1999) This model is not provided with the correct categories of the training documents.

The ways of dealing with automated classification can be divided into the following two based on whether or not considering structural relationships among categories :

- **Flat classification approach :** In flat classification, categories are considered as independent and the relationships among categories such as a hierarchical structure in a classification scheme are ignored. Thus, the computation cost is proportional to the number categories. In a large scale of categories, its cost increases to an infeasible level. Classification models find the most relevant class, based on the assumption that all classes are in a flat structure. (Larson, 1992; Dolin, 1998; Thompson et al., 1997).
- **Hierarchical classification approach :** In hierarchical classification, a large-scale classification task can be decomposed into smaller tasks. Such a divide-and-conquer strategy greatly reduces the computation cost. Classification models find the most relevant class which a new record is the most likely to belong to while considering the hierarchical structure of a classification scheme.