Online Information Retrieval

Course Code: **LIS-6210**

Course Instructor:

Asim Mehmood Khan

Lecturer,

Department of Library & Information Sciences

University of Sargodha

Outline

- Features of an information retrieval system

- Elements of an information retrieval system

- Kinds of an information systems

- Design issues

- Discussion

Features of an information retrieval system

- An information retrieval system (IRS) is designed to enable users to find relevant information from a stored and organized collection of documents.

- The concept of information retrieval (IR) presupposes that there are some documents or records containing information that have been organized in an order suitable for easy retrieval.

- The documents or records we are concerned with contain bibliographic information, which is quite different from other kinds of information or data.

Features of an information retrieval system

- Conventional database management systems, i.e., Access, Oracle, MySQL, and so on, deal with structured data, where the organization or structuring of data takes place depending on the specific attributes of the data elements.

- As opposed to a conventional database management system, an IRS is designed to deal with unstructured data.

- The major objective of an IRS is to retrieve the information – either the actual information or the documents containing the information – that fully or partially match the user's query.

Features of an information retrieval system

- Whatever the nature of the database may be – bibliographic, full-text or multimedia – the system presupposes that there is a group of users for whom the system is designed.

- Although historically IRS were designed to help people find information from bibliographic and textual databases, in today's world we use IRS in almost every aspect of our daily lives, i.e., to retrieve a message or information in our mobile, specific information in sent and received e-mails, to find something or someone on the web, search in an online catalogue or in digital library, to find a video or song on YouTube or similar media, etc.

Elements of an information retrieval system (IRS)

➦ IRS may comprise one or more different types of documents and can contain text as well as multimedia information.

➦ All the documents are processed to create an index, which is searched for retrieval of information.

➦ In simplest form, the index can be considered as a back-of-the-book index, but in reality it is much more complex than that.

➦ Users interact with IRS through an interface where they are usually expected to express their information needs in the form of a query.

➦ Information is retrieved whenever the search terms match the index terms.

Elements of an information retrieval system (IRS)

- **Purpose**
  - An IRS aims to collect and organize information in one or more subject areas in order to provide it to users as soon as they ask for it. Belkin describes how IRS are used in the following way:
    - A writer presents a set of ideas in a document using a set of concepts
    - Somewhere there are users who require the ideas but may not be able to identify them
    - IRS match the writer's ideas expressed in the document with the users' requirements or demands.

Elements of an information retrieval system (IRS)

- **Functions**
  - An IRS deals with various sources of information and users' requirements.
  - An IRS must;
    - Analyse the contents of the sources of information as well as the users' queries, and then,
    - Match these to retrieve those items that are relevant

Elements of an information retrieval system (IRS)

- **Functions**
  - To identify the information (sources) relevant to the areas of interest of the target users' community
  - To analyse the contents of the sources (documents)
  - To represent the contents of analysed sources in a way that matches users' queries
  - To analyse users' queries and represent them in a form that will be suitable for matching the database
  - To match the search statement with the stored database

Elements of an information retrieval system (IRS)

➡ **Functions**

➡ To retrieve relevant information

➡ To make continuous changes in all aspects of the system, keeping in mind the rapid developments in information and communication technologies (ICTs) relating to changing patterns of society, users and their information needs and expectations.

Elements of an information retrieval system (IRS)

- **Components**
  - It is obvious that on the one side of an IRS there are documents or sources of information and on the other there are users' queries. These two sides are linked through a series of tasks.
  - Lancaster mentions an IRS comprises six major subsystems;
    - Documents subsystems
    - Indexing subsystems
    - Vocabulary subsystems
    - Searching subsystems
    - User-system interface
    - Matching subsystems

Kinds of information retrieval systems (IRS)

- IRS can be broadly categorized into two categories: in-house and online.

  - In-house IRS are set up by a particular library or information centre to serve mainly the users withing the organization, i.e., OPACs

  - By online IRS we mean those that have been designed to provide access to a remote database (s) to a variety of users. These are mostly commercial services and there are number of vendors that handle them.

Kinds of information retrieval systems (IRS)

➤ Another grouping could be made on the basis of the content, purpose and functions of IRS. In this approach four distinct types of IRS can be identified:

➤ OPACs

➤ Online databases

➤ Digital libraries and web-base information services

➤ Web search engines

Design issues

- A system can be defined as a set of interacting components, under human control, operating together to achieve an intended purpose.

- A system carries out processing on inputs to produce required outputs; the agents of this processing are people and machines.

- System design may be viewed as a series of choices from which the designer selects each element and tries to fit it with the proposed objective of the system.

Design issues

- The life-cycle approach to system design suggests the following basic stages I in the life of a system:
  - An analysis has to be conducted in order to establish the requirements of a system, and to learn the various options available.
  - Next comes the design phase, which eventually gives rise to a specific system to match the requirements.
  - Next comes the implementation stage, which leads into the operating evolution during which the system fulfils its objective and is modified from time to time to match the minor changes in requirements.

Design issues

- Eventually the system becomes less effective, for a number of reasons including mechanical faults, arrival of new technologies and major changes in the requirements and in the environment.
    - This stage leads to decay, which finally leads to replacement of the system – stating at step one again.

Design issues

- Liston and Schoene suggest that an effective IRS must have provisions for:
  - Prompt dissemination of information
  - Filtering of information
  - Providing the right amount of information at right time
  - Active switching of information
  - Receiving information in the desired form
  - Browsing

Design issues

- Liston and Schoene (continue)
  - Getting information in an economical way
  - Current literature
  - Providing access to other information systems
  - Interpersonal communication
  - Offering personalize help
- The above mentioned requirements have become even more essential for success in today's web-based information retrieval environment.

Discussion

- Development in information retrieval can be viewed from two different perspectives:
  - The computer-centred view, which deals with building efficient computer systems for storage, organization and access to information, and focuses on areas such as building up efficient access mechanisms, query processing, ranking algorithms and display and delivery of search results.
  - The user-centred view, which focuses on the stud of human information behaviour, understanding of human needs, information context and use, and so on.