

DIS 401: Information Retrieval

Topic 2: Basic concepts of Information Retrieval Systems



Lecture Topics

- a) What is an information retrieval system?
- b) Features of an information retrieval system
- c) Elements and functions of an information retrieval system
- d) Kinds of information retrieval systems
- e) Design issues

a) What is an information retrieval system?

➤ Definition:

An Information Retrieval System (IRS) can be defined as:

- ▶ a system which handles and interprets the contents of the information items and user queries to retrieve the information needed to satisfy information needs of users (<http://www.ijcee.org/papers/104.pdf>)
- ▶ A system that allows people to communicate with an information system or service in order to find information (the course textbook, page 1)

➤ Goal

An information retrieval system is designed to enable users to find relevant information from a stored and organized collection of documents (the course textbook p. 2)

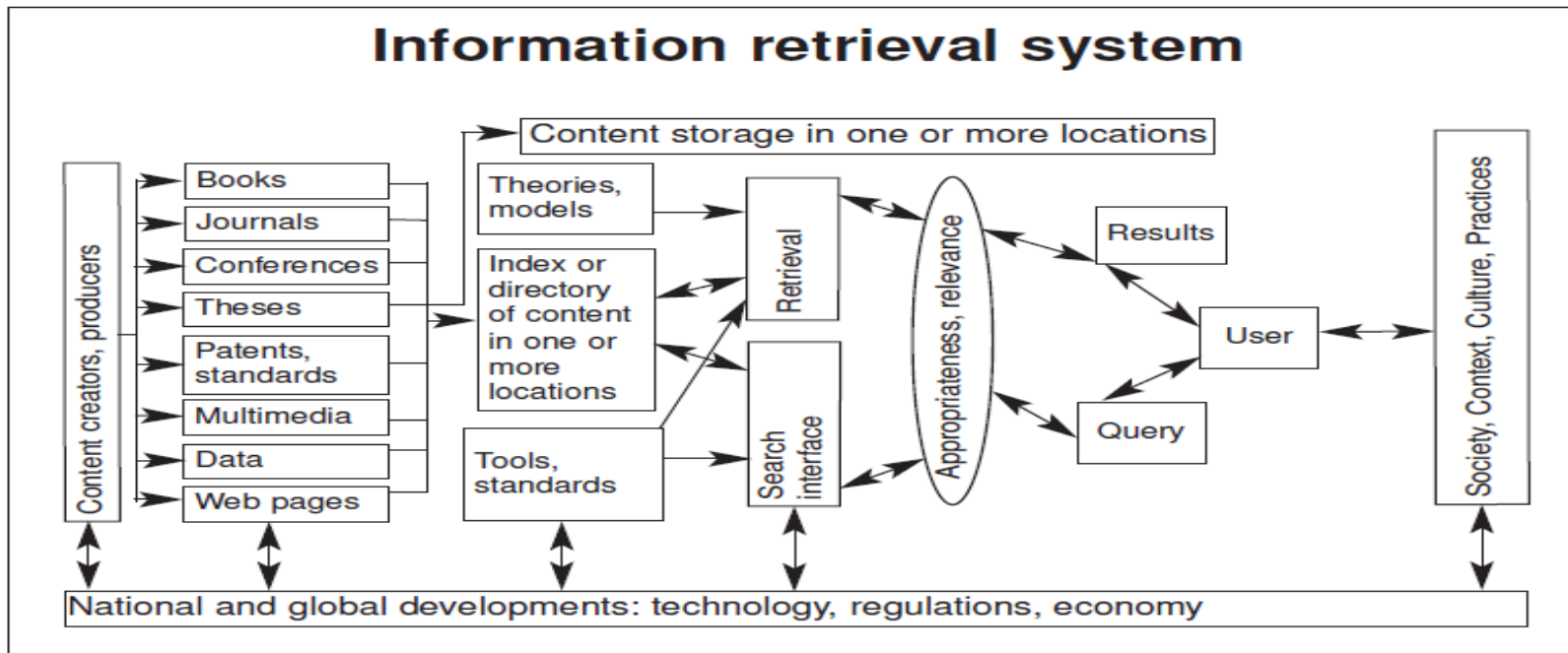
b) Features of an information retrieval system

- Modern information retrieval systems usually provide the following features:
 - ▶ (Automatic) **indexing**
 - ▶ Capability of **searching** and **retrieving all types of information items** (text, image, audio, video, etc.)
 - ▶ **Free (full)-text** and **field** search capability
 - ▶ **Boolean** and **proximity** matching
 - ▶ **Keyword** and **phrase** search techniques
 - ▶ **Multi-language** retrieval
 - ▶ **Query assistance** (recommending system)
 - ▶ Documents **ranking**
 - ▶ **Relevance feedback**
 - ▶ **Filtering capabilities** (cluster based search method)

c) Elements and functions of an information retrieval system

➤ Components of an information retrieval system:

- ▶ Document subsystem
- ▶ Indexing subsystem
- ▶ Vocabulary subsystem
- ▶ User system interface
- Searching subsystem
- Search interface
- Matching system



➤ Functions of information an retrieval system:

- ▶ Identify information (sources) relevant to the interests of targeted user community
- ▶ Analyze the content of information sources (documents)
- ▶ Represent the contents of analyzed sources in a way that matches users' queries. Currently, this is done by automatically creating one or more index files
- ▶ Analyze users' queries and represent them in a form that will be suitable for matching databases. Currently, this is made in a number of ways, through the design of sophisticated search interfaces including those that can provide some help to users for selection of appropriate search terms, by using dictionary and thesauri, automatic spell checkers, a predefined set of search statements, and so forth
- ▶ Match search statement(s) with information (documents) in databases. Information retrieval models (will be discussed later) are usually adopted by IRS developers to enable their IRS do the matching between search statements (user queries) and document terms (metadata, descriptors, subject headings, etc.,)
- ▶ Retrieve information (documents). At the end, IRS should retrieve relevant documents that satisfy user needs. Different tools & techniques are used by IRS to determine relevancy (will be discussed later)

d) Kinds of popular information retrieval systems

- Major information retrieval systems:
 - ▶ OPACs
 - ▶ Online Retrieval Systems
 - ▶ Digital libraries
 - ▶ Web search engines

▶ Online Public Access Catalogs (OPACs)

- ▶ OPACs are typical example of limited features of information retrieval systems
- ▶ An OPAC is mainly used to search for and retrieve information from and about published documents (books, journals, Theses, CDs, DVDs, etc)
- ▶ Mainly used by libraries but can be useful for any organization with a collection of published information sources
- ▶ Provide very limited bibliographic entry searches such as title, author (creator), publisher, etc
- ▶ OPACs usually have less sophisticated search and retrieval that other categories of information retrieval system

The screenshot shows the library's search interface. At the top left is the University of Sydney crest and logo. The word 'LIBRARY' is displayed in the top right. A navigation bar contains links for 'LIBRARY', 'ASK US NOW', 'CONTACT US', and 'SITEMAP'. Below this is a yellow menu with categories: 'catalogue', 'databases and electronic resources', 'borrowing', 'libraries', 'guides', 'tutorials', and 'about the library'. On the right side, there are buttons for 'CrossSearch', 'MyLoans Login', and 'Help'. The main search area is titled 'Advanced Keyword Search' and includes a warning: 'WARNING - Title search is retrieving irrelevant records'. It features four search input fields, each with a dropdown menu set to 'Any Field' and an 'And' dropdown. To the right of these fields are 'Submit' and 'Clear Form' buttons. Below the search fields, there are sections for 'Limit results' and 'Ctrl + click to select multiple options:'. Three dropdown menus are visible: 'Location' (with options: ANY, Badham, Camden, Con Music, Dentistry), 'Material Type' (with options: ANY, Music score, Music recording, USYD thesis, DVD/VHS/Image), and 'Language' (with options: ANY, Arabic, Australian Aboriginal, Chinese, Dutch). At the bottom, there is a 'Year:' section with 'Within' and 'After' options.

▶ Online Retrieval Systems

- ▶ Online retrieval systems appeared in the early years of computer application in information retrieval (60s & 70s)
- ▶ Examples of these systems were Dialog, Ovid and RBS.
- ▶ Online retrieval systems are designed to work on many large live databases containing millions of records and some times their documents
- ▶ Compared to other systems, online retrieval systems provide very advanced search and retrieval features
- ▶ Online retrieval systems are usually fee- or subscription-based services; and they provide access to peer viewed, quality (often scholarly) information sources



Collection Thesaurus

Search education resources [Advanced Search Tips](#)

Peer reviewed only Full text available on ERIC



50 Years of ERIC

The Education Resources Information Center (ERIC) is celebrating its 50th Birthday! First opened on May 15th, 1964 ERIC continues the long tradition of ongoing innovation and enhancement.

[Learn more about the history of ERIC here.](#) 

▶ Digital Libraries

- ▶ A digital library is a digital service normally available on the web that provides search, access, and delivery of organized digital collections of text, image, audio or video files
- ▶ Digital libraries are different from typical online retrieval systems in that they are often free and can be accessed by virtually anyone through the web

DIGITAL LIBRARY of GEORGIA
SHARING GEORGIA'S HISTORY AND CULTURE ONLINE

Home About DLG Educator Resources Reference Shelf Help

Search

Advanced Search
Previous Results

Browse Collections

Browse by Topic

- The Arts
- Business & Industry
- Education
- Folklife
- Government & Politics
- Land & Resources
- Literature
- Media
- Peoples & Cultures
- Religion
- Science & Medicine
- Sports & Recreation
- Transportation

Browse by Time Period

Browse by County

Browse by Institution

Browse by Media Type

Collections A-Z

Welcome to the Digital Library of Georgia

The Digital Library of Georgia is a gateway to Georgia's history and culture found in digitized books, manuscripts, photographs, government documents, newspapers, maps, audio, video, and other resources.

The Digital Library of Georgia connects users to a million digital objects in more than 200 collections from 60 institutions and 100 government agencies. Though this represents only a fraction of Georgia's cultural treasures, the Digital Library of Georgia continues to grow through its partnerships with libraries, archives, museums, government agencies, and allied organizations across the state.

Based at the [University of Georgia Libraries](#), the Digital Library of Georgia is an initiative of [GALILEO](#), the state's virtual library. Please also visit the [Civil Rights Digital Library](#), built in partnership with the DLG and GALILEO.

Partners & Sponsors

- [GALILEO](#)
- [Georgia Public Library Service \(Board of Regents of the University System of Georgia\)](#)
- [Georgia Humanities Council](#)
- [Institute for Museum and Library Services](#)
- [National Endowment for the Humanities](#)
- [The University of Georgia Libraries](#)
- [More Partners »](#)

DLG News

- [Visit DLG blog](#)
- [Recent Additions to the DLG](#)
- [Ga. Gov't Publications database redesigned](#)

Featured Collection

South Georgia Historic Newspapers Archive

The South Georgia Historic Newspapers Archive consists of over 81,000 fully searchable newspapers pages from 1845 to 1922. The papers include the *Albany News* (1870-1883), *Albany Patriot* (1845-1866), *Americus Times Recorder* (1881-1921), *Sumter Republican* (Americus, 1870-1885), *Thomasville Times Enterprise* (1873-1922), and *Valdosta Times* (1908-1912).

[View the Web site.»](#)

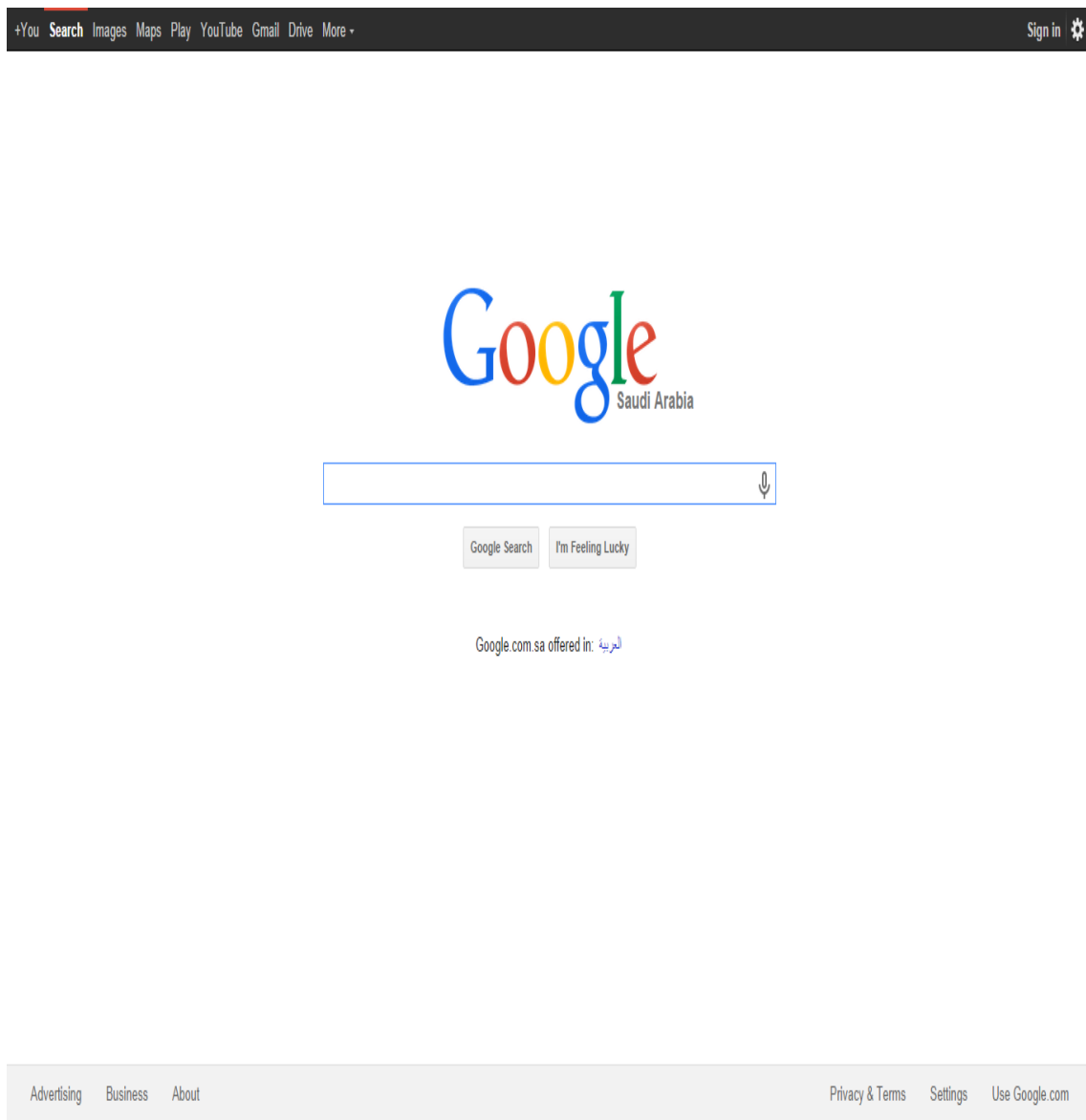
[Contact Us](#) | [Site Map](#)

The Digital Library of Georgia is part of the [GALILEO Initiative](#)
© 2011 Digital Library of Georgia

Last modified: September 17, 2012

▶ Web Search Engine

- ▶ Search engines are special tools (computer programs) on the web which are designed to help people search for and retrieve information stored on websites
- ▶ Web search engines often help people find information on the web, but they do not guarantee access to the resources they retrieve
- ▶ Search engines normally are free to use by anyone as search tools



e) Design and Selection Issues

1) Analysis

An analysis has to be conducted in order to establish the requirements of a system, and to learn the various options available

2) Design Selection

Next comes the design phase, which eventually leads to having a specific system to match the requirements.

3) Implementation

Next comes the implementation stage, which leads into the operating stage during which the system fulfils its objectives and is modified from time to time to match the minor changes in requirements

4) End & Replacement

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 – starting at step 1 again