

# Chap. 1: Introduction to IR

Jean-Pierre Chevallet & Philippe Mulhem

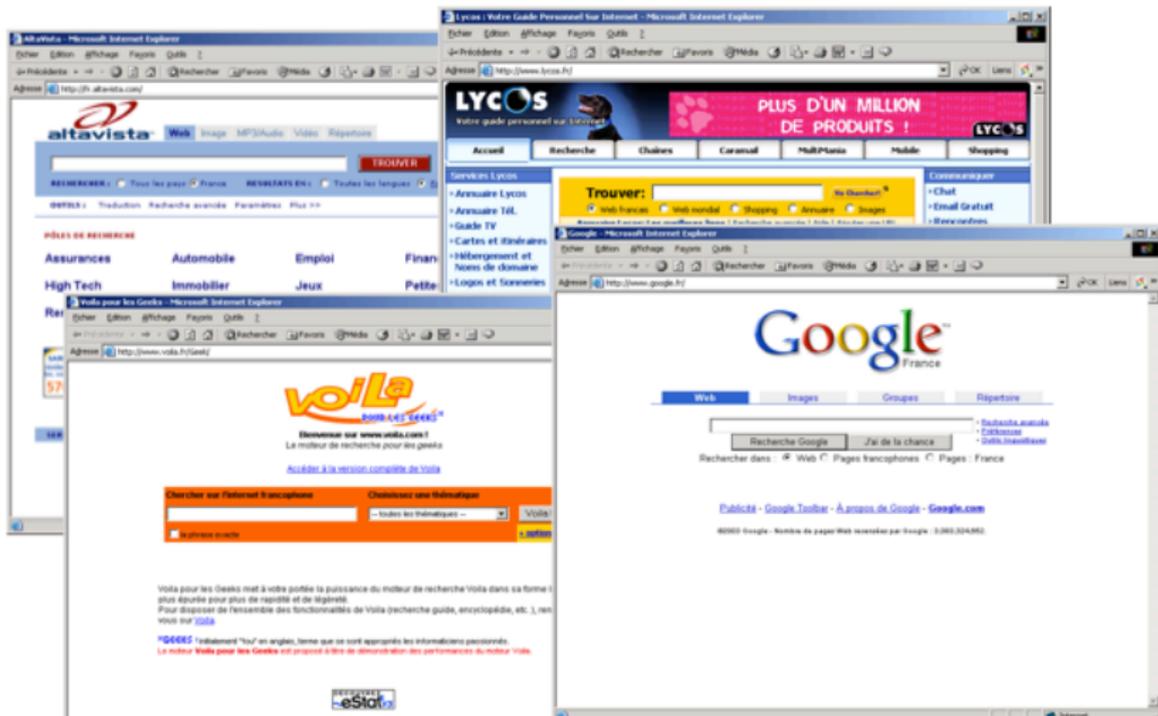
LIG-MRIM

Oct 2014

# Outline

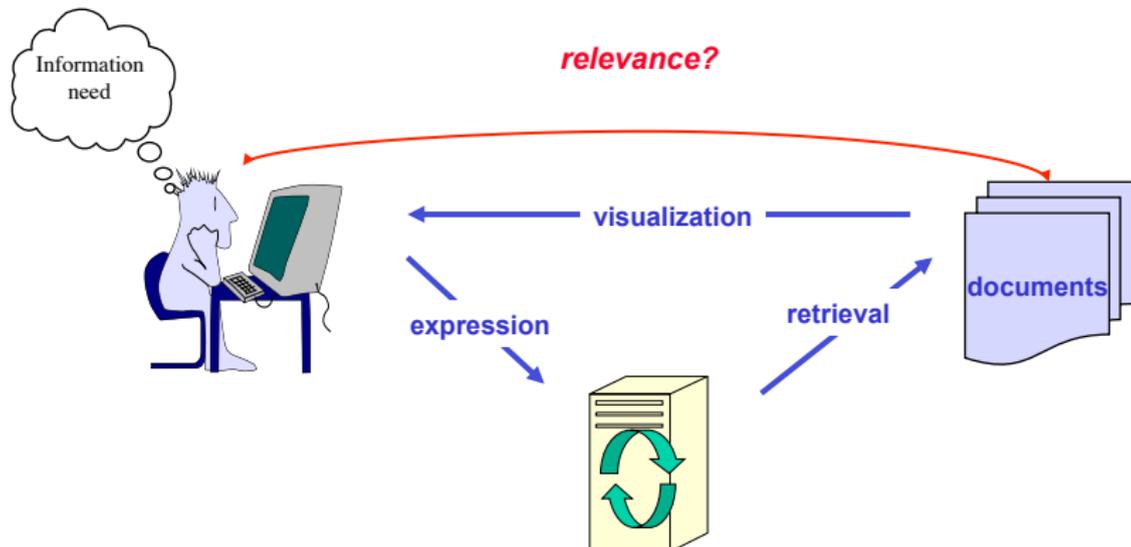
- 1 Introduction
- 2 Key points in IR
- 3 Relevance
- 4 IR Context

# Search Engine in the WEB



# Challenge of IR

Challenge of Information Retrieval:  
Content base access to documents that satisfy an user's  
information



# Mobile Information Access



# Important Notions 1

- IR definition: "Information retrieval (IR) deals with the representation, storage, organization of and access to information items" [1]
- What about:
  - Information ?
  - Document ?
  - User need

## Important Notions 2

- Central elements for IR:
  - Documents
  - Document content
  - Information need of a user
  - Satisfaction or the user
- Information:  
Is what a user gets from documents using his own knowledge

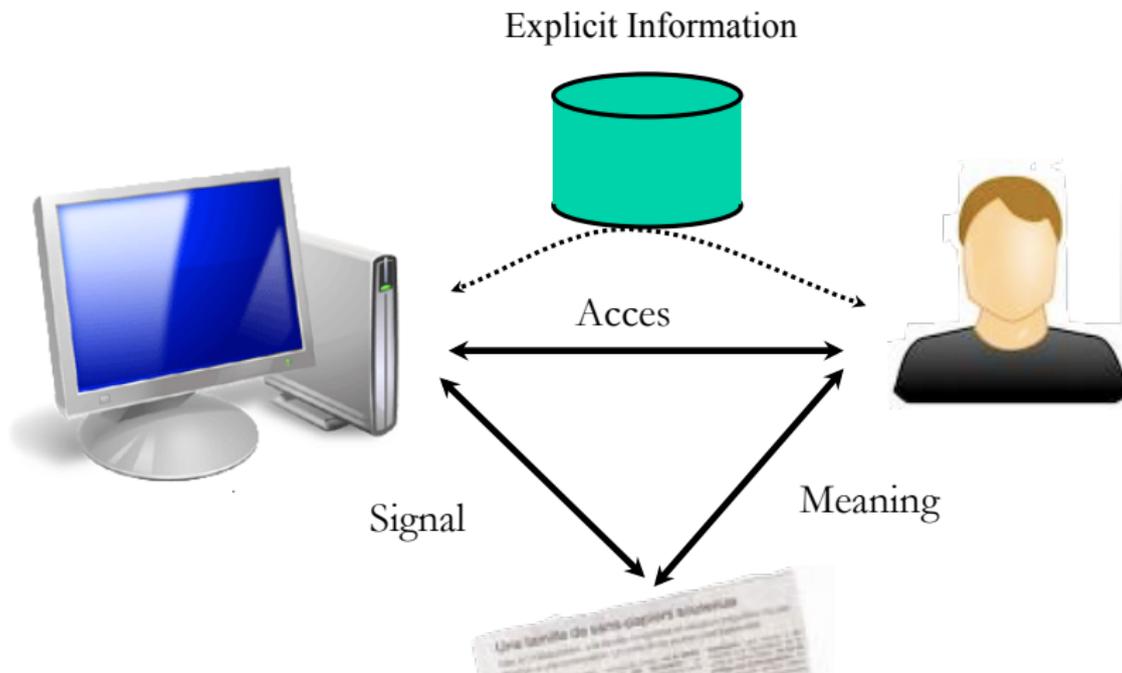
# Information

- Which kind of information ?
- transform a trace into information (ex: fire place)

# Role of IRS

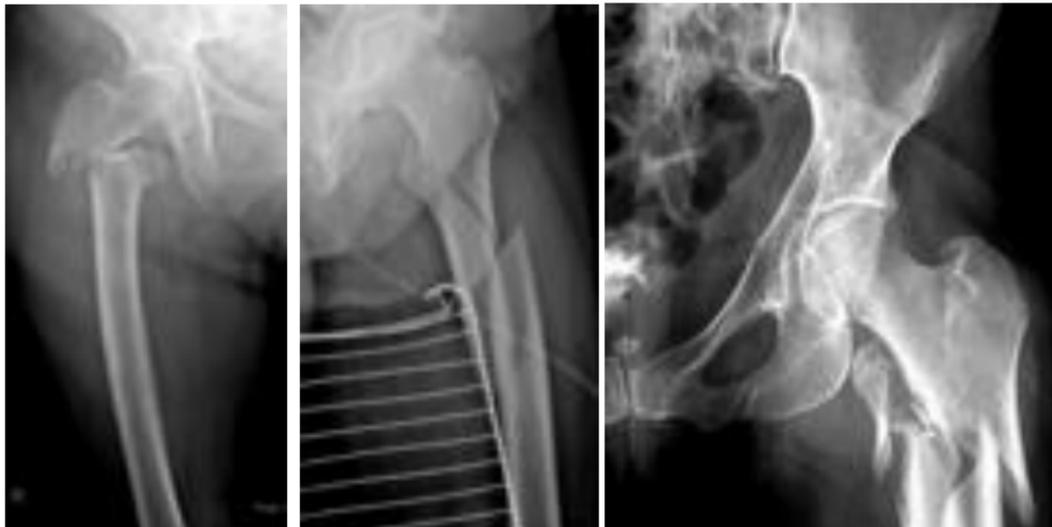
- The role of an IRS:  
An automatic mediator between user and documents
- How to match user need and document ?  
Express user need into a query  
Can we compute a match between query and document  
whitout external informations ?

# Relevance



## Multimedia queries

- Show me x-ray images with fractures of the femur.
- Zeige mir Röntgenbilder mit Brüchen des Oberschenkelknochens
- Montre-moi des fractures du fémur.



# Documents

Document as a form:

## Media

- Text, still image, video, structured documents

## Type

- Text : book, article, letter, ...
- Image : X-Rays, Photographs, Graphics,

## Granularity and structure

- Text : whole document, structure element (chapter , section, paragraph, sentence), passage (window of x words in a text), notion of **doxel** as documents atoms.
- Video : whole video, a shot, an image of the video

## Documents aspects

- Physical (form):** An object material or not, a proof function, an information support, a structure, digital = need a tool to be read
- Meaning (content):** A sign with a meaning and an intension, context is part of the meaning construction
- Social (medium):** A medium for social relationship, a trace, constructed or found, of a communication that exists outside space and time; at the same time, it is an element of identity systems and a vector of power.

# Documents

From : "Document: Form, Sign and Medium, As Reformulated for Electronic Document" Roger T. Pédaouque, STIC-CNRS

- An electronic document is a data set organized in a stable structure associated with formatting rules to allow it to be read both by its designer and its readers.
- An electronic document is a text whose elements can potentially be analyzed by a knowledge system in view of its exploitation by a competent reader.
- An electronic document is a trace of social relations reconstructed by computer systems

## Document in IR

- As an object:** the item returned as an answer. Still clear in the digital word ?
- As a sign:** the content that interests the reader, the aspect to be analyzed and indexed. A sign is supposed to have a meaning, and it is this meaning that counts for an IRS user.
- As a medium:** for social relationship, a trace, also used for collaborative works

# Document content

2 classes of information

**As an object:** Meta-Information (information about the document)

Attributes: title, author, creation date, etc.

Structure (content organization): logical and physical structures, links, etc.

**As a sign:** Content

Raw content : the initial document

Semantic content: extracted information from the raw content

# Document content

## 2 classes of information

- As object:** Meta-Information (information about the document)  
Attributes: title, author, creation date, etc.  
Structure (content organization): logical and physical structures, links, etc.
- As sign:** Content Raw content : the initial document Semantic content: extracted information from the raw content

# User Query

User's information need

Use of queries according to a predefined language

- Constrains on meta-information

Attributes : **novel written by Victor Hugo**

attribute on document type and author

Structure : **article on football containing a photograph**

Structure of links between text and image

- Constrains on the content

Raw content : **letter with the text "I came, I saw, I  
conquered "**

Retrieval on character strings

Semantic content : **documents about information retrieval,**  
retrieval symbolic descriptions

# Satisfaction of the user

Some criteria for user satisfaction:

- The system should be simple to use
- The system must give the best possible answers, and these answers must be relevant to the user
  - System relevance versus user relevance
  - Granularity of relevant information
- The system must return "reasonable" quantities of answers
- The system must give fast answers

Very hard to satisfy all these points

## User's need

Taking into account of the expertise of a user

### Domain expertise of the user

One information need expressed the same way by two persons should not necessarily give the same answers.

## User's need

Taking into account of the external context of the information need

### Temporal

One information need expressed at two different moments does not give the same answers: "tsunami" at the end of december 2004 is "obviously" related to what happen in Asia.

### Geographical

One information need expressed at two different places does not have the same meaning: "restaurant" in Grenoble do not necessarily need to give restaurants of New-York in the answer.

## Relevance at Document Side (Mizzaro 97)

### Document

The physical entity that the user of an IRS will obtain after his seeking of information.

### Surrogate

A representation of a document.

### Information

What the user receives when reading a document.

## Relevance at User Side

### Information need

A representation of the problem in the mind of the user.

### Request

A representation of the information need of the user in a "human" language, usually in natural language.

### Query

A representation of the information need in a "system" language, for instance Boolean.

## Relevance: definition

Relation between:

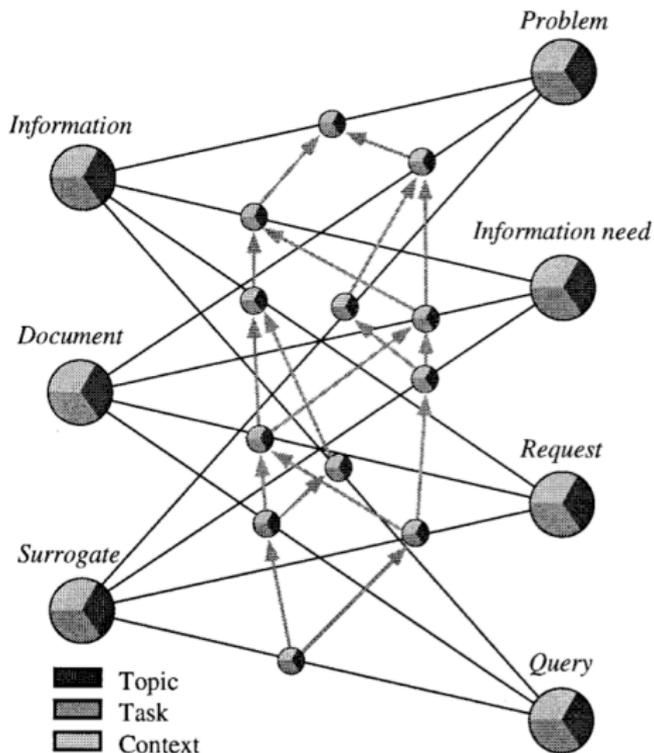
### Information Source

Document  
Surrogate  
Information

### Information Target

Information need  
Request  
Query

# Relevance Relations (Mizzaro)



From Mizzaro 97

# IR Context

IR context can be relate to:

## Cognitive state of the user

IR context > Seeking context > User task context

Interactive searching

Personalisation in search

## Physical state of the user+device

Physical context as measured by sensors

# IR Context

Tacking into account the context

To enhance precision of access (classical IR)

Filter answer using context.

For more interactivity (filtering IR)

System can "push" information triggered by context

## Context layers

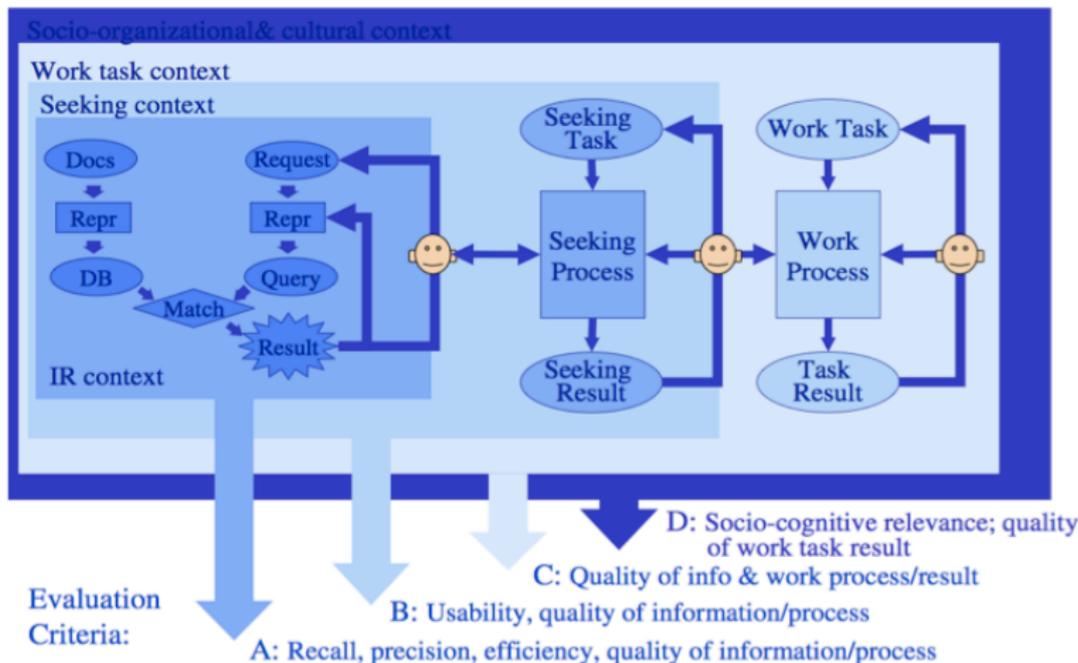
**IR context:** Related to only one query, against the IRS

**Seeking context:** A session of several queries related to the same seeking task

**Work task context** Several sessions of information seeking for a given task

**Social and cultural context** What the task is used for ?

# Context



# Context

