



SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE NAME : 19CS732 INFORMATION RETRIEVAL TECHNIQUES

IVYEAR / VII SEMESTER

Unit 1- INTRODUCTION

Topic 2 : Early Developments and The IR Problem , The User's
Task



Early Developments and The IR Problem , The User's Task - Problem



- **Internet speed:** Slow speeds of Internet delay the retrieval system.
- **Mental dissatisfaction:** Online Information Retrieval system use computer technology for retrieving information which creates many physical and mental problems.
- **Irrelevant information:** Sometime this system retrieve many irrelevant information, this may dissatisfy user.



Early Developments and The IR Problem , The User's Task

- For more than 5,000 years, man has organized information for later retrieval and searching This has been done by compiling, storing, organizing, and indexing papyrus, hieroglyphics, and books
- For holding the various items, special purpose buildings called libraries, or bibliothekes, are used The oldest known library was created in Elba, in the Fertile Crescent, between 3,000 and 2,500 BC
- By 300 BC, Ptolemy Soter, a Macedonian general, created the Great Library at Alexandria



Early Developments–Cont..



- Nowadays, libraries are everywhere
 - In 2008, more than 2 billion items were checked out from
 - libraries in the US—an increase of 10% over the previous year
 - Since the volume of information in libraries is always growing, it is necessary to build specialized data structures for fast search — the indexes
 - For centuries indexes have been created manually as sets of categories, with labels associated with each category
 - The advent of modern computers has allowed the construction of large indexes automatically



Early Developments–Cont..



- During the 50's, research efforts in IR were initiated by pioneers such as Hans Peter Luhn, Eugene Garfield, Philip Bagley, and Calvin Moores, who allegedly coined the term Information Retrieval
- In 1962, Cyril Cleverdon published the Cranfield studies on retrieval evaluation
- In 1963, Joseph Becker and Robert Hayes published the first book on IR
- In the late 60's, key research conducted by Karen Sparck Jones and Gerard Salton, among others, led to the definition of the TF-IDF term weighting scheme



Early Developments –Cont..



- In 1971, Jardine and van Rijsbergen articulated the cluster hypothesis
- In 1978, the first ACM SIGIR International Conference on Information Retrieval was held in Rochester
- In 1979, van Rijsbergen published a classic book entitled Information Retrieval, which focused on the Probabilistic Model In 1983, Salton and McGill published a classic book entitled Introduction to Modern Information Retrieval, which focused on the Vector Model



Libraries and Digital Libraries



- Libraries were among the first institutions to adopt IR systems for retrieving information
- Initially, such systems consisted of an automation of existing processes such as card catalogs searching
- Increased search functionality was then added Ex: subject headings, keywords, query operators
- Nowadays, the focus has been on improved graphical interfaces, electronic forms, hypertext features



The IR Problem



- Users of modern IR systems, such as search engine users, have information needs of varying complexity
- An example of complex information need is as follows:
- Find all documents that address the role of the Federal Government in financing the operation of the National Railroad Transportation Corporation (AMTRAK)



The IR Problem –Cont..



- The IR Problem The key goal of an IR system is to retrieve all the items that are relevant to a user query, while retrieving as few nonrelevant items as possible
- That is, the IR system must rank the information items according to a degree of relevance to the user query



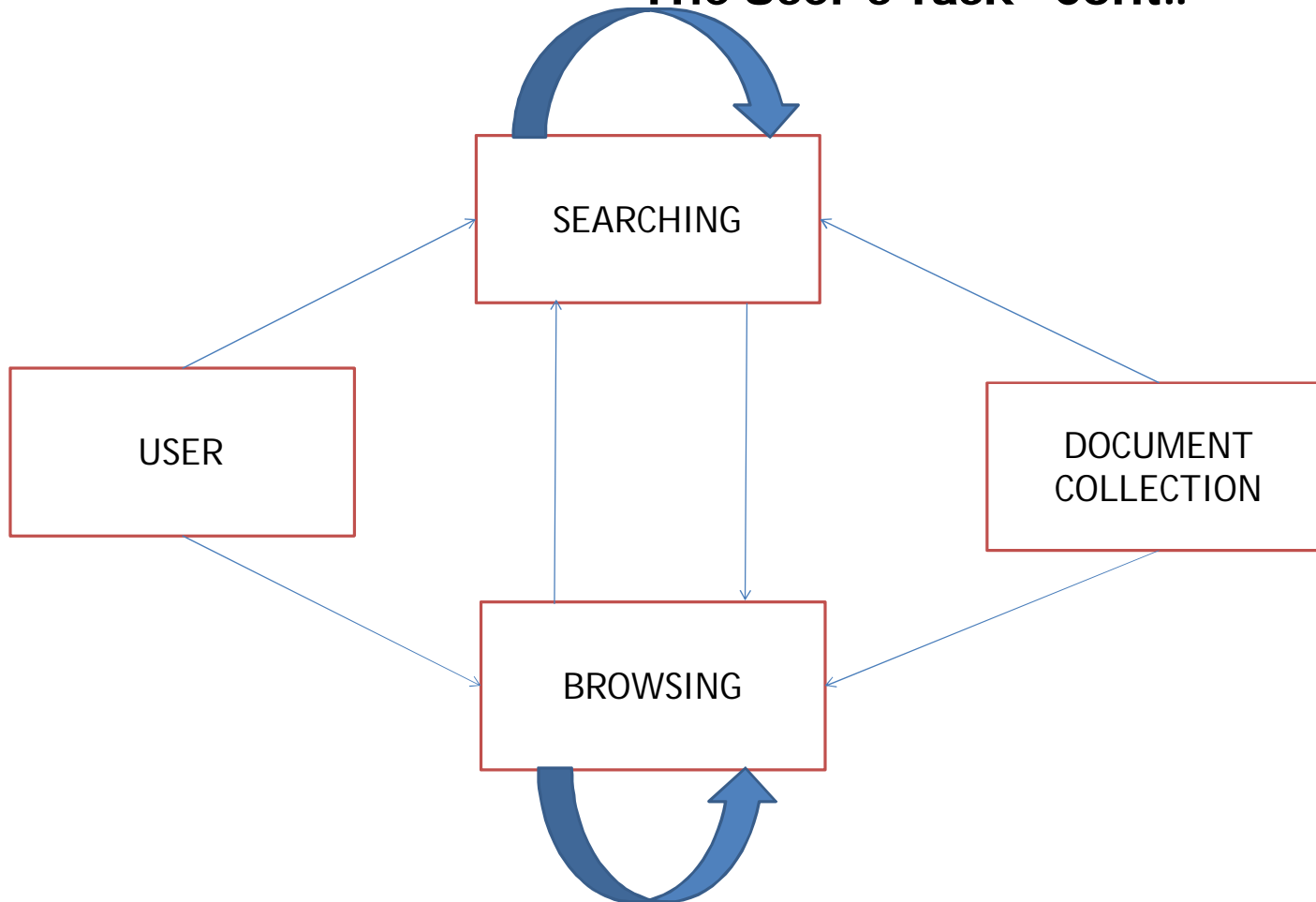
The User's Task



- **Consider a user who seeks information on a topic of their interest**
- This user first translates their information need into a query, which requires specifying the words that compose the query
- In this case, we say that the user is searching or querying for information of their interest
- **Consider now a user who has an interest that is either poorly defined or inherently broad**
- For instance, the user has an interest in car racing and wants to browse documents on Formula 1 and Formula Indy
- In this case, we say that the user is browsing or navigating the documents of the collection



The User's Task –Cont..





Activity



Disadvantages



- Primary cost
- Lack of Budget
- IT knowledge
- Lack of training facility
- Electric problem



Advantages



- Time saves
- Easy to understand
- Current information
- Database
- Multi database search
- Multiple concepts



Assessment 1



1. List out the Advantages of Early Developments

- a) _____
- b) _____
- c) _____
- d) _____

2. Identify the Applications of Early Developments

- a) _____
- b) _____
- c) _____
- d) _____





TEXT BOOKS:

1. Ricardo Baeza-Yates and Berthier Ribeiro-Neto, –Modern Information Retrieval: The Concepts and Technology behind Search, Second Edition, ACM Press Books, 2011.
2. Ricci, F, Rokach, L. Shapira, B.Kantor, –Recommender Systems Handbook||, First Edition, 2011.

REFERENCES:

1. C. Manning, P. Raghavan, and H. Schütze, –Introduction to Information Retrieval, Cambridge University Press, 2008.
2. Stefan Buettcher, Charles L. A. Clarke and Gordon V. Cormack, –Information Retrieval: Implementing and Evaluating Search Engines, The MIT Press, 2010.

THANK YOU