# UNIT 14 ISAR SYSTEMS: OBJECTIVES, TYPES, OPERATIONS AND DESIGN

**Structure**



## 14.0 OBJECTIVES

In the previous Units of this course we have discussed the basics of information processing and organization. All the areas such as indexing, bibliographic description and formats, vocabulary aspects etc. are the sub-systems of Information Storage and Retrieval System (ISAR). Now we will be taking a holistic view of the total bibliographic information storage and retrieval system. In this Unit, the objectives, types and design aspects of ISAR system will be discussed.

After reading this Unit, you should be able to:

- define an Information Storage and Retrieval System;
- know different type of Information Storage and Retrieval System (ISAR);
- understand knowledge representation aspects of ISAR system designed according to information requirements;
- explain the components of ISAR system; and
- appreciate uses of ISAR system in libraries and information centres.

## 14.1 INTRODUCTION

Information Storage and Retrieval (ISAR) system deals with three basic aspects:

- Information representation
- Information storage and organisation
- Information access.

One of the best examples of ISAR system is library system, where information is stored, processed organised and retrieved on demand. Information could be stored in a book, audio-video, images and so on. The library and information centres endeavours to organize knowledge available in documentary form. The multi-faceted universe of knowledge is represented in libraries in linear form using some classification scheme. At the time of retrieval, specific aids like cataloguing and indexing are used and meaningful information is retrieved.

A lot of emphasis is given on improving the performance of the system. For that librarians have developed classification schemes for helpful arrangement of documents on the shelf so that retrieval can be facilitated. The tools like library catalogue or indexes are developed and further modified to satisfy the different approaches of users toward information. Automated systems reduce the effective time of users in searching for information which, in effect, further improves the performance of the system. Therefore, tools like Database Management System (DBMS) are used for keeping records of a holding of library, which is known as Online Public Access Catalogue (OPAC).

## 14.2 USERS AND THEIR INFORMATION NEEDS

Several classes of users consume different kinds of information for different purposes. A user when searches for information, may want either advanced level of information or semi-expert level of information or preliminary level of information according to his information needs. For example, a physician may want advanced level of information on medicine in his area of specialization. But the same person may want preliminary level of information on legal matters and semi-expert level of information on biomedical instruments. So the information needs of a user differ significantly in different situations.

In general, the 'Information Gathering Habit' surveys reveal that there are at least four types of information needs of the users. These may be categorised as:

a) Current Information Need,

b) Exhaustive Information Need,

c) Every day Information Need, and

d) Catching-up or Brushing-up Information Need

You have already learnt about these information needs in Unit 13 of course MLII-101. An ISAR system is designed to meet these types of information needs. Recognition of actual information needs and also potential information needs of the users is a problem faced by the ISAR system developers, as it has been recognized that the 'need' and 'want' do not connote the same meaning. Users may also have different approaches towards of a library. A user might make an approach by:

- name of author
- title
- subject

The system, thus, should have provision of field level search, not only for browsing facility according to author, but also, title, subject and so on. This is quite an imaginative job from the point of view of a system designer. In a typical library automation system, the front-end should have facilities of generating different kind of reports of library usage.

But most commercial ISAR systems cater to either exhaustive information needs or current information needs of the users. The different classes of users can be professionals, students, researchers, policy makers, managers, and common citizens. Many times such information may be available in sources other than the documentary sources. For every profession and every vocation, there is demand of information to successfully carry out professional or vocational work. For example, a farmer may require weather-related information as well as information on best farm practices. A journalist may want background information for his current assignments, but not necessarily from the documentary information sources.

The ISAR system is designed to meet varieties of needs of different classes of users. Nowadays ISAR system is designed on modular basis, which has different components or subsystems. The user interface is designed in a friendly manner (user friendliness), so that ease of use can be ensured. An ISAR system usually offers different approaches of searching, which may be achieved through strong indexing capabilities and facilities. Users also have choice to search through simple search, or advanced search mechanisms. In simple search, user enters a single search parameter such as title of document or keyword or name of author. In advanced search, user enters a combination of search parameters, such as, more than one keyword, or a keyword and a name of author, or, a keyword and year of publication, etc. In advanced search, Boolean operators, like AND, OR, NOT, SAME, etc. can be incorporated to refine a query. The system which is to be developed should have all these capabilities to search and retrieve information.

## 14.3 OBJECTIVES OF ISAR SYSTEMS

The principle objective of ISAR system is to provide correct information to the user in least time with least efforts. Thus, while designing any ISAR a system designer should keep following objectives in view:

### Information Facilitator

The ISAR system should act as facilitator between the information (contained in document) and the users. If a user approaches with the subject term, name of contributors or title of the document and so on, the system should be helpful to give him the desired information. The information could be exact information or the reference of a document which contains information.

### Non-Ambiguous

The system should be so organized that ambiguity of information is avoided so that search result is free from any kind of ambiguity. This requires identification of terms, setting their context and their proper indexing. For example, search for a term 'screw driver' should not bring results like 'truck driver', 'hardware driver' and so on.

### Minimum Time

The system should be so designed that minimum effort and time are spent to interrogate the system.

Searching through the system should take minimum time, meaning thereby that the ISAR should be capable of performing fast search. Not only that, it is best

to have an online ISAR so that users do not need to walk to library. They should get whatever they want at there work place.

### User Friendliness

Ease of use is an important consideration for any ISAR system.
Any ISAR should have user friendly interface. The important aspects of ISAR should be highlighted. Before a user uses the system he/she should be properly introduced to the system with all its features, i.e., informing users about the scope of system, available search options, and most importantly how to perform search with the system. It is only this interface through which a user operates an ISAR system. Take an example of a Library OPAC. It should have following features:

- Introduction to library
- Scope of collection
- Instructions for performing search

The search interface should facilitate framing the search like,

- Keyword search
- Author and title search
- Combination search (using Boolean operators)
- Proximity search, etc.


**Fig. 14.1: Web OPAC of Library of Congress**

While designing an ISAR system the following aspects should be recognised to achieve the objectives of the system:

- The desirability of making systems as readily usable as possible for their clienteles;
- The need to recognise basic features of retrieval system; and

- To incorporate coordinating features such as vocabulary control, search strategies, user-interface, information modelling aspects in general, etc.

**Features of ISAR System**

Keeping in view the objectives mentioned above and recognizing the aspects to be considered in designing a system, an ideal ISAR system should incorporate one or more of the following features:

- The competence and compatibility for consolidated searching and retrieval of information from any client terminal from any database within the system.
- It should be able to narrowcast or broadcast or relate the information need in a variety of associations to get optimum retrieval performance.
- It should have access facilities at multi-points.
- It should have common command language facility to retrieve information from several databases of the system.
- It should be able to handle information access from entity-related or object-oriented approaches. It may also provide all other associations for accessing information.
- In a bibliographic or full-text database, the surrogates chosen should have indicative as well as informative features that are sufficient enough to select or reject the retrieving information based on end-users' needs.
- It should have the ability to select, classify, process and consolidate the analysed information into a cohesive text ready for assimilation by the end-users.
- It should have ability to orient the information to specialist needs of the users from time to time. This calls for understanding the processing of user profiles.
- It should be able to retrieve maximum information with minimum number of clues. The fuzzy approaches of end-users must be able to get clarified and ultimate result should provide satisfaction to the searcher.
- It should have capacity to interchange the information available in one database or another for purposes of retrieval relevance end usage.
- It should have bibliographic data interchange capacity (using Z39.50 or similar standard) to meet consolidation to a chosen format for networking and other purposes. Compatibility with standards at all levels must be the goal.
- It should have ability to search simple information quickly in an easy manner and also have the ability to multi-track the complex questions and present them in a simple easy manner. User-friendly presentations are very important.

## 14.4 TYPES OF ISAR SYSTEMS

ISAR systems are used by a wide range of users. According to different kinds of needs and purpose of use, different types of automated systems may be designed. Such types may be:

- Database Management System (DBMS)
- Text Retrieval System
- Management Information System (MIS)
- Decision Support System (DSS)
- Knowledge Based System (KBS)

### 14.4.1 Database Management System (DBMS)

Any automated system is based on a collection of stored information or documents

in a database which is amenable for access. A DBMS is primarily concerned with data storage, maintenance and retrieval and is used to keep control and manipulate data within the database.

The distinguishing characteristic in DBMS is the definite structure of the stored information, instead of dealing with natural language text. In DBMS, normally files of data are described by a small set of pre-specified attributes. For example, in a Salary database of an organization, name of person, designation, salary, etc. are attributes. Similarly, in the context of records of books author, title, publisher, year etc., may be the attributes. Each attribute carries some kind of value in it. Therefore, a DBMS can be defined as set of records and each record contains fields (attributes) which in turn contain data (value). A database may contain textual, numeric, statistical and graphical information.

OPAC is a kind of DBMS often built of some kind of Bibliographic Database Management Systems (BDBMS). The typical example of BDBMS is the one built by CDS-ISIS/WINISIS developed by UNESCO. The data fields may contain author, title, place, publisher, year of publication and so on.

### 14.4.2 Text Retrieval System

In contrast to DBMS, text retrieval systems are designed for unstructured data such as full text documents. Queries are usually language based here such as, keywords and a number of advanced search techniques (such as proximity search) can be used. However, systems may also handle discrete structured data.

### 14.4.3 Management Information System (MIS)

Management Information System is a kind of database management system designed to cope up with the needs of managements who need to have information about different alternatives related to his/her interest to facilitate his work. Though built on DBMS platform, information are subjected to special processing. In such a system information is available with different alternatives. In the business environment managers needs to take complex and rapid decisions. Under such circumstance MIS is a kind of sophisticated tool which provides them timely information to take decisions.

Collection of data is critical in MIS because the information comes from different sources, i.e., within the organization or from outside organization. Collection of data not only needs defining that how the data would be captured but also the estimation of cost involved in data collection. Once the data are collected and organized such system generates reports for usage. The reports could be generated in printed or electronic form depending on desired format. Such systems also generate reports upon different intervals if it is desired.

### 14.4.4 Decision Support System (DSS)

Decision Support systems help top-level management in arriving at decisions. There is very little difference between MIS and DSS. The former generates reports in anticipation or on demand and collect facts, whereas the later provides possible alternative solutions. These are interactive computer-based systems that provide the user easy access to decision models and data in order to support semi-structured and unstructured decisions. In management parlance a structured decision means use of rules and norms for making decision. Such systems help managers in identifying the problem, analyzing alternatives and choosing possible solution. However, the typical decision-making takes place with shared effort of human and machine. In true sense, a DSS cannot take decision, rather it amplifies decision maker's capability by providing resources and facilities. In other words it provides intellectual support. These systems are integrated with powerful tools like generating charts, preparing tables and presentation tools.

### 14.4.5 Knowledge Based System (KBS)

Specialised computer programs, modeled in the same way as human experts tackle problems and arrive at solutions, are called 'Expert Systems'. Such systems rely upon a store of specialised knowledge for solving problems and hence referred to as Knowledge Based Computer Systems (KBCS) or Knowledge Based systems (KBS).

Expert systems are sophisticated computer programs that manipulate knowledge to solve problems efficiently in a narrow problem area. Knowledge based systems enhance the value of expert knowledge by making it readily and widely accessible. Like human experts, these systems use symbolic logic and heuristics to find solutions. They are also capable of learning from experience through inferencing mechanism.

KBS systems are domain specific and are backed up by a strong knowledge base. In these systems each bit of information is not only stored but they are also linked. This linking is used to preserve the context. The context is used to draw the inference from a query. They are capable of providing solutions and replace the human intervention. Expert systems are thus data driven. But it is also important that how the given information is utilized to achieve the goal. Human experts possess procedural knowledge (How To), which helps them to flowchart the courses of action to be taken in solving problems. Accordingly, rules for manipulation of the knowledge have to be incorporated in the expert systems to get the desirable results. However, this procedure involves other related tasks such as intuitive inferencing power, learning and updation of knowledge. Solutions to problems can be achieved if only all tasks are executed in a coordinated way.

**Self Check Exercises**

1) What are the objectives which should be kept in mind while designing an ISAR?

2) Name different types of ISAR systems.

**Note:** i) Write your answers in the space given below.
ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

## 14.5 DESIGN OF ISAR SYSTEMS

From the foregoing discussion, it should be clear that a number of features should be ingrained in an ISAR system. Thus, an ISAR system is an integrated one combining various aspects.

### 14.5.1 Components of An ISAR System

An ISAR system comprises of following components:

- User Interface
- Knowledge Base
- Search Agent

The features to be incorporated in each components are discussed below :

#### 14.5.1.1 User Interface

User interface is the part which puts users across the ISAR system. It is the front-end which enables user to put a query and displays results. Basically, user interface is of two types:

- Query Interface
- Result Interface

##### 14.5.1.1.1 Query Interface

This is the end from where users enter his/her search terms. It is one of the major components which initiate communication between users and the system. The Query Interface should have following features:

a) *Understanding the user input statement*

It is also known as front-end. The interface should be able to capture keywords given by users which should be passed on to the search program. The front-end should have understandable look and feel. At the time of designing the system, one should also seriously consider use of different colours, and instructions for performing search and limitation of search.

b) *Refining the problem statement*

Sometimes users start with broader domain and further refine the search. The interface should have flexibility for further refining any query within the displayed search results. In other words, system should provide facilities for further modification of search statement. It should also display some kind of arrangement among topical terms which further facilitate browsing through the system.

c) *Search statement to search strategy translation*

Any knowledge base accepts a query in a particular format. For example, a Relational Database Management System (RDBMS) accepts search statement in Structured Query Language (SQL). It is the system front-end which translates a search statement and formulates a search strategy in the language which is understood by Search Agent.

d) *Modification of search strategy*

If one does not get desired output from the database, ISAR system should have procedure for further modification of search strategy. The modification should be interactive. Vocabulary control devices can also be added as an aid for users to locate the term of his/her interest.

##### 14.5.1.1.2 Result Interface

Display of search results is another important aspect of searching. It should be in user friendly manner. Not only that the result should cater the needs of individual users but the display should also be customized. Search results should also display the ratings in the light of search terms. For this purpose statistical techniques can be used.

#### 14.5.1.2 Knowledge Base

The store house of any ISAR system is its Knowledge Base. It contains list of facts or related facts (information). Any kind of query is answered based on the facts stored in the Knowledge Base. A Knowledge Base could be a Database Management System (DBMS). Retrieval of information from storage depends on two important aspects of Knowledge Base:

- Knowledge Representation
- Indexing and clustering

#### 14.5.1.2.1 Knowledge Representation

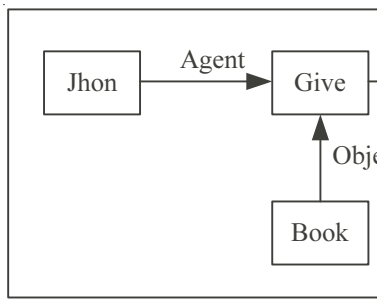
The first and foremost objective in constructing an ISAR system is representation of facts within the Knowledge Base. There are different ways of representation of knowledge:

***Semantic Network Knowledge Representation***

Semantic network is a method of knowledge representation based on a network structure. A semantic network contains points called *nodes* connected by links called as *arcs*. The nodes represent objects, concepts or events - in other words documents or information. The arcs are used to represent the relations between the nodes. Arcs build a kind of hierarchies in the Knowledge Base. Arcs usually represent relations like *is_a* or *has_part.* For example, Universe of Knowledge à Library Science à Cataloguing

Semantic networks are a useful way to represent knowledge in domains that use well established taxonomies to simplify problem solving. Semantic networks are useful in representation of sentences of natural language.

For example, a sentence *John gave a book to Richard* can be represented as,




***Frame Based Knowledge Representation***

It is an object-oriented approach. A frame represents an object (document or information) or class of objects (collection of documents or information) or several facts. When they represent a class of objects, they generalise certain groups identifying overall properties of those groups, it shares. The pointers where properties are stored are known as *slots*. Similarly, if frame represents an object, slots represent the properties or attributes of the object. Slots contain value for that particular attribute. For example, a book in a library is an object, therefore it can be represented as *frame*. The properties of book, i.e., Title, Author, Place, Publisher and so on are stored as slots and each slot would have corresponding value.

Frames are also very helpful in representing hierarchies. In a frame base, slots can store procedures or relations. Relations are used for storing taxonomy or genealogical data or parent child information. For example, if a person is represented by a frame, different slots can store value who is the father or who is the son of the person. Similarly, slots can also store procedures that means how a frame or object should react or the way it should be used or it can operate with other objects.

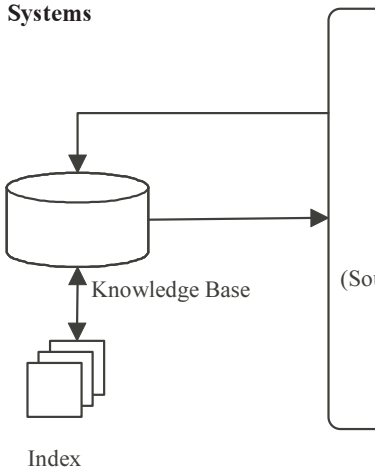



**Fig. 14.2: Diagrammatical representation of an information storage and retrieval system**

### *Rule-Based Knowledge Representation*

Rule based representation is a popular approach. Rules are employed to state the way in which the inferencing has to be done. Rules provide a formal way of representing recommendations, directives, or strategies. Rules are appropriate when the domain knowledge results from empirical associations developed through years of experience in solving problems in a given area. Rules are expressed in the form of IF-THEN statements. For example,

| | |
|---|---|
| If search is in collection of BOOKS | THEN display Title, Author, Place, Publisher, Year, Physical Description, ISBN |
| If search is in collection of ARTICLES | THEN display Title, Author, Name of Journal, Volume, Issue, Year, ISSN |

In a rule based expert system, the domain knowledge is represented as a set of rules that are checked against a collection of facts or knowledge about the current situation. When the IF portion of the rule is satisfied by the facts, the action specified by the THEN portion is performed. When the condition is satisfied the rule is said to 'fire' or 'execute'. A rule interpreter is used to compare the IF portions of rules with the facts and execute the rule whose IF portion matches the facts.

### *Indexing*

Many of the systems follow a kind of keyword indexing. Unfortunately the keyword indexes are sometimes good in recall not in precision and in some cases vice-versa. Such indexes fail in preserving the context of search term. WWW is a good example of use of keyword indexing. There has been different efforts made in this direction such as using clustering techniques like keyword clustering to attach semantics with a keyword. In such a technique, relation among the terms are used like Broader term, Narrower term and Related term. This technique heavily uses Thesaurus. Such relations can be stored in the form of sequential inverted files or using B-tree structures.

#### 14.5.1.3 Search Agent

Any ISAR system should be backed up by a Search Agent. It is a program which takes input from Search Interface and searches in the Knowledge Base

using existing index. A good ISAR system means efficient retrieval. Thus, a good search agent must be equipped with following features:

- facility of using Boolean operators
- context setting to search terms
- use of clustering algorithms
- use of soundex and metaphone algorithms

#### 14.5.1.3.1 Boolean Operators

Three Boolean operators are AND, OR and NOT. These operators are used to generate combinational search. AND and NOT operators increase precision whereas OR increases recall of search results. The shaded area represents retrieved records in the following example.

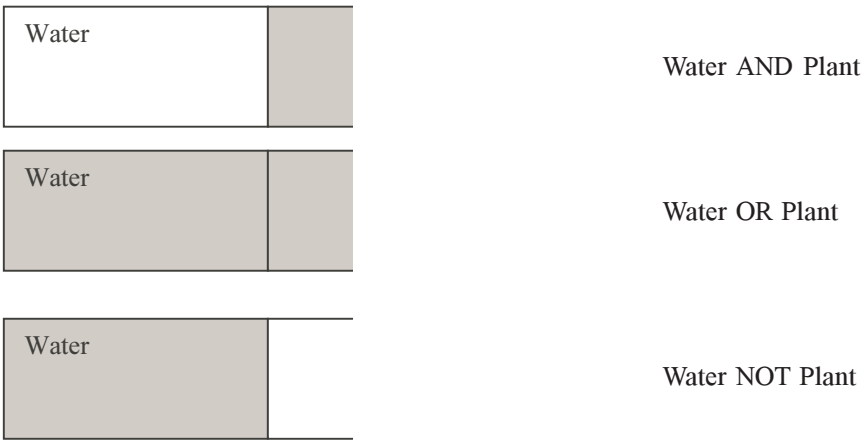



**Fig. 14.3: Boolean operators**

#### 14.5.1.3.2 Context Setting

Context Setting requires content analysis of document. Here one analyses document manually or automatically in order to preserve the context of each term in the index. It can be done in two ways, i.e., Conceptual Analysis and Relational Analysis. Conceptual analysis can be thought of as *frequency of concepts*. Concept can be represented by texts as well as pictures where text is very common. To analyze the concept one looks for the appearance of words in the text. It is not necessary that same word appears always, there may be synonymous terms present. For example, if one is analyzing with a hypothesis that a certain document is about *freedom* then one should look for the related words like *liberation*, *independence*, etc. In contrast, relational analysis goes one step further by examining the relationships among concepts in a text. In relational analysis we look for what are the related words appearing next to the word in question. For example, to see what are the words that appear next to *freedom* and then determine the related concepts.

#### 14.5.1.3.3 Clustering Algorithms

Clustering is a method by which large sets of data is grouped into groups or clusters of smaller sets of similar data based on some characteristics. For example, in a group of players one can cluster players according to their specialisation of game, like those who play cricket, those who play hockey and so on. A clustering algorithm attempts to identify natural groups of components or data based on some similarity in a given population. In other words, it is a method to create sub-class in a given class. The first thing in such algorithms are identification of core entity which is also known as *centroid*. Around centroid similar kind of entities

or data are collected which are called as members of cluster. To determine cluster membership and size, most algorithms evaluate the distance between each entity or data and the cluster centroids. Statistical techniques are used for generating clusters as output.

#### 14.5.1.3.4 Soundex and Metaphone Algorithms

Soundex and metaphone algorithms are almost the same kind of algorithm. Both these algorithms are based in the way pronunciation of a word is made. In soundex algorithm, a numeric code is assigned to each character used in a word and when search is performed, words with similar codes are also brought out in search result. Metaphone is also same kind of algorithm but unlike soundex which encodes on letter-by-letter basis, it encodes groups of letters i.e. a word. Metaphone embodies more accurately the rules of pronunciation in language. Such algorithms are well established for English as a language. Both algorithms return all the words that exactly match the desired word as well as all similar sounding names.

### 14.5.2 Functionality of An ISAR System

The objective of an ISAR system is storage, processing and retrieval of information. In order to perform the above mentioned tasks, such systems are equipped with user friendly interface, powerful database management system and search agents. Information is represented inside the system in a machine readable format using frames or semantic networks or as rule base. Often such systems are combination of above three or two methods. The overall objective of the system depends upon the need of users. Once needs of users are defined, it is easy to frame a system for the purpose.

System stores information in a form of structured record which in turn is indexed using automated or manual technique. Often the search engines over Internet use Keyword indexes using automated programs called Robots or Spiders. It goes through various web pages on Internet and indexes them word-by-word. This index is stored inside databases similar to library catalogues. Any kind of search query is searched through the index created by Robots.

From the perspective of users, he/she approaches to ISAR system with some query in his/her mind. Then he/she transfers the query into a search strategy with fixed vocabulary or key term or in a fixed format (for example, SQL query). Once query is made, he/she is given a list of search results out of which one chooses the best ones of his/her interest.

At this level a user goes through the search results and evaluates the displayed result in the light of his/her query. If the result displayed does not suit to his/her needs then he/she can go for further refinement of search strategy.

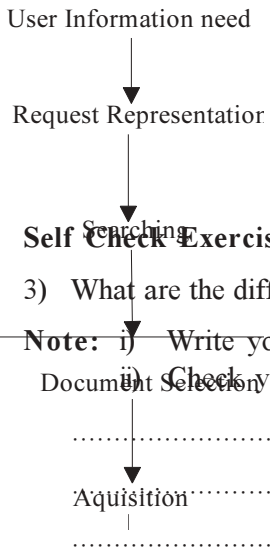



**Fig. 14.4: Functional view of ISAR**

## Self Check Exercise

3) What are the different components of ISAR system?

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

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

.....

.....

.....

.....

.....

### *Steps in the Development of ISAR Systems*

The development of an ISAR system starts with the recognition of the need for developing a system in relation to the needs of the users, as all the subsequent activities are dependent upon these. When designing, ISAR systems should follow system development life cycle (SDLC) for greater efficiency and effectiveness of the systems. In the design of the systems, following steps need to be followed :

a) Recognising the need for development of ISAR system.

b) Recognising the information needs of the users.

c) Identification of users need.

d) Deciding the type(s) of databases to be incorporated into the system.

e) Deciding about the features for incorporation in the databases.

f) Preparation of structured queries.

g) Design and development of various components of the system such as user interface, search agent, etc.

h) Evaluation of the system.

i) Re-designing/Modification of ISAR system, if needed.

## 14.6 EVALUATION OF ISAR SYSTEMS

In an ISAR system search can be performed in manual or automated environment. An example of manual system is library shelf-list or card catalogue whereas OPAC or Digital Libraries are examples of automated systems. According to Claverdon and others any ISAR system can be evaluated based on the following points:

- The coverage of the collection, that is, the extent to which the system includes relevant matter;
- The time lag, that is, the average interval between the time the search request is made and the time an answer is given;
- The form of presentation of the output;
- The effort involved on the part of the user in obtaining answers to his search requests;
- The recall of the system, that is, the proportion of relevant material actually retrieved in answer to a search request; and
- The precision of the system, that is, the proportion of retrieved material that is actually relevant.

If the view of Claverdon is taken and further analyzed one gets following check points, on which any ISAR can be evaluated:

- **Coverage**
- **Cost Benefit Analysis**
- **Time**

The important issue which should be taken into consideration while evaluating an ISAR, is the Time factor. The first and foremost thing which should be observed, How long a user takes to get the satisfactory answer to his question? If a user spends more time over system searching for his/her answer, it means that the system has to be modified in such a way that one gets his/her answer immediately. Searching over an ISAR system depends a lot and varies from on person to person. Therefore such kind of measurement should be supported by a long observation.

Other evaluation criteria are:

- **Completeness and Relevance**

- **Novelty Ratio**
- **Noise**

These have been discussed in detail in Unit 5 of this Course. The evaluation parameters discussed therein in the context of indexing systems are equally applicable for ISAR System as a whole or a sub-system of it.

**Critical Aspects of Recall and Precision**

There is a problem of determination of the total number of relevant documents in collections unless one scans the whole collections completely, which is almost unfeasible. Another thing is that relevancy cannot be measured only by yes/no terms, but also by the degree of relevancy. Recall and precision have inverse-relationship, where both cannot achieve the highest degree at a point of time. The recall and precision are influenced by the following factors:

- queries that imperfectly represent information needs;
- indexing factors;
- search strategy factors; and
- vocabulary factors.

Lack of specificity, lack of exhaustivity, lack of specific terms, inadequate hierarchical cross-reference structure, defects in hierarchy, failure to cover all reasonable approaches to retrieval, etc. are the major hindrance to achieve higher degree of relevant retrieval from an ISAR system which also affect recall and precision ratios.

### 14.6.1 Systems Criteria for Evaluation

The basis for evaluation of an ISAR system could be done at every stage with every component of it. Once again it centres around the users, information sources, intermediaries, the tools, techniques, methodologies and overall environment. B.C. Vickery and Alina Vickery have provided a detailed structure for it.

After setting a framework for evaluation of ISAR system, they present a panorama and yet analyse view of ISAR system which is on the basis of quality and value as beneficial aspects of the evaluation system. User demands and their satisfaction through provision of right information will be ultimate aim of any evaluation. They have structured their thoughts into following aspects:

Table 14.1: Criteria for Evaluation

| S.No. | ASPECTS | PARAMETERS |
|---|---|---|
| 1. | Criteria for Evaluation | Quality and value |
| 2. | A Frame Work of Evaluation | Correlation of Information Provision, Information Content and Information use |
| 3. | Relevance and Assessment | Source and Receiver coexistence |
| 4. | Service Qualities | Simplicity, ease of use, personal attention, internal decor and presentability, success/ failure analysis |
| 5. | Evaluating Performance | Choice of characteristics for Performance. Measure of Performance |
| 6. | System Efficiency, Cost, Manpower and Cost-Effectiveness | Labour, Expenditure on documents consumables, Equipment External Charges, Service overhead management and development |
| 7. | Coverage in Acquisition Search | Messages acquired/ Messages omitted |
| 8. | Retrieval from Store | Relevant/ Not relevant, Precision of retrieval |
| 9. | Evaluation of an Information System | Study of Hits/ Non Hits<br>Relevance of Hits<br>Search Qualities |
| 10. | Operational Current Awareness Service | User assessment of processing, indexing and people studies |
| 11. | Online Search Service | Value of computer-based information searches<br>Value of Internet-based information searches |
| 12. | Experimental Study of Retrieval | Intellectual analysis and its role on operations in retrieval |
| 13. | Availability on Demand | Compatibility between Database and the User's demands |
| 14. | Variables affecting Availability | Information and studies |
| 15. | Document-delivery Test | Actual demand potential demand time.<br>Time lag studies. |
| 16. | The Effect of Service Delay | Impact on access<br>Impact on queuing<br>Impact on service |
| 17. | Degradation of Performance | Queue imputes<br>Set priorities in the queries |
| 18. | Value of Information | Reference<br>Introduce self-service personal knowledge<br>Synthesis<br>Measurement of Relevance |
| 19. | The Perceived Value of Information Services | Matrix of value for professional literature |

Nowadays advanced technologies are used in modern ISAR systems. The full-text and multi-media contents are now common in modern ISAR systems. Also hybrid systems coexist that provide access to physical collections as well as electronic collections. The criteria for evaluation, as suggested by B.C. Vickery and Alina Vickery, are still relevant in modern ISAR systems, but require some modifications to cope up with new techniques, new technologies and new types of documents.

## 14.7 EXAMPLES OF ISAR SYSTEMS

### 14.7.1 Online Public Access Catalogue (OPAC)

Library catalogue has changed its form to the electronic form with the advent

of computerisation and today it is called the OPAC (Online Public Access Catalogue). OPAC provides access to the documents by different approaches of users, such as Author, Collaborator, Title, Subject, Keyword, etc. An OPAC is built on Data Base Management System.

An OPAC should provide facilities for searches like Boolean, truncation and proximity searches so that users can locate document without much effort. OPAC uses bibliographic standards like MARC21, CCF (Common Communication Format), UNIMARC etc.

### 14.7.2 Digital Library

Online Public Access Catalogues (OPAC) provides only bibliographic details not full-text information. This limitation led to development of full-text databases in digital form. This kind of database is known as Digital Library. Digital libraries are available over network often using WWW.

The costs of creating, storing, and transmitting digital information have been decreased, and the technology to support distribution and access is widespread. Rising acquisition and subscription fees (not to mention shelving and processing costs) have forced libraries to seek other ways to make information available.

Perhaps more importantly, digital libraries support service improvement. Information search and navigation across electronic information resources are faster, with enriched points of access and alternative methods for browsing and exploration. The resources themselves can be segmented, rearranged, annotated, and enhanced in ways not possible before. Digital libraries provide remote access to information resources to the users on their desktops. Also, digitization of documents presents opportunities for long-term preservation of bodies of knowledge, if not of the original carriers of that knowledge.

### 14.7.3 Search Engines

WWW is itself is a big source of information. Almost everything and anything can be found over Internet. Search Engines provide a kind of interface for users to search the web. A Search Engine basically has three components:

- a Robot or Web Crawler
- a Database
- an Agent

Web Crawler goes to each and every site over Internet and indexes each word present in the page or sometimes few lines from the page. This index is stored in search engines database with corresponding URL (Uniform Resource Locator). When a search query is given it searches in databases of search engine and result is generated.

Search Engines can be categorised into three main types:

- Individual Search Engines – text or image based search engine, for example, Google (www.google.com)
- Subject Directories – subject-based search engines, for example, Yahoo! (www.yahoo.com)
- Meta-search Engines – search engine of search engines, for example, Askjeeves (www.askjeeves.com)

## 14.8 SUMMARY

Libraries are store house of information. In a big collection, it is very difficult to locate a document or any piece of information in the library. For this purpose libraries had evolved tools like catalogue and index. With the advent of Information Technology and development of automation of libraries the tools like catalogue and index are being transformed into Automated Information Storage and Retrieval (ISAR) Systems. A typical ISAR system has following components:

- User interface
- Knowledge base
- Search agent

In course of time and with the expansion of IT, several applications of ISAR has been developed. The most important product of automated ISAR system is Digital Libraries.

## 14.9 ANSWERS TO SELF CHECK EXERCISES

1) A system designer should keep in mind following points while designing an ISAR:
   - The ISAR should act as facilitator between the information (contained in document) and the users
   - Search results must be free from any kind of ambiguity
   - Searching through the system should take minimum time
   - ISAR should have user friendly interface.

2) There are different kinds of ISAR systems such as:
   - Database Management System (DBMS)
   - Text Retrieval System
   - Management Information System (MIS)
   - Decision Support System
   - Knowledge Based System.

3) An ISAR system has following components:
   - User Interface
   - Knowledge Base
   - Search Agent.

## 14.10 KEYWORDS

**Boolean Query** : A query that is a Boolean combination of terms. For example, *Information AND Retrieval, Vision OR Sight*.

**Classification** : The process of deciding the appropriate category for a given document.

**Collection** : A group of documents that a user wishes to get information from.

**Information Retrieval** : The study of systems for indexing, searching, and recalling data, particularly text or other unstructured forms.

**Inverted File** : A representation for a collection that is essentially an index.

**Precision** : A standard measure of IR performance, precision is defined as the number of relevant documents retrieved divided by the total number of documents retrieved.

**Query** : A string of words that characterizes the information that the user seeks.

**Recall** : A standard measure of IR performance, recall is defined as the number of relevant documents retrieved divided by the total number of relevant documents in the collection.

**Relevance** : An abstract measure of how well a document satisfies the user's information need. Ideally, system should retrieve all of the relevant documents.

**Robot** : Any browser program which follows hypertext links and accesses web pages but is not directly under human control. Examples are the search engine spiders, the 'harvesting' programs which extract e-mail addresses and other data from web pages and various intelligent web searching programs.

**Spider** : Also called a robot, a program that scans the web looking for URLs. It is started at a particular web page, and then access all the links from it. In this manner, it traverses the graph formed by the WWW. It can record information about those servers for the creation of an index or search facility. Most search engines are created using spiders. The problem with them is, if not written properly, they can make a large number of hits on a server in a short space of time, causing the system's performance to decay.

**URL (Uniform Resource Locator)** : A URL (Uniform Resource Locator) is the address of a file (resource) accessible on the Internet, e.g., http://www.dell.com

## 14.11 REFERENCES AND FURTHER READING

Bailey, Charles W. (1990). The intelligent reference information system. *CD-ROM Librarian,* 5(8), 10–19.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Massachusetts: MIT Press.

Chowdhury, G.G. (1994). *Information retrieval system*. Calcutta: IASLIC.

Chowdhury, G.G. (1994). *Introduction to modern information retrieval system*. 2nd Edition. London: Facet Publishing.

Cleverdon, C.W. (1970). Progress in documentation: evaluation of information retrieval systems. *Journal of Documentation*, 26, 55-67.

Cleverdon, C.W. (1972). On the inverse relationship of recall and precision. *Journal of Documentation*, 28, 195-201.

Cleverdon, C.W., Mills, J. and Keen, M. (1966). *Factors determining the perfor-*

*mance of indexing systems*, Vol. 1, Design, Vol.II, Test Results, ASLIB Cranfield Project, Cranfield.

Doyle, L.B. (1965). Is automatic classification a reasonable application of statistical analysis of text?. *Journal of the ACM*, 12, 473-489.

Doyle, L.B. (1975). *Information retrieval and processing*. Los Angeles, California: Melville Publishing.

Ghosh, S.B. and Satpathi, J.N. (eds.) (1998). *Subject indexing systems: concepts, methods and techniques*. Calcutta: Indian Association of Special Libraries and Information Centres (IASLIC).

Gopinath, M.A. (1999). *Evaluation of information storage and retrieval (ISAR) systems*. In: MLIS-03, Block 3, Unit 12 course material. New Delhi: Indira Gandhi National Open University.

Gopinath, M.A. (1999). *Objectives of information storage and retrieval system*. In: MLIS-03, Block 3, Units 9 course materials. New Delhi: Indira Gandhi National Open University.

*Knowledge representation in computer – supported environment*. <http://www.uia.org/knowledg/knowrep.htm>.

*Knowledge representation*. <http://blackcat.brynmawr.edu/~dkumar/UGAI/kr.html>.

Lancaster, F.W. (1991). *Indexing and abstracting: theory and practice*. London: Library Association.

Lancaster, F.W. (1968). *Information retrieval systems: characteristics, testing and evaluation*. New York: Wiley.

Leibowitz, J. (1990). *Dynamics of decision support systems and expert systems*. Chicago: Dryden.

Maron, M.E. and Kuhns, J.L. (1960). On relevance, probabilistic indexing and information retrieval. *Journal of the ACM*, 7, 216-244.

*Methods of knowledge representation*. <http://busfa.vu.edu.au/awenn/kbs/week3/KnowRep.html>.

Minsky, M. (1968). *Semantic information processing*. Massachusetts: MIT Press.

Minsky, M. A. (1975). *Frame for representing knowledge*. In: *Psychology of Computer Vision*. New York: McGraw–Hill.

Paice, C.D. (1977). *Information retrieval and the computer*. London: Macdonald and Jane's.

*Refdesk search engine*. <http://www.refdesk.com/newsrch.html>.

Salton, G and McGill, M. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.

Senko, M.E. (1969). *Information storage and retrieval systems*. In: *Advances in Information Systems Science*. New York: Plenum Press.

Sparck Jones, K. (1971). *Automatic keyword classification for information retrieval*. London: Butterworth.

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

Waterman, D.A. (1986). *A guide to expert systems*. Massachusetts: Addison-Wesley.

Weil, C.B. (1968). Automatic retrieval of biographical reference books. *Journal of Library Automation*, 1(4), 239–249.

Zarri, Gian Piero. (1985). Interactive information retrieval: an artificial intelligence approach to deal with bibliographic data. *Advances in Intelligent Information Retrieval: I* Informatics, 8, 101–149.