

4.2 SUBJECT:- INFORMATION STORAGE & RETRIEVAL

UNIT – I

Introduction of Information Storage and Retrieval System

1.1 Definition of ISRS :- It is a system that is designed to store and manage information and then make it available to users when they need the information.

ISRS are used in a wide variety of application including, archives, museums, businesses, and government agencies. They can be used to store and retrieve a wide variety of information including text, images, audio and video.

Characteristics & Objectives of ISRS :-

- 1) ISRS to improve access to information.
- 2) Make it easier for users to find the information they need.
- 3) To save user's time & effort.
- 4) Increase efficiency.
- 5) Improved accuracy.
- 6) To find up to date information.
- 7) To reduce cost of storing & managing information.

Components of ISAR System :-

Here are the key components of an information storage and Retrieval system.

1) Users Interface :- It's the gateway through which users interact with the system.

It Provides features :-

Entering Queries
Browsing Content

A well designed interface :-

Easy to use
Guides users effectively
Supports various search strategies

2) Knowledge Base :- It's the heart of the ISAR system it stores and organizes the information to be retrieved.

Forms of Knowledge bases.

- Textual documents
- Images
- Audio files
- Video Recoding
- Structured data

3) Search Agent:- It's the component that processes user queries and retrieves relevant information from the knowledge base.

- Analyzing queries of users.
- Matching queries
- Result/ Relevance

4) Indexing & Retrieval Mechanism:-

- To create structured representation of information
- Techniques like – keyword Retrieval

5) Evaluation & Feedback:-

Assesses the systems effectiveness in retrieving relevant information.

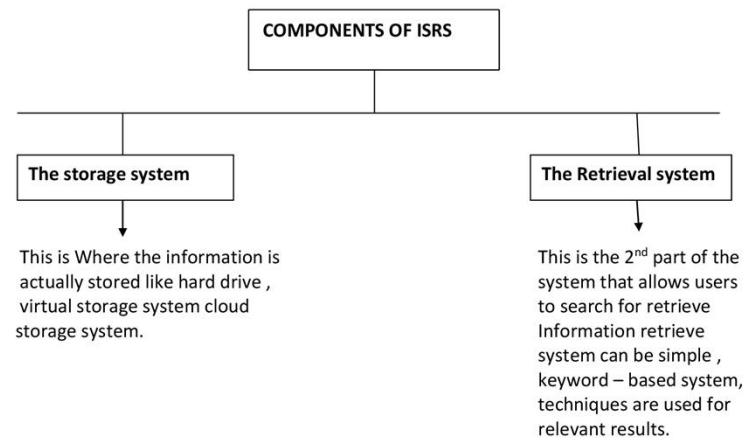
Accuracy -> Completeness -> User satisfaction

Feedback – It is for the improvement of the system.

6) Additional Components :-

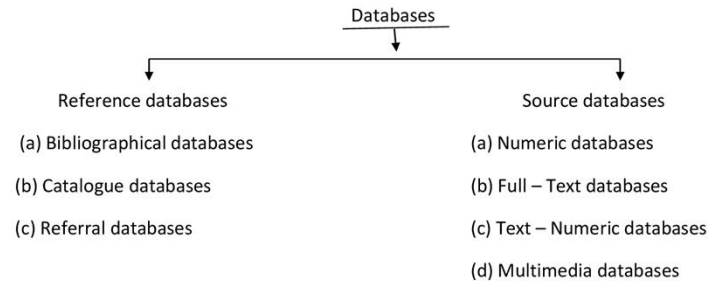
Data Management

Security – Protects information



1.2 Definition of Database :- A database is an organized collection of structure information or data, typically stored electronically in a computer system. It's designed to make data easily accessible, management, and updatable.

Types of Database :-



Numeric databases :- Numeric databases contain numerical, statistical or survey types information to give answers of numerical queries such as diameter of Mars, number of countries in the world etc.

Bibliographic databases :- Bibliographic databases contains bibliographic details like document number, title, author, ISBN, publisher, year imprints, abstracts, indexing words etc.

CD – ROM Databases :-

- Physical storage: Data is stored on a CD-ROM disc, which must be inserted into a computer's CD-ROM drive to access the database.
- Limited capacity: Typical capacity is around 650-700MB, restricting the amount of data that can be stored.
- No internet connection required: Can be accessed offline, making them useful in areas with limited or no internet access.
- Fixed content: Data is fixed at the time of production and cannot be updated
- Internet connection required : Must have an active internet connection to access the database.
- Frequently updated : Content can be updated regularly ,ensuring access to the most current information.
- Widely used today : The predominant form of database access due to internet ubiquity and advantages.
- Examples : Google scholar , PubMed , JSTOR, LexisNexis, ProQuest, EBSCOhost.

Key Differences:

Feature	CD-ROM Database	Online Database
Storage medium	Physical CD-ROM disc	Remote servers accessed via internet
Capacity	Limited (around 650-700MB)	Large (terabytes or more)
Internet connection	Not required	Required
Content updates	Fixed at production	Frequently updated
Common usage period	Pre-internet era	Widely used today

Online Databases

- Remote storage :Data is stored on remote servers, accessed through the internet.
- Large capacity : Can hold vast amounts of data, often in the terabytes or more.

Purposes of Databases :-

- Storing and managing large amounts of information.
- Retrieving specific data efficiently.
- Analyzing data to make informed decisions.
- Tracking transactions and records
- Maintaining customer information
- Supporting e-commerce websites.
- Powering social media platforms.
- And many more applications in various industries.

1.3 Information Retrieval Process and Search Strategy :-

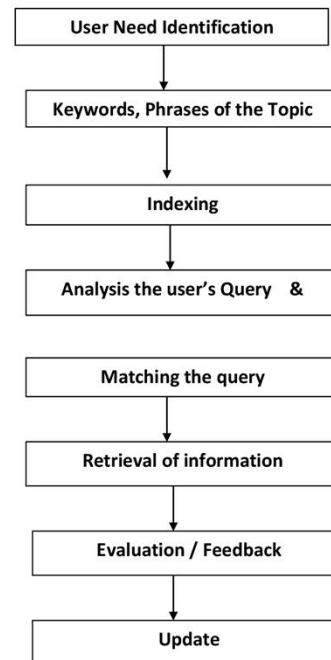
A search strategy in information retrieval is a systematic plan. It involves several steps & techniques.

Here are some points for a good search strategy:

- 1) Define the research question or topic
 - To clear the questions & topic
 - Break down the topic into key concepts
 - Identify the main facet of topic
- 2) Select appropriate information sources
 - Select database & websites
 - To consider news articles, Journals, Conference proceeding & Industrial Report.
- 3) Develop search terms & keywords :
 - Use synonyms
 - Use subject heading or thesauruses
 - Consider controlled vocabularies
- 4) Apply search techniques:
 - Use Boolean techniques (And, OR, NOT)
 - Use truncation (*)
 - Use Indexing & Abstracting , Sources
- 5) Evaluation the Finding:-
 - Analyze the results
To check of the topic relevant finding

A well defined search strategy is efficient & effective information retrieval.

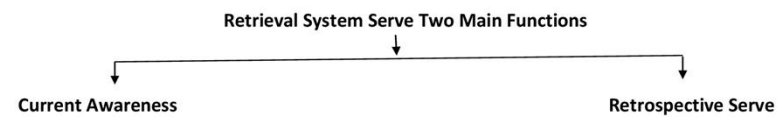
Information Retrieval Process



1.4 Evaluation of ISAR System :- Information retrieval is the process of recovering specific piece of information form stored data. It the an integral part of the communication process.

Information retrieval has four Phases

- 1) Word Retrieval - Word means
- 2) Reference Retrieval – Library Catalogue, Index
- 3) Document Retrieval - Complete copy of the document
- 4) Data Retrieval - Sought Information form the document



A System should be design to serve both functions.

In the construction of an information retrieval system there are there terms.

1) Descriptors:- Descriptors are the keywords which analyse the thought content of the documents.

2) Specifications:- It means here: Pinpointed information about the author, title, publisher, pagination etc.

3) Address:- Here it means – code Number/call No. Shelf Number, File Number or which facilities retrieval of documents.

All the three terms taken together complete the description of documents and is an integral part of the information storage & retrieval system.

(A) Evaluation an ISRS effectively requires many key aspects.

1) Effectiveness: - How well do retried documents match the users information need.

Metrics like precision and recall are crucial.

Precision:- Proportion of relevant documents retrieved.

Re-Call:- Proportion of relevant documents found.

2) Completeness: - It means that No one relevant document is missing and the system retrieve all relevant documents.

3) Accuracy: - Are retrieved documents correct and reliable?

(B) Efficiency

(1) Speed: - Response time is less and user satisfaction is must.

(2) Resource Utilization: - Does the system use computerised resources efficiently.

Flexibility: Can the system have more search option.

Adapted to individual users needs.

(C) Other factors:-

1) Is ISRS reliable

2) Dose the system protects user data and security and privacy.

3) Cost- effectiveness: Is the system affordable to implement and maintain?

UNIT – I**QUESTION BANK****Short answer type question**

Q.1 Define ISAR System.

Q.2 Write the Objective of ISAR system.

Q.3 Define Database.

Q.4 Write any two types of Database.

Q.5 Write the full form of Database.

Q.6 Write the meaning of online Database.

Q.7 what is meant by CD – Rom Database.

Long answer type question

Q.8 Define Database. Explain the types of database.

Q.9 Write the characteristics of ISAR system/

Q.10 Describe the Components of ISAR system.

Q.11 How to evaluate the ISAR system.

Q.12 Describe the search strategy of ISAR system.

Q.13 Differentiate between C D – ROM database & online database.

UNIT- II Bibliographic Description

2.1 Principles and Evolution of Bibliographic Description

Bibliographic Description:

A bibliographic description is like an ID card for a piece of information, like a book, article, website, or any published material. It's a standardized way to identify and describe a resource in a clear and consistent format, making it easier for people to find it and understand what it's about, even if they don't speak the same language or follow the same citation style.

Here are some key things to know about bibliographic descriptions:

Structure: They follow a set of rules called the international standard bibliographic description (ISBD), developed by the international federation of library Associations and institutions (IFLA).

Content: The description typically includes elements like:

- **Title:** The full title of the resource, including subtitles.
- **Author(s):** Names of the people who created the content.
- **Publication details:** Publisher, Publication date, edition, etc.
- **Physical descriptions:** Number of pages, format (e.g., print, online) etc.
- **Notes:** Additional information about the resources, like translations, awards, etc.

Benefits:

- **Standardization:** Makes it easier to find and organize information across different sources.
- **Clarity:** Provides a concise and accurate overview of the resource.
- **International Access:** Helps people understand resources even if they don't speak the same language.
- **Citations:** Used to properly cite sources in academic work and other publications.

Principles and Evolution of Bibliographic Description

The principal and evolution of bibliographic description are fascinating topics that have undergone significant development over centuries, influenced by :

1 Changing technologies

2 User needs

3 Information landscapes

Principles of Bibliographic Description

1) **Clarity & accuracy:** Information provided should be precise and verifiable.

2) **Ease of understanding (Consistency)**

3) **Specificity**

4) **Objectivity**

5) **User friendliness**

6) Economy: Concise descriptions while

7) International

Evolution of Bibliographic Description

- 1) Early years: Initial focus on physical characteristics like title, author, publication details for physical cataloguing.
- 2) 20th century: Emergence of standardized cataloguing codes like AACR-2 & ISBD to achieve consistency.
- 3) Late 20th century: Rise of new formats and information needs like ISBD for different materials.
- 4) 21st century: Focus on user centric approaches.
- 5) The future: Continued adaptation to the digital age with emphasis on interoperability, machine readable formats and in tegration with new information retrieval technologies.

2.2 Rules for Bibliographic Description

- 1) Choose a recognized standard like (ISBD or RDA, MARC-21)
- 2) Identify the of Resource (Book, Article, Website)
- 3) Focus on Essential elements (Title, Author, and Editor)
- 4) Use Specific punctuation marks. (Colons, Semi Colons)
- 5) To take Information directly from the source.
- 6) Use note for additional Information.
- 7) To consider User's needs.
- 8) Follow the order.

2.3 Standards for Bibliographic Record Formats

1) ISBD:- International Standard Bibliographic Description

Purpose of ISBD:- Establish a standardized format for cataloguing at International Level

Feature of ISBD :-

- 1) Bibliographical detail – Title, Author, Publisher etc.
- 2) Use of Specific Punctuation marks.
- 3) Applicable to various resources like-books, Maps, Periodical etc.

2) MARC-21 :- Machine Readable Cataloguing

Purpose: - To facilitate exchange the bibliographical record among libraries. It is a machine readable Cataloguing format.

Key Features of MARC-21

- 1) It is a machine readable format of cataloguing.
- 2) It defines field & subfields for different data elements.
- 3) Common base format for bibliographical detail like author, title, publisher etc.

3) CCF :- Common Communication Format

Purpose of CCF :- To exchange the bibliographic information among libraries particularly scientific & technical fields.

Key Features of CCF

- 1) Focus on content rather than presentation.
- 2) Common format for cataloguing.
- 3) To facilitate data exchange between two or more libraries.

4) Doubling Core

Purpose:-

- 1) Simple Set of metadata elements for describing digital resources.
- 2) To promote resource discovery on the web

Key Features:-

- 1) 15 Core elements (Title, Creator, Subject etc.)
- 2) Widely used for web-based resources.

5) BIBFRAME:- Bibliographic Frame work

Purpose:- In digital environment to like the data model for bibliographic information.

Key Feature:-

- 1) To represents bibliographic entities as interconnected web resources.
- 2) To facilitate data sharing & reuse.

6) ISO – 2709:- International standard for the structure of bibliographic records.

7) RDA:- Resource description and access.

It is a cataloguing standard & to provide guidelines for cataloguing of documents.

The choice of format depends on factors such as type of resource. Types of library & need of the library & Information Centre.

2.4 Metadata Concepts, Metadata Standards: Dublin Core, MARC 21 etc.

Metadata Means :- “data about data” metadata refers to the information that describes the another piece of data. It is label of box of files.

Which describe what's inside when it was created and who put it.

Types of metadata

- 1) Descriptive:- Title, Author, Subject, Keywords, Date, File size.
- 2) Technical: - File Format, Encoding, Resolution, Compression.
- 3) Structural: - Relationships between different data elements.
- 4) Administrative: - Access rights, Security Information, Version control.

Benefits of Metadata

- 1) It makes data easier to find & retrieve.
- 2) It provide context & meaning to the data.
- 3) To facilitate data sharing and reuse.
- 4) Data management & long-term preservation.

Metadata Standard:- A vast amount of data generated globally. Metadata standards define common ways of structuring and formatting metadata.

Types of Metadata Standards

- 1) **Dublin Core**:- Its widely used for general descriptive metadata across various types of content.
- 2) **Library of congress**:- Cataloguing data standards for bibliographic description libraries (MARC-21)
- 3) **Federal Geographical data committee.**
- 4) **Clinical data interchange standards.**

Metadata standard help us manage, Interpret, better collaboration & analysis.

- 1) **Consistency**:- Data is descended in a uniform manner, regardless of its origin.
- 2) **Data can be exchanged & understood by different systems & application.**
- 3) **Accuracy and Completeness.**

Choosing the appropriate metadata standard depends the specific context & purpose of data.

Dublin Core :- The Dublin core metadata element set (DCMES), commonly known as Dublin core, is a set of 15 core metadata elements used to describe digital or physical resources. Developed by the Dublin core metadata initiative (DCMI), it aims to provide a simple and interoperable way to describe these.

Contributed to the Resource

- **Data**: When the resource was created or modified.
- **Type**: The resource's category (e.g., book, image, website.)
- **Format**: The file format or physical medium.
- **Identifier**: A unique identifier for the resource (e.g., ISBN, URL).

- **Source:** Where the resource originated from.
- **Language:** The language of the resource.
- **Relation:** Relationship to other resources.
- **Rights:** Intellectual property rights associated with the resource.

The 15 Core Elements:

- **Title:** Identifies the resource.
- **Creator:** The person or organization responsible for the resource.
- **Subject:** The topic or theme of the resource.
- **Description:** A summary of the resource.
- **Publisher:** The entity that made the resource available.
- **Contribution:** Individuals or organizations who contributed to the resource.
- **Date:** When the resource was created or modified.

Limitations of Dublin Core:

- **Limited expressiveness:** The core elements may not be detailed enough for complex resources.
- **Lack of domain – specific terms:** Requires additional vocabularies for specialized fields.
- **Incomplete implementation:** Not always used consistently or comprehensively.

Overall, Dublin core remains a valuable and widely used metadata standard for general resource description, its simplicity and interoperability make it a great starting point for organizing and sharing information effectively.

Benefits of Using Dublin Core:

- **Simplicity:** Easy to understand and implement, even for non-technical users.
- **Interoperability:** Compatible with various systems and applications, facilitating data exchange.
- **Flexibility:** The core elements can be extended with additional terms to meet specific needs.
- **Standardization:** Promotes consistent and accurate descriptions of resources.
- **Widespread adoption:** Used by libraries, museums, archives, and other organization globally.

UNIT-II

Short Answer Type Question

- Q.1. Define Bibliographic Description.
- Q.2. Write the meaning of Bibliography.
- Q.3. Write the full form of ISBD.
- Q.4. Write the name of standard for Bibliography Description.
- Q.5. Write the full form of CCF & MARC.
- Q.6. Write the types of metadata.
- Q.7. MARC-21& doubling core are _____ of Bibliographic Description.

Long Answerer type question

- Q.6. Write the benefits of Bibliographic Description.
- Q.7. Write the Principle of Bibliographic Description.
- Q.8. Write the rules of Bibliographic Description.
- Q.9. Explain the criteria of evaluation of Bibliographic Description.
- Q.10. Describe MARC 21.
- Q.11. Describe Dublin core in brief.
- Q.12. Explain the Metadata with its benefits.
- Q.13. Write the types of metadata standard.
- Q.14. Write the benefits of Dublin core.

UNIT – III Vocabulary Control

3.1 Definition and Purposes of Vocabulary Control

3.2 Tools of Vocabulary Control: Classification Schedules, Subject Heading List & Thesaurus

Classification Schedule.

Definition:

1. In Library & Information Science, a Classification Schedule is a systematic and hierarchical scheme used to organize books, Journals, and other materials within a library.
It consists of a series of main class, sub-classes and further divisions that represent different subjects.

Purpose of classification schedules.

1. To organise the reading material in a library.
2. To facilitate browsing and retrieval of materials by users.
3. Related subject books together on the shelves.
4. Optimize space utilization within a library.
5. It saves time if different libraries using the same classification schedule.

Example of Classification schedules.

1. DDC (Dewey Decimal Classification)
2. UDC (Universal Decimal Classification)
3. CC (Colon Classification)
4. Bliss Bibliographic Classification.

Subject Heading :-

Subject headings are like tags or labels that describe the content of an information resource.

Such as a book, article or website. They are controlled vocabulary terms.

Definition: Subject Headings are specific terms or represent their main topics.

* They are used in library catalogues, databases and other information systems to help users find relevant resources.

Definition of SUBJECT HEADING

A subject heading is part of a systematic list of terms that describe a given subject. According to Reitz, a subject heading is "The most specific word or phrase that describes the subject, or one of the subjects, of a work, selected from a list of preferred terms (controlled vocabulary) and assigned as an added entry in the bibliographic record in the library catalogue."

Functions & Purpose of Subject Heading

- 1) Readymade subjects heading can fully-satisfy the requirement of users.
- 2) The determination of the name of the subject.
- 3) To assist the cataloguer on how to determine the specific subject.
- 4) Reader's satisfaction.
- 5) To save the time of users.
- 6) To helpful in added entries.
- 7) See also reference are given to inform the users. What alied material is available in the library for which he is looking for.
- 8) Uniformity in subject heading.

Three popular subject heading lists are : 1.Sears List of Subject Headings (SLSH

2 Library of Congress Subject Headings (LCSH)

3 Medical Subject Heading (MESH).

SEARS LIST OF SUBJECT HEADING

3.3 Thesaurus – Definition and Types :-

Define Thesaurus:

A Thesaurus is a reference book that list & words and their synonyms and antonyms.

Synonyms – Similar words

Antonyms – opposite meaning words.

Entomology - Some thesaurus also provide information on the origin and history of words.

Definition of Thesaurus:- Thesaurus is a dictionary that lists words along with their meanings synonyms, etc. In a alphabetical order, but it assembles all words related to an ides at one place.

Purpose of Thesaurus :

1. It help us to find words that have the same or similar meaning.
2. It also helps to fine words that have the opposite meaning.
3. Some thesaurus also provides information on the origin and history of words.
4. Improve vocabulary: - Using a thesaurus can help you expand your vocabulary and use more precise language.
5. Avoid repetition.
6. Clarity and conciseness.
7. It can help to understand the meaning of complex words.
8. Thesaurus has different shades meaning, so we choose the right word.
9. It gives new ideas for writing because it have more synonyms and antonyms.

Types of Thesaurus :- On the basis of the nature of terminology and control thesaurus are mainly of two types (a) controlled thesaurus, and (b) free language thesaurus. The controlled

thesaurus allow only one term, whereas free language thesaurus allows use of all terms to denote a concept to be used for indexing and searching.

- (i) General Thesauruses (ii) Subject-Specific Thesaurus.

Functions & purpose of Thesaurus :-

- 1) It provides a standard vocabulary for a given subject field.
- 2) It shows the relationship between the terms.
- 3) It helps the indexer and the searcher in choice of preferred terms.
- 4) It provided hierarchical display of terms.
- 5) It increases the speed of retrieval by use of terms.
- 6) It provides a map of a given subject field.

Vocabulary Meaning :- The complete set of index terms in a subject indexing system to represent what the document is about, is referred to as the Vocabulary or index language of that system.

Definition of Vocabulary Control :- Vocabulary control means the control over the vocabulary, i.e., over the terms which are used in any indexing system in information retrieval system. It may be defined as a list of terms showing their relationships and used to represent the specific subject of a document.

Indexing Language :- Indexing language is the language which is used to describe a subject or other aspects of information in an index or in a library catalogue. The indexing language can be alphabetical language or may be a classification scheme.

UNIT - III

Questions Bank

- Q1. Define Thesaurus.
- Q2. Write the example of Thesaurus.
- Q3. Describe the types of Thesaurus.
- Q4. Write the purpose of Thesaurus.
- Q5. Define Library classification Schedule.
- Q6 Describe the Purpose/Objective/Functions of classification Schedule.
- Q7 Write the example of Classification Schedule.
- Q8. Define Subject Heading.
- Q9. Describe the Purpose/Objective/Functions of Subject Heading.
- Q10. Write any five benefits of Subject heading.

Q11. Write the example of Subject heading list.

Q12. Define the Vocabulary control.

Q13. Describe the methods of Vocabulary control.

Q14. Write the purpose of Vocabulary control.

UNIT-IV

Indexing and Abstracting

4.1 Indexing: Definition, Characteristics, Types of Indexes

Indexing:- The word index is derived from the Latin word indicare which means to indicate or to point out. An index is a systematically arranged list giving enough information about each item to enable it to be a link between a specific subject or topic and the identified document.

Definition of Indexing:- 'B.Guha'. Indexing is a method of document representation, it naturally follows that indexing has to represent or exactly specify the content of documents.

Types of Indexing:- The main objective behind preparing an index is to satisfy the different approaches of the users.

1) Author Index:- An index prepared for satisfying the author approach is called the author index. The name of the authors are arranged alphabetically to facilitate easy search of the documents.

2) Affiliation index:- This type of index plays an important role as it adds the value of an article as the details about the author guide us to know about the present status in an institution, which proves his work more authentic. Thus this affiliation indexes are more useful and important for the scientific work.

3) Subject Index:- A subject index mainly concerns itself with the subject approach of the users, which is one of the most important indexes as subject approach is the common approach of the scholars. It is always easy for a reader. Subject indexes are of the types (i) General subject index, (ii) Specific subject index.

Examples: (a) Social science index, New York : Wilson, 1947.

(b) Applied science and technology index, New York: Wilson, 1913.

4) Key Words Index:- These terms represent subject of the document. An index based on these keywords of a document is termed as keyword index. These index has become a popular means of document representation and a number of techniques based on these indexes have developed. They are PRECIS, POPSI, KWIC, KWOC, Uniterm etc.

5) Other Indexes:- Besides the above indexes, there are also some special indexes to satisfy some of the specific subject approach of the users.

Examples:- (a) Formula index (Chemical Abstracts).

(b) Locality index (British Geological Literature).

(c) Periodical index (Scientific American cumulative Index).

Characteristics of Indexing

- 1) They use controlled vocabularies, like thesauri and classification schemes.
- 2) Subject heading are used for key themes & concepts.
- 3) Multiple access points.
- 4) Cross reference: They provide links between related heading for information retrieval.
- 5) Hierarchical structure for the subject heading.
- 6) Quick browsing & easy access.
- 7) Increase resource utilization.

Salient features of a good Indexing System

1) Recall :- Recall is the retrieval of documents by the system relevant document is known as the recall power of the system. It can be define by the formula.

$$R = \frac{\text{No of items retrieved and relevant}}{\text{Total items relevant in collection}} \times 100$$

$$= \frac{5}{10} * 100 = 50\%$$

Thus the system has achieved a 50% recall value.

2) Precision :- The precision is a measure of irrelevance.

Precision Ratio :

$$P = \frac{\text{No of items retrieved and relevant}}{\text{Total No items retrieved in collection}} \times 100$$

$$= \frac{7}{10} \times 100 = 70\%$$

Precision ratio = 70%

3) High level of exhaustively of indexing tends to ensure high recall.

4) Index language to describe the topic precession.

5) An cant or new information for the users.

6) Cost- effeteness:- The cost effectiveness include the cost every relevant items and the cost per new relevant item.

4.2 Outlines of Indexing System: Pre – Coordinate Indexing System: Chain Indexing, PRECIS, POPSI

Difference between pre-coordination indexing system and post-coordinate indexing system.

	pre-coordination indexing system	post-coordinate indexing system
1.	It is conventional system mostly found in printed indexes.	This system is non – conventional system.
2.	In this system coordination of terms in the index is decided before the particular request.	The word and phrases, terms selected at the time of indexing at the time of searching the index.
3.	Chain indexing, PRECLS, POPSI are the method of pre indexing system. In this type of system a document is represented in the index by pre-defined heading.	KWIC and KWOC are the type of post-coordinate indexing system.
4.	Leading term determines the position of entry in the index and other terms are subordinate it.	This system allows dynamic terms combination.
5.	All pre-coordinate indexing systems are based on classificatory principles such as facet analysis.	It is more Flexible system.
6.	Coordination at indexing stage.	Co-ordination at retrieval stage.
7.	Controlled vocabulary.	More flexible vocabulary.
8.	Time consuming to update.	Easier to update.
9.	Can be more precise due to controlled vocabulary.	Can be less precise due to term ambiguity.
10.	Lower recall	Higher recall.

Chain Indexing:- One of the most important development is subject indexing is the evolution of chain procedure. It was designed by Dr. Ranganathan 1938. Chain of a class number is translated into its verbal equivalents to formulate a subject heading that presents the subject

contents of the document. As this chain is used for deriving subject entries on the base of a set of rules & procedures. This new system user called chain procedure. Its main purpose for retrieving the document form such a classification, the ingenious method of chain indexing to create subject index entries form the chain. Thus chain indexing is a semi automatic method of deriving subject entries from the chain of successive sub division.

Chain Procedure :- One of the most important development in subject indexing is the evolution of chain procedure.

Definition:- On the basis of a definite set of rules and procedures simple, compound and complex subjects are handled.

Merits of chain Indexing :-

- 1) Chain indexing is a technique used in information retrieval.
- 2) Saves time & effort.
- 3) It provide multiple access points for users searching for information.
- 4) It is a systematic & mechanical method of deriving entries.
- 5) It increase pinpointed retrieval efficiency.

Limitation

- 1) It may not be suitable for all types of information & retrieval system.
- 2) It depends open the classification scheme.

PRECIS :- Preserved context Indexing System

It was designed & developed by Derik Austin by about 1970 for the purpose of deriving subject heading and subject entries.

PRECIS is based on principles of context dependency. PRECIS was the first computerized pre-coordinating system.

Various steps in PRECIS

- 1) Content analysis of a documents.
- 2) Select appropriate indexing terms.
- 3) Alphabetical arrangement of entries.

Merits of PRECIS

- 1) PRECIS is now a recognised indexing model.
- 2) It is also based on the principle of classification. But specially designed to be amenable to computer main pulsation.
- 3) High precision.
- 4) Flexibility.
- 5) User friendliness.

POPSI Indexing Syastem :- A powerful tool for subject Indexing.

POPSI, or postulate – based permuted subject indexing, is a pre-coordinate indexing system developed by G. Bhattacharyya. It's designed to efficiently index and retrieve information based on its key features:

Principles:-

- **Analytic – synthetic methods:** POPSI breaks down complex subject into their fundamental concepts (analytic) and them builds subject strings by recombining these concepts (synthetic).

- **Ranganathan's postulates:** The system is grounded in S.R . Ranganathan's five fundamental postulates of library classification subject representation.
- **Elementary categories (ECs):** POPSI uses a set of predefined categories like entity (E) Relation (R), Attribute (P), Action (A), Space (S), and TIME (T) to structure subject terms.

Components :

- **Classaurus:-** A controlled vocabulary that ensures consistency and eliminates ambiguity in terminology.
- **Indexing procedure:-** Involves identifying key concepts, assigning ECs, formulating subject strings, and generating index entries.
- **Permutation:** Each concept in the subject string become an "approach term." Resulting in multiple index entries for different access points.

Advantages:

- **Flexibility:** Handles complex subjects and retrieves information based on various search terms and combination.
- **Specificity:** Provides precise subject representation and retrieval compared to keyword indexing.
- **Theoretical foundation:** Rooted in established principles of library classification, ensuring logical structure.

Disadvantages:

- **Complexity:** Requires training and skill to apply effectively.
- **Time – consuming:** Indexing process can be more involved compared to simpler system.
- **Software dependence:** Often relies on specialized software for efficient implementation.

Applications:

- Libraries and information centres.
- Digital libraries and repositories.
- Bibliographic databases.
- Subject – specific information systems.

Overall, POPSI is a powerful indexing system that offers and flexible subject representation and retrieval. While it requires some expertise to use, its advantages outweigh the drawbacks for organizations seeking advanced information organization capabilities.

4.3 Post – Coordinate Indexing System: KWIC and KWOC etc.

KWOC Full form:- Keyword out of context.

It is type of automatic indexing system used to provide access to text based information. It is invented by Hans Peter Luhn in 1958.

How it works

1) Keyword Extraction (Excluding the words a, an, the)

2) Context preservation

3) Alphabetical Arrangement of keywords.

There are two main formats for displacing KWOC entries:

1) Single Line Format

2) Two Line Format

Example

Title : The impact of artificial intelligence on education

KWOC INDEX

1) Artificial Intelligence:- The impact of artificial intelligence on education.

2) Education:- The impact of artificial intelligence on education.

3) Impact:- The impact of artificial intelligence on education.

Advantages of KWOC indexing

1) KWOC indexes are useful for searching scientific literature, news articles and other text collections.

2) To find relevant documents.

3) Simple & efficient technique.

4) It provide context around the keywords.

5) Easy to use.

Disadvantages of KWOC

1) Limited control over keywords.

2) Can be bulky for large datasets.

Overall KWOC is a valuable tool for libraries, researchers who needs to navigate large collections of documents.

Difference between KWIC & KWOC indexing method.

	KWIC	KWOC
1.	Keyword in context.	Key out of context.
2.	Each keyword appears as the central entry on a line (Presentation)	Presentation:- Each keyword appears on a separate line
3.	Benefit:- Providing clarity & understanding to users.	Benefit:- More spacer efficient than KWIC for long titles.

4.	Drawback:- Requires more space than KWOC	Drawback:- Lacks immediate context around the keyword. (read the entire text)
5.	Keyword placement:- Center of each line	Keyword placement:- Separate line.
6.	Context:- Surrounding words displayed.	Context:- Entire text follows.
7.	Space Efficiency:- Less efficient	Space Efficiency:- More efficient.
8.	Clarity:- High.	Clarity:- Lower.

4.4 Defination of Abstract :- “An Abstract is a summary of a publication or article”.

Quality of a good Abstract:-

- 1) Complete, Concise, Clear, accurate information about the artical.
- 2) It should present new data.
- 3) It should be as brief as possible.
- 4) It should include author’s own view.
- 5) Good orienting & Compactness.

An Abstract is an abbreviated, accurate representation of the original document. It is a sufficient basis for the reader to decide whether or not to refer to the original. The abstracts are arranged in classified order.

Need/Importance/Purpose/ of Abstracts

- 1) Provide an overview to the readers.
- 2) To help the users to understand the main content.
- 3) To help reader decide whether the complete article is relevant to their needs.
- 4) To provide concise & condensed knowledge about the article.
- 5) Summarize key points.
- 6) Indentify relevant information.
- 7) Abstract save reading time.
- 8) Enhance search results.
- 9) Facilitate communication when people form different disciplines are working together.
- 10) To help the readers to develop their own understanding & Critical thinking of the topic.
- 11) Assess quality: Abstract can be used to asses the quality of a work- By reading the abstract, readers can get a sense of the author’s research method, finding.
- 12) Enhance learning efficiency.
- 13) Decision making tool.

- 14) Search & retrieval keys in databases & library.
- 15) Abstracts facilitate Literature searches.
- 16) Abstracts help overcome the language barrier.
- 17) Abstracts improve indexing efficiency.

Difference Between an Index and an Abstract :- The abstract provides a clue to the relevance of the material. An index only gives a key to where the material is located.

Elements of the Abstracts

- 1) Purpose : Why & What of the study.
- 2) Scope :Extent of the work.
- 3) Method : Techniques.
- 4) Specialized Content.
- 5) Results
- 6) Conclusions

Types of Abstractor :-

- 1) Author Prepared abstract.
- 2) Abstracts prepared by the subject expert.
- 3) Abstracts prepared by the professional abstractors.

Types of Abstract :-

Abstracts are various types depending on the amount of information.

1) Indicative Abstracts :- Indicative abstract is often little more than a broad statement of content, an extension of the title.

Example :- Such abstract are thus often used in information retrieval system, e.g., Chemical abstracts, Biological abstracting.

2) Information Abstracts :- Informative abstract summarizes all relevant arguments, data, methods, finding and conclusions of the original.

Example :- Bibliographies, Handbooks.

3) Listing Abstracts :- Listing abstracts list the contents of the documents conference proceedings.

4) Critical Abstract :- In a critical abstract, the abstractor functions as an evaluator. For critical or evaluative abstracts, the abstractor injects his opinions and analysis. Extra care has to be taken in preparing critical abstracts.

Example :- Critical Abstracts are found in Applied mechanics Review.

5) The Title or Expanded Title :- A good title is often sufficient to indicate. Whether the paper is of interest.

6) The Summary or Synopsis :- To give an outline of the whole of the work described. Many journals require authors to write a summary of their papers. These summaries are sometimes republished in certain abstracting periodicals.

7) Abstracting With A Slant :- Abstracts are usually prepared keeping in view the purpose of the users. The area of research, and depth of information required.

8) Graphic Abstracts :- In the field of chemistry where a precise language of communication is available, a new type of abstract has been introduced by the current Abstracts of chemistry and Index Chemicus, making full use of this language. A chemist finds it easier to get the essence from such an abstract. Indexing becomes easier.

9) Telegraphic Abstracts :- Telegraphic abstract is a detailed index to a graphic record which is composed of (i) significant words selected from the documents; (ii) Code symbols called role indicators which supply a context for the selected words; (iii) Punctuation symbols.

10) Statistical or Tabular Abstracts :- The abstract is a summary of the data presented in tabular form.

11) Auto – Abstracts :- The use of a computer to extract abstracts from the text of the document.

12) Extracts :- Key sentences are selected excerpted; and arranged in a logical order to present the reader with the thesis its substantiating information, and the tone or style of the document, charts, diagrams, tables, and other graphic aids may be used.

UNIT – IV

Short Answer Type Question

Q.1. The word index is derived from Latino word _____

Q.2. Write the type of index.

Q.3. Write any two merits of chain indexing.

Q.4. Write the full form of PRECIS.

Q.5.

Q.6. Write any one limitation of chain indexing.

Q.7. KWIC stands for _____

Q.8. KWOC stands for _____

Q.9. Define abstract.

Q.10. Write the difference between index & abstract.

Long Answer Type Question

Q.11. Define indexing.

Q.12. Explain the type of indexing.

Q.13. Describe the salient feature of indexing system.

Q.14. Differentiate between pre-coordinate & post-coordinate indexing system.

Q.15. Describe the chain indexing with its merits.

Q.16. Explain the KWOC indexing methods in brief.

Q.17. Explain the KWIC indexing system.

Q.18. Differentiate between KWIC & KWOC indexing methods.

Q.19. Write the quality of good abstract.

Q.20. Write the need & purpose of Abstract.

Q.21. Describe the types of abstracts.

UNIT- V

Trends in Information Retrieval

5.1 Web Information Retrieval – Search Engines: Definition

Search Engines Definition :- A search engine is a powerful tool that helps & navigate the vasts of information available online. Its like a digital librarian who tirelessly scours () the world wide web to find what you’ are looking for.

Example: Google, Bing, Yahoo, Duck Duck Go.

Working process of search engine.

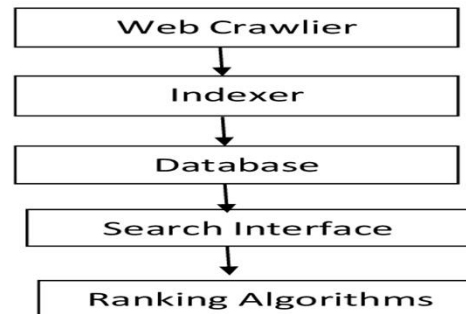


5.2 Functions and Components

Functions of Search Engines:

- 1) Finds information:- It search for web pages, images, videos, titles that match your search query.
- 2) Sorts Results:- It uses complex algorithms to rank the result based on their relevance, quality and other factors. This ensures to get the most helpful information first.
- 3) Presents Results:- It displays the retrieved information in a clear and organised manner.

Components of Search Engines:



5.3 Meta Search Engines, Subject Directories

Definition of Mata Search Engines:- A meta search engine is a powerful information retrieval tool that aggregates results from multiple search engines into one unified list.

Key Characteristics of Meta Search Engines

- 1) Combine results from multiple search engines. It act a super – search engines from various sources.
- 2) Provide a single point of access.

- 3) It can combine results into a single set.
- 4) It saves time & effort.
- 5) Brooder coverage.
- 6) Convenience.
- 7) Comprehensive research.
- 8) Comparing results from different sources
- 9) Aggregate results.
- 10 Ranks & Filters results.

Popular Meta Search Engines

- 1) Dog pile
- 2) Meta Ger
- 3) Zap Meta

Difference between subject directories search engines

Subject directories and search engines are both tools for finding information online, but they work in different ways. Search engines use algorithms to crawl and index the web while subject directories rely on human editors to categorize websites. This means that subject directories can be more accurate for specific topics, while search engines can be more comprehensive for general searches.

5.4 Subject Gateways and

Definition definition of subject Gateway:- A subject gateway is an online service that provides organized, curated access to high quality, relevant information resources with a specific subject area. It serves as specialized portal or gateway to navigate the vast information landscape on the internet.

Key features of subject gateways:-

- **Subject Specificity:-** They focus on a well-defined subject area, ensuring the relevance of the resources they catalog.
- **Resources selection and Quality:-** Resources are carefully selected and evaluated by experts in the field, ensuring their quality and credibility.
- **Organization and structure:-** Resources are organized in a structure way, often using controlled in a vocabularies and classification systems, enabling efficient browsing and searching.
- **Search and browsing capabilities:-** Users can search for specific information using keywords or browse resources by subject categories.

Value-Added Features:- May offer additional feature like:

- Resource descriptions and annotations.

- Links to related resources.
- News and updates in the field.
- Discussion forums or communities.

Benefits of subject Gateway:

- **Improved Resource Discovery:-** Facilitate finding relevant, high-quality information within a subject area, saving time and effort.
- **Subject Expertise:-** Leverage the knowledge of experts who select and evaluate resources, ensuring their trustworthiness.
- **Organization and Structure:-** Provide a structured way to navigate information, enhancing browsing and discovery.
- **Filtering and Quality Control:-** Help filter out irrelevant or low-quality information, ensuring users find reliable sources.

Common examples of subject gateways:

- INFOPORT (INFLIBNET subject Gateway for Indian Electronic Resources)
- Intute (formerly a UK-based gateway for education and research)
- BioMed Central (gateway for life sciences research)
- Science.gov (gateway for science, technology, and engineering research)
- ERIC (Education Resources Information Centre)
- PsycINFO (gateway for psychology research)
- Social Care Online (gateway for social care resources)

Subject Directories

- Organized collections of websites, categorized by topic or subject matter.
- Curated by humans, often experts in the field, ensuring quality and relevance.
- Offer a structured way to browse and explore information making them ideal for research and learning.
- Contrast with search engines, which index billions of pages using algorithms.

Benefits:-

- **Quality control:** Human creation ensures websites are accurate and reliable.
- **Focused exploration:** Narrow down your search to specific topics or fields.
- **Discovery:** Uncover unexpected resources you might not find through search engines.
- **Expert guidance:** Benefit from the insights of subject matter experts.

Common types:

- General directories: Cover a wide range of topics, e.g., Yahoo! Directory (defunct), DMOZ (open directory project).
- Specialized directories: Focus on specific fields, e.g.,:
 - Education: Academic Info
 - Health: Medline Plus
 - Law: Find Law
 - Business: Business.com
 - Science: Science Direct

- Literature: The WWW Virtual Library

How to use subject directories:

1. **Identify relevant directories:** Choose directories aligned with your interests or research needs.
2. **Browse categories:** Explore the hierarchical structure to find related websites.
3. **Utilize search features:** If available, use keyword searches to refine results.
4. **Evaluate websites:** Assess credibility and Quality before using information.

Image examples:

- Subject directory
- Search Engine
- Deep Web Search

Key Features:

- Hierarchical structure: Websites arranged categories and subcategories, like a tree branching out.
- Annotations: Brief descriptions of websites, helping users determine relevance.
- Search functionality: Some directories allow keyword searches within their catalog.

5.5 Definition of Institutional Repository :- An Institutional Repository is a digital archive maintained by an Institution typically a university, college and research centre.

Purpose: To collect, preserve and provide open access to the intellectual output of its community, it serves as a central platform and manage the institution's scholarly work, research data, educational materials and other form of intellectual property.

Functions of IR

1) Collecting:- Gathers wide range of scholarly and institutional materials, Including:

- Journal articles.
- Conference papers.
- Research reports.
- Theses & dissertations.
- Working papers.
- Books.
- Data sets.

Multimedia content (Images, audio, videos.)

Educational resources (Course materials, syllabi, lectures)

Administrative documents (Reports, policies guidelines)

2) Preserving:- Ensures long term accessibility and integrity of digital contents.

3) Disseminating:- Provides open access to deposited materials a making them freely available to:

- Institutional community members.
- Global research community
- General Public
- Online search interfaces
- Metadata standards (Dublin core)
- Integration with search engines and subject-specific repositories.

4) Managing:- Organizes and maintains the repository through:

- Metadata application (descriptive information about items)
- Access control (setting permissions for different)
- Version control (tracking changes to items)
- Usage statistics

Benefits of IRS:

- Increased visibility & Impact of research
- Preaervation of institutional knowledge
- Enhanced collaboration and knowledge sharing
- Improved access to research for both institutional and external users.
- Support for research management

Example of IR software

1) D-Space

2) E-Prints

3) Digital Commons

Functions of Institutional Repositories in IR

1) Collecting and organizing research output.

- To gather a wide range of scholarly material provided by the institutions researchers.
- Organizes them using metadata (descriptive information) for easy retrieval.

2) Providing open access:- online access easy lie and removing barrier to knowledge sharing.

- Increases Usability and impact of research.

3) Preserving Institutional knowledge:- Long term preservation of digital content.

4) User tiredly search interfaces & Browsing tools.

5) Integrates with discovery services and search engines.

6) Enables users to find relevant research efficiently.

7) Increases exposure of research outputs to a wider user.

8) To promote collaboration & knowledge sharing → connects researchers within and across institutions.

9) Enabling research assessment:-

10) Supporting research management.

11) Preserving institutional History.

UNIT - V

Short answerer type question

Q.1. Define search engine.

Q.2. Difference between subject directories & search engine.

Q.3. Write the name of Meta search engine.

Q.4. Write the example of information retrieval software.

Long answerer type question

Q.5. Write the function of search engines.

Q.6. Describe the components of search engine.

Q.7. Describe Meta search engine.

Q.8. Define Institute repository. Write its purpose & functions.

