INFS 324
INDEXING AND ABSTRACTING

Session 2 – Origins and Functions of Information Retrieval Systems

Lecturer: S. Nii Bekoe Tackie, School of Information and Communication Studies, Department of Information Studies
Contact Information: snbtackie@ug.edu.gh

UNIVERSITY OF GHANA
College of Education
School of Continuing and Distance Education
2014/2015 – 2016/2017

Session Overview

The huge volumes of information that is churned out on a daily basis, only a fraction of it is used for the purposes of teaching, learning, decision making, manufacturing and entertainment among other uses. The bulk of the information goes into storage to be retrieved as and when any aspect is required by a user. Searching for that specific information from the mass of information available out there without any means of easily identifying that item of information would be a formidable task indeed. To circumvent the problem libraries and other information management units have designed ways of organizing information for storage, retrieval, and dissemination.

- Organizing information for storage, retrieval and dissemination is called Information Retrieval (IR). In this Session, I would be looking at what information retrieval systems are and how they organize information for storage, retrieval and dissemination.

Session Objectives

By the end of this section, you should be able to:

- Learn how information retrieval systems begun.
- Understand the purpose of an information retrieval system.
- Identify the functions of an information retrieval system.
- See how an information retrieval system works
- Understand the role of indexes and abstracts in information retrieval

Session Outline

The key topics to be covered in the session are as follows:

- Topic One: The Beginnings of Information Retrieval
- Topic Two: Definitions and Functions of an Information Retrieval System
- Topic Three: What are Indexes and Abstracts?
- Topic Four: Concepts relevant to Indexing and Abstracting
- Topic Five: Definitions and Types of Indexes

Reading List

- TOPIC ONE

  THE BEGINNINGS OF INFORMATION RETRIEVAL

Origins of Information Retrieval

- I have been making references to mechanisms that will help information seekers get easy access to information.
- Facilitating easy access to information is called information retrieval.
- Indexes and abstracts play very important roles in information retrieval.

Ancient Greek and Roman Scholars

- Information Retrieval began with ancient Greek and Roman scholars who began producing large works containing different types of information.

Origins of Information Retrieval(Cont.)

- Having the good sense to know that people might not be interested in reading everything contained in the book, sought a way of organizing it so that one can retrieve only what he/she needed.
- One of the first things they did in this direction was to provide a table of contents.

- For example, Pliny the Elder wrote what he called ''The Natural History in 37 Volumes". It was a sort of encyclopedia.

- In order to make the work user- friendly, the first volume acted as the table of contents listing volume by volume all the subjects that had been treated in the subsequent Volumes.

- This acted as a directional aid to help identify where in each volume a particular subject was located.

Origins of Information Retrieval(Cont.)

- Alphabetization was another method used to facilitate the retrieval of information.
- It was devised by the Greek scholars in the third century before the birth of Christ (C3rd BC) at the library of Alexandria in Egypt.
- The system was to help them to organize the increasing number of Greek literary works in the library.

Topic Two:

DEFINITION AND FUNCTIONS OF INFORMATION RETRIEVAL

Definition of Information Retrieval

- Information Retrieval may be defined as the process of exploiting the information and other contents of documents.

- "Information Retrieval" as a term first came into use in 1952 but gained popularity in 1961 among the research community. Its purpose is to facilitate the dissemination of the right information to the right user in the quickest possible time.

Functions of Information Retrieval Systems

The functions of an information retrieval system are:

- to identify the sources of information that are relevant to the areas of interest to the target community
- to analyze the contents of the sources
- to represent the contents of the analysed sources in a way that will be suitable for matching users' requests
- to analyze users' queries and to represent them in a form that will be suitable for matching with the database

Functions of Information Retrieval Systems(Cont.)

- to match the search statement with the stored database
- to retrieve the relevant information
- to make necessary adjustments in the system according to feedback from users

In order to facilitate the information retrieval process a number of tools have been devised.

- Indexes and abstracts are two major tools that are used in information retrieval.

Topic Three:

What are Indexes and Abstracts?

Document Representations

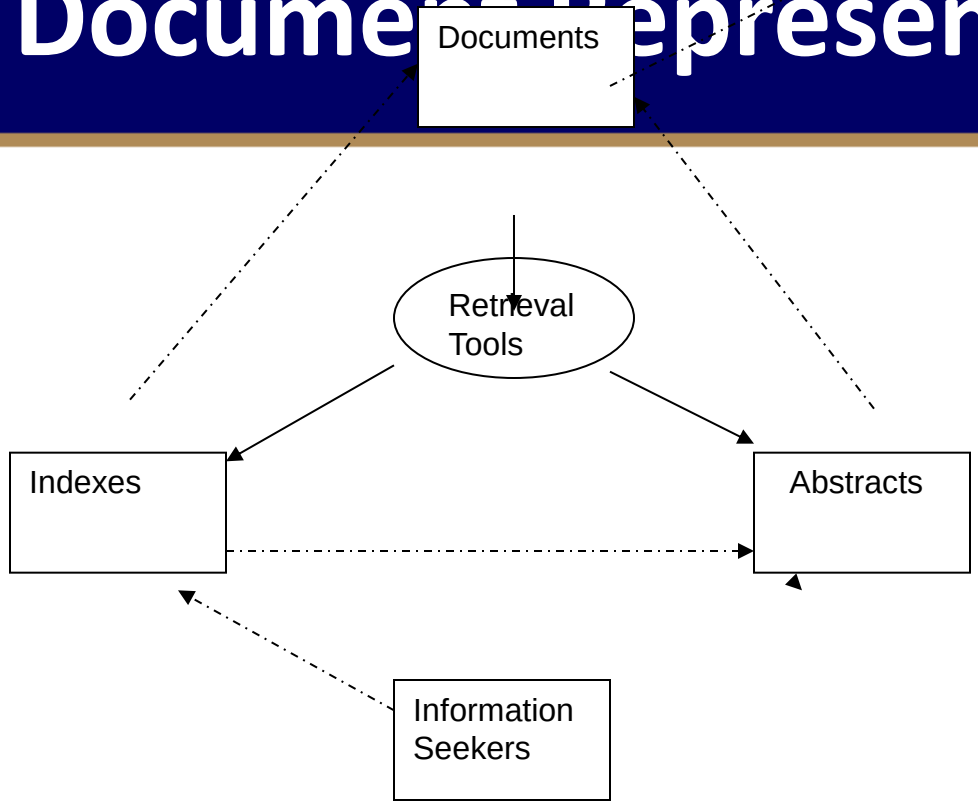
Indexes and Abstracts are:

- Document representations or document surrogates
  - used to organize literature or information so that a specialist can identify relevant documents or information more easily.
- This function of indexes and abstracts has long been recognized as valuable in the science and technology fields.
- In the last forty to sixty years they have been increasingly also recognized for their information content
  - used in business, management, social sciences and economics.

Document Representations(Cont.)

- The process of creating indexes and abstracts is what is called indexing and abstracting.
- From the discussion so far it would be seen that information retrieval is basically about searching for information or documents from a collection of documents or a store of information (database).
- Indexing and abstracting are, therefore, interwoven with searching. The diagram below illustrates the relationship between indexing, abstracting and searching.

Document Representations(Cont.)



Interrelationship of Indexing, Abstracting and Searching (Cleveland and Cleveland,….)

Document Representations(Cont.)

- The diagram shows that documents are represented by retrieval tools namely indexes and abstracts hence the straight lines.
- The broken lines show how information seekers access the document or the information.
- The information seeker may use an index or an abstract to get the document sought for.
- The line from the index to the abstract shows that sometimes an abstract may be accessed through an index.

Document Representations(Cont.)

- The effectiveness of the relationship between indexing, abstracting and searching is manifested in the user's satisfaction with the system.
- Information seekers generally are concerned with the length of time it takes for information requests to be satisfied.
- Unreasonably long delays in fulfilling an information request will discourage information seekers from using the system whether manual or computer-based.

Purpose of Indexes and Abstracts

- The basic purpose of an index or abstract is to provide effective and efficient access to information through either structured records like books and databases or random stores of information found with internet search engines. Indexes are road maps to both known and unknown information (Cleveland and Cleveland, 2013).

Cleveland and Cleveland note four things that may happen to the information seeker in consulting an index or an abstract . These are:

- Information is not found though it may be there
- Information is found though it is not what was expected

Purpose of Indexes and Abstracts(Cont.)

- Only a part of the information is found while the other part is missed
- The information is exactly what was needed and expected.

Information professionals in creating tools for information retrieval minimize the occurrence of the first three and maximize the fourth scenario.

Topic Four:

CONCEPTS RELEVANT TO INDEXING AND ABSTRACTING

Information

- Indexes and abstracts as you may have seen already are tools used in searching for information or documents in a database.
- They are used to represent the information or subject content of a document.
- From what I have said briefly I am sure that you can identify three main issues here, namely “information”, “document” and “document representation” or “document surrogate”.
- I am going to show you how important these three concepts are to indexing and abstracting.

Information(Cont.)

The first of the concepts is Information

- Primary theoretical concept that underlies the discipline of library and information science.
- It is at the heart of the information profession. Information professionals deal with the collection, description, classification, storage, retrieval and dissemination of information upon request.
- It is very difficult to arrive at a universal definition of information. For example Machlup and Mansfield (1983) provide 30 different definitions of information some of which are:
  - -Increasing the state of knowledge of a recipient of information
  - -Resolving uncertainty
  - -Value in decision-making
  - -The physical surrogate of knowledge
  - -All published and unpublished knowledge about any given subject
  - -Body of knowledge

Information(Cont.)

- Buckland (1991), on the other hand identifies three aspects of information namely:
  - Information as a process
  - Information as knowledge, and
  -Information as 'thing'

Buckland (1991), goes on to divide information into two broad groups namely **information that is tangible** and **information that is intangible**.

Information(Cont.)

Tangible information

- is information as ‘thing’. It represents physical objects that can be seen and touched for example, documents.
- Thus tangible information is information that is recorded in the form of written documents, printed materials or digitalized materials.
- These may appear in formats like books, journals, newspapers, diaries, letters, manuscripts, tapes, diskettes, compact discs, databases, artefacts, microforms etc.

Information(Cont.)

Intangible information

- information that is described as process or knowledge.
  - abstract forms of information.
- Information is described as a process when it is performing the function of 'informing'
  - which involves transmitting information from a source to a recipient.
- Information is described as knowledge when it is performing the function of imparting knowledge to an individual
  - where it reduces uncertainty.

From the foregoing, information may best be described than defined. Describing information involves identifying its properties.

Information(Cont.)

Properties of Information:

- **Information may be described as :**
  - **True or False**
  - **Current or old**
  - **Raw or processed,**
  - **Valuable or worthless,**
  - **Confidential or open**
  - **Private or public**

It may exist in large or small quantities.

It may also exist in a wide variety of media.

- For information professionals, our primary concern is information that has been recorded in certain media. These may be print, images on film, digital signals in a computer memory, or it may be scratches on a wall (graffiti) etc.

DOCUMENTS

A document is a carrier of information.

- It is a substantive entity that carries information.
- Documents come in all manner of forms.

  - They include:

annual reports,

cartographic drawings,

photographs,

brochures,

letters,

Carriers of Information(Cont.)

diaries,
videos,
voice prints,
books, cartoons,
catalogues,
contracts,
films,
invoices,
journals,
manifests,
flags etc.

Documents(Cont.)

Classification of Documents

- Documents may be classified into groups like
  - textual and non – textual;
  - published or unpublished.
- Some may be on paper; others on film or diskette etc.
- Sometimes those on paper may be transferred onto other media
  - For example:

a paper document may be put on a P.C. to prevent the loss of the information contained in it as a result of deterioration.

- I am sure you have learned about how paper documents deteriorate in your Preservation course.

DOCUMENT SURROGATES

Providing access to documents is a very essential aspect of the preservation of human civilization.

- Most of the documents produced by mankind may be lost or may be of very little use if they were not put into storage, organized and retrieved for use by information users.
- A document surrogate is a brief representation of the original document.
- It may be found attached to the original document
  - For example

    a book index also called the back of the book index;

Document Surrogates(Cont.)

They may be separated from the original document

- For example

a general index or a bibliography, or a catalogue.

- Document surrogates definitely have advantages over the original document:

-Firstly, they are easy to store.

- Secondly, they can be browsed more easily than the original document.

- TOPIC FIVE

  DEFINITIONS AND TYPES OF INDEXES

Definitions of an Index

Rowley (1992)

- "an organized series of access points which lead from information known to the user to additional previously unknown information".

The American Standard Institution( )

- "a systematic guide to items contained in or concepts derived from a collection, that is documents, or groups of documents, or sets of objects. It is arranged in a known or stated order usually different from that of the items or concepts within the collection itself.

The British Indexing Standard( )

- "a systematic arrangement of entries designed to enable the location of information in a document."

Types of Indexes

Book Index or Back of the book index

- it's the most common form of index. It is
- normally restricted to one document as opposed to a general index that may cover several documents.
- it's a list of terms derived from the subject matter of the book.
- arranged in alphabetical order.
- Each term has a page number or numbers as the case may be on which it appears in the book listed in front of the term.
- Therefore back of the book index provides an access point to information contained in the book.

Types of Indexes(Cont.)

Author Index

- series of alphabetically arranged names of:
  - individual authors;
  - corporate authors;
  - government agencies;
  - organizations; and
  - institutions including universities.

Each entry will have either a class number or call number or accession number by which we can locate information in the index.

Types of Indexes(Cont.)

Subject Index

- based on the conceptual analysis of the contents of a document.
- also alphabetically arranged.
- involves identifying concepts or ideas and using them to represent topics or subjects in the document.

Types of Indexes(Cont.)

- Other types of indexes include
  - classified index;
  - co-ordinate index;
  - cumulative index(used for journal literature); and
  - computer databases.

Types of Indexes(Cont.)

- Computer databases may be categorized according to the physical form of the data entered into the computer.
- Two broad divisions may be identified.
  - reference databases
  - source databases.
- Reference databases link the information seeker to the source of the information which may be a document or a person or an organization.
- Source databases provide the actual information for the information seeker.

Types of Indexes(Cont.)

Reference databases may be grouped into three categories namely:

- Bibliographic databases.
  - most common form of databases.
  - consist of:
    name of the author of the document,
    place and date of publication,
    title,
    publisher,
    descriptors or access terms.
- **Sometimes they may contain abstracts of the documents.**
  - Example: **LISA** (Library and Information Science Abstracts).

Types of Indexes(Cont.)

Catalogue databases.

- catalogue of a particular library or a group of libraries in a network.

  - example is UGcat of the Balme library of the University of Ghana.

Referral databases.

provide references to information like

name,

address,

specialization, etc.

Types of Indexes(Cont.)

of individuals,

institutions,

information systems etc.

- example is **EXPERTS** which is a database of high level manpower in Ghana.

- was created by the Institute of Scientific and Technological Information (INSTI), one of the institutes of the Council for Scientific and Industrial Research (CSIR).

Types of Indexes(Cont.)

Source databases also grouped into three categories namely:

- Numerical databases.
  - provide factual;
  - Statistical; and
  - survey data among other kinds of information in the areas of
    business,
    economics,
    industry etc.

Example: **FAME** (Financial Analysis Made Easy).

Types of Indexes(Cont.)

- Full-text databases.

  - contain the full text of documents.

- Text-numeric databases.

  - contain a combination of textual and numeric data like company annual report and handbook data.

- You will be learning more about databases in your database management course

References