About the course · · ·
Primitives of IR · · · · · ·
IRS · ·
IR Application ·
IR Terminology · ·
Topics and Conferences of IR · ·

1985

جامعة محمد بوضياف - المسيلة
Université Mohamed Boudiaf - M'sila

Information retrieval (IR)

By: **Dr. LOUNNAS Bilal**

Slide 01: Introduction and contents - Content

- ▶ Introduction of IR
- ▶ IR literateur
    - ▶ Architecture
    - ▶ Retrieval type
    - ▶ Doc - Query function
    - ▶ Indexation
- ▶ IR models
    - ▶ Type
    - ▶ Models:
        - ▶ Boolean
        - ▶ Vectorial
        - ▶ Others
- ▶ Web IR
- ▶ Evalution of an IRS
- ▶ Conclusion

Slide 01: Introduction and contents - Resources

- 🌐 **WebSite** :   https://sites.google.com/view/lounnasbilal
- **Books** :
  - Introduction to information retrieval.
  - Information Retrieval Implementing and Evaluating Search Engines.
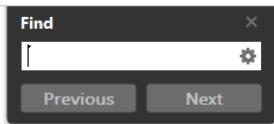 - Search Engines Information Retrieval in Practice.

Slide 01: Introduction and contents - Assignments (TDs) and Evalution

- ► Exercises on data indexation
- ► Exercises on IR models (Boolean, vectorial..)
- ► Exercises on IR Evaluation.
- ► **Implementtion** :
  - ► Datastructure review course.
  - ► Programming of some IR functionalities
  - ► Search Engines Information Retrieval in Practice.
- ► **Evalution** :
  - ► Presence and participation = 8
  - ► Pre exam = 12
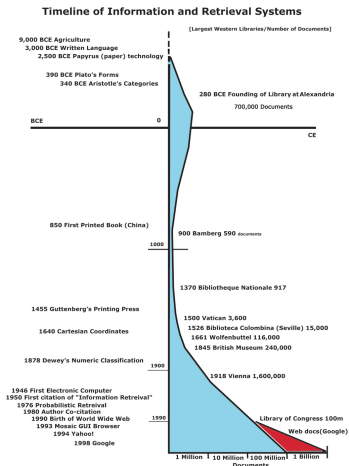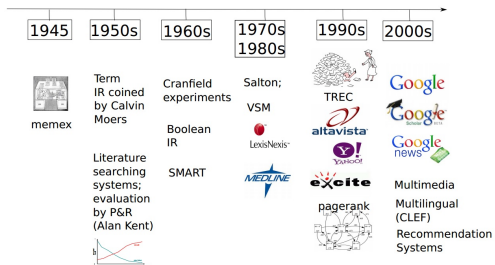
what this cours about

Web Search Google

PDF Find content

Find ×

Previous Next

File Explorer search

Windows ×

History

Timeline of Information and Retrieval Systems

► The idea of using computers to search for relevant pieces of information and that was popularized in the article "As We May Think" by Vannevar Bush in 1945
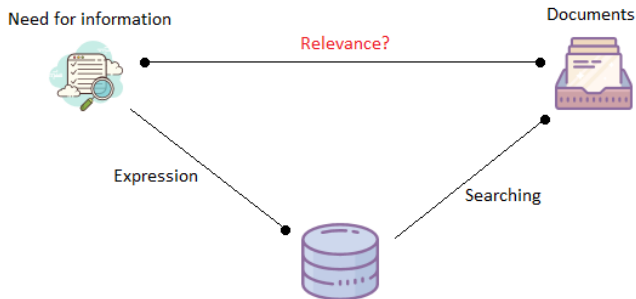
History

- ▶ Before 70 ies : Manual IR in libraries: manual indexing; manual categorization.
- ▶ Between 70 and 80 ies : Automatic IR in libraries.
- ▶ After 90 ies : IR on the web and in digital libraries.

IR Problem

> ### Problem of Information Retrieval
>
> Content-based access to documents that meet a user's information needs

Definition

Definition

Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).
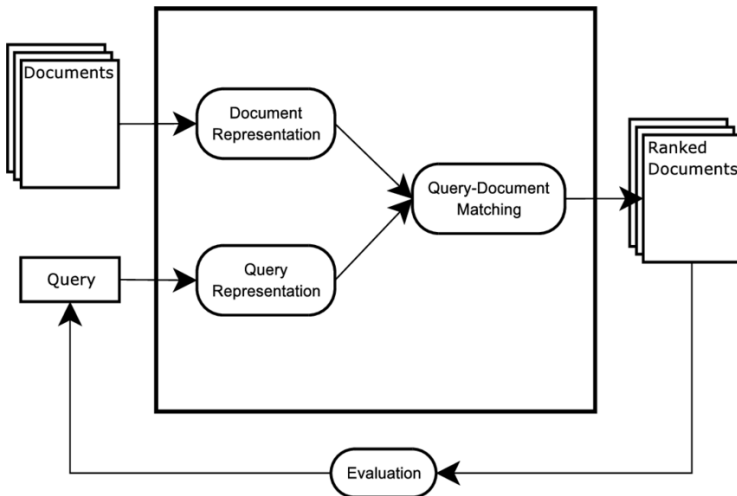
▶ Also

Information retrieval (IR) is the activity of obtaining information resources relevant to an information need from a collection of information resources

IR central element

What are the central elements of the information retrieval?

- ▶ Documents.
    - ▶ Text, Images, Video (Pdf, web page, letter, email...)
- ▶ Content of documents.
    - ▶ Meta-information: title, content...
    - ▶ Raw content and indexed content.
- ▶ Query (information need)
- ▶ Simularity or results (Docs,Query)
    - ▶ Be easy to use
    - ▶ Provide the best possible answers, and these answers must be "relevant" to the user
    - ▶ System relevance versus user relevance
    - ▶ Provide a reasonable number of answers
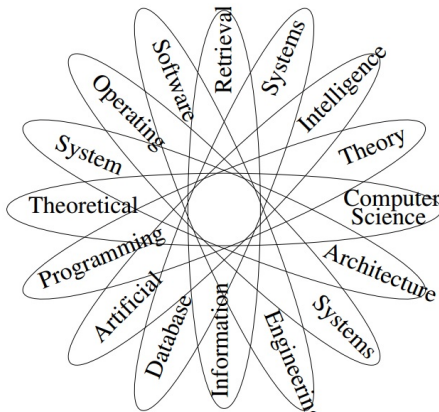   - ▶ The system must provide quick answers

IR System

IR System

The need of an efficient IRS.

- ▶ Large volume of data (web version, social media, documents)
  - ▶ Average Number of Tweets Sent Per Day: 500 million - 2 Billions queries per day on twitter
  - ▶ Every minute 510,000 posted comments FaceBook
  - ▶ 45 milliards (Google), 25 milliards (Bing)
  - ▶ 672 Exabytes - 672,000,000,000 Gigabytes (GB) of accessible data.
- ▶ Most business related information on those kind of data.
- ▶ Time and Efficiency are the most important feature.

Automated information retrieval

Information retrieval in computer science

IR Application

Contexts of use:

- ▶ Office automation
- ▶ Technical applications: equipment maintenance
- ▶ Medical informatics
- ▶ Software engineering
- ▶ Online sales
- ▶ Tourism
- ▶ Scientific research
- ▶ etc.

Terminology

- ▶ **General**: Information Retrieval, Information Need, Query, Retrieval Model,Retrieval Engine, Search Engine, Relevance, Relevance Feedback, Evalua-tion, Information Seeking, Human-Computer-Interaction, Browsing, Inter-faces, Ad-hoc Retrieval, Filtering
- ▶ **Related**: Document Management, Knowledge Engineering
- ▶ **Expert**: term frequency, document frequency, inverse document frequency,vector-space model, probabilistic model, BM25, DFR, page rank, stemming,precision

About the course
000

Primitives of IR
000000

IRS
00
0

IR Application
0

IR Terminology
0●

Topics and Conferences of IR
00

Terminology

A great glossary has been written by the Berkeley University titled by

The Modern Information Retrieval Glossary.

Topics of IR

- ▶ Retrieval models
- ▶ Text processing
- ▶ Efficiency, compression, MapReduce, Scalability
- ▶ Distributed IR
- ▶ Multimedia: image, video, sound, speech
- ▶ Web retrieval and social media search
- ▶ Cross-lingual IR (FIRE), Structured Data (XML),
- ▶ Digital libraries, Enterprise Search, Legal IR, Patent Search, Genomics IR

Conferences of IR

- ▶ SIGIR: Conference on Research and Development in Information Retrieval
- ▶ ECIR: European Conference on Information Retrieval
- ▶ CIKM: Conference on Information and Knowledge Management
- ▶ WWW: International World Wide Web Conference
- ▶ WSDM: Conference on Web Search and Data Mining
- ▶ ICTIR: International Conference on Theory of Information Retrieval
- ▶ TREC: Text REtrieval Conference