Fachhochschule Nordwestschweiz
Hochschule für Wirtschaft

Information Retrieval and Knowledge Organisation

Knut Hinkelmann

Fachhochschule Nordwestschweiz
Hochschule für Wirtschaft

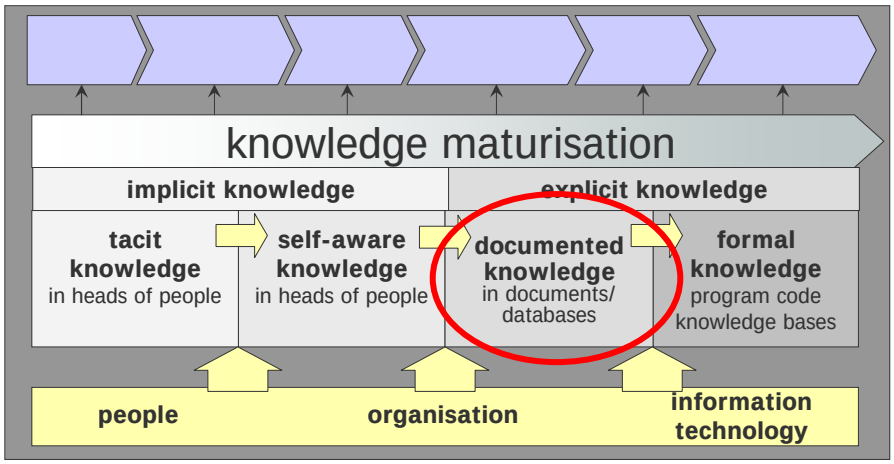
Content

- Information Retrieval
  - Indexing (string search and computer-linguistic aproach)
  - Classical Information Retrieval: Boolean, vector space model
  - Extensions of classic retrieval methods
  - Evaluating search methods
  - User adaptation and feedback
- Knowledge Organisation: Thesaurus
- Associative Search
- Metadata and meta knowledge
- Classification schemes
- Information extraction
- Case-based reasoning
- Topic Maps

1 Introduction

Fachhochschule Nordwestschweiz
Hochschule für Wirtschaft

Information Retrieval and Knowledge Management: Process-based Organisational Memory



Information Retrieval (IR) deals with the representation, storage, organisation of, and access to information items (= explicit knowledge)

Information Need

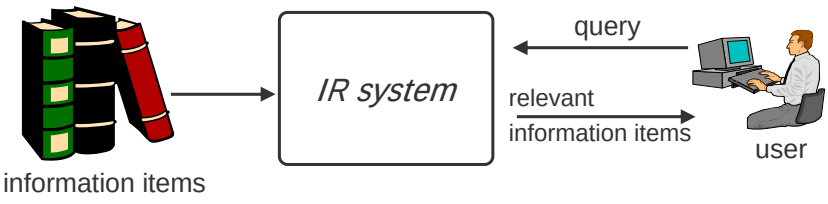
- Representation and organisation of information items should provide the user with easy access to the information items, he/she is interested in
- Characterisation of the ***user information need*** is not a simple problem
- Example:

  > Find all the pages (documents) containing information on college tennis teams which: (1) are maintained by an university in the USA and (2) participate in the NCAA tennis tournament. To be relevant, the page must include information on the national ranking of the team in the last three years and the email or phone number of the team coach.

- The user has to translate the information need into a query which can be processed by a search engine (IR system)

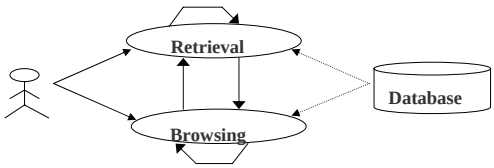
Key Goal of Information Retrieval

Key goal of Information Retrieval: Given a user query, retrieve information that might be relevant to the user.



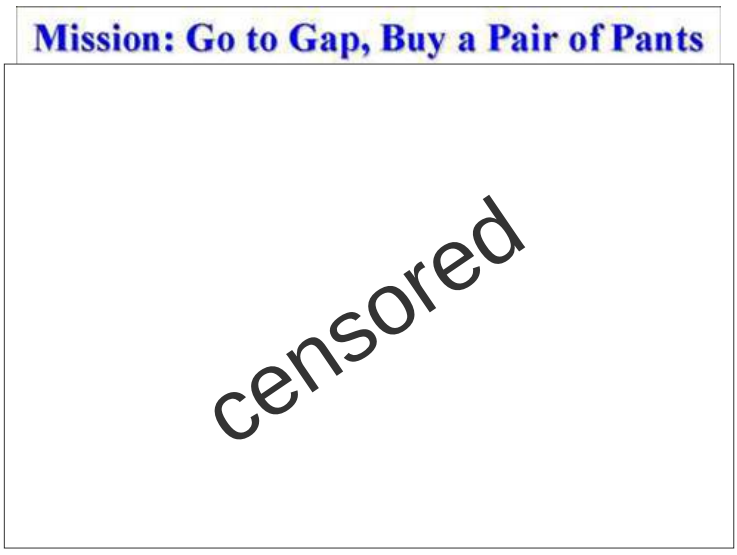
- In its most common form, the query is a set of key words which summarize the description of the information need.

Basic Concepts: Retrieval vs. Browsing



- Retrieval
  - Information need can be specified
  - A ***query*** can be formulated which contrains the information that might be of interest
- Browsing
  - Main objective not clearly defined or very broad
  - Purpose might change during interaction with the system
  - Example: User is interested in car racing in general. When finding information about formula 1, he might turn his interest into a specific driver or car manufacturer

Retrieval vs. Browsing



Filtering relevant information

- Retrieval and browsing are ***pulling*** actions
  - the user requests the information in an interactive manner
- Alternatively, information can be ***pushed*** towards the user
  - Examples: periodic news service, RSS feed
- In push scenario, the IR system has to ***filter*** information which might be relevant for the user

- What is the difference between querying a database (data retrieval) and querying a a document repository (information retrieval)?

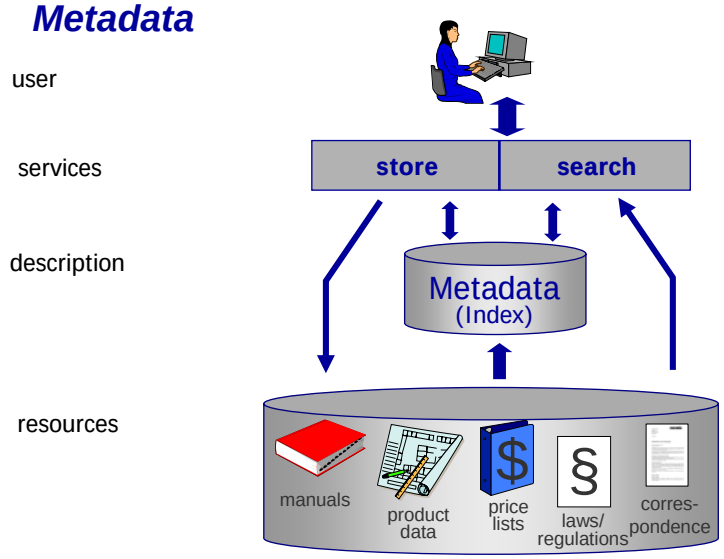
Basic Concepts: Data Retrieval vs. Information Retrieval

- data
  - typed data (integer, string, ...)
  - well-defined structure and semantics
- query results are exact
  - result: information items directly satisfying information need
  - retrieve all items that satisfy clearly defined conditions
  - a single erroneous object means failure
- data queries refer to the structure
  - attribute, table etc.
  - syntactic processing based on structure

```
select c.name, c.phone
from customer c, order o
where c.customerno =
      o.customerno
```

- information
  - is often unstructured (text)
  - may be semantically ambiguous
- query results are approximations
  - result: documents containing information satisfying information need
  - notion of relevance is in the center of IR
  - retrieved objects might be inaccurate, small failures are tolerable
- information retrieval deals with content
  - systems interprets content of information items
  - find relevant documents in spite of differences in formulation and vocabulary

Metadata

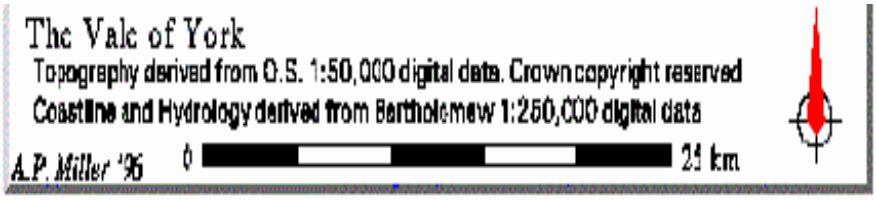


(An index is the internal representation of metadata for efficient processing)

Use of Metadata

- Search for information resources using suitable criteria
- organisation and effective access to electronic resources (e.g. document management systems, digital libraries)
- unique identification of resources
  - (storage) location, e.g. URL
  - unique ID
- distinction of different resources
- data exchange between systems with different data structures and interfaces

Meta-data Example



The meta-data of a map containing the name of the resource, the creator, die resolution, the copy right and a description.

Structured Metadata – Examples

user data (document)

metadata

name:	ELENA-Ber
creation:	18.3.2001
modification:	25.6.2001
format:	Word
document type:	project report
recipient:	All Life Insurance Inc.
author:	Smith

Examples for Meta-data:

- *library catalogue* with description of books: author, title, publication date, publisher, key words, location
- *document management systems* distinguish between user data (resources, documents) and meta-data
- *skill databases / yellow pages* contain descriptions of people

General vs. application-specific metadata

- General metadata
  - can be used for any kind of information
  - Examples: author, date of creation, key words
- Application-specific metadata
  - specific attributes
    - examples:
      for a photograph: resolution
      for a piece of music: the style or the album
  - specific attribute values
    - examples: predefined key words, project names

Types of Metadata

- descriptive
  - provide information about the content and the objective of the resource (e.g. using key words)
- structural
  - describe the structure/composition of the resource
- administrative
  - administrative information to deal with the resources (e.g. date of creation, access rights)

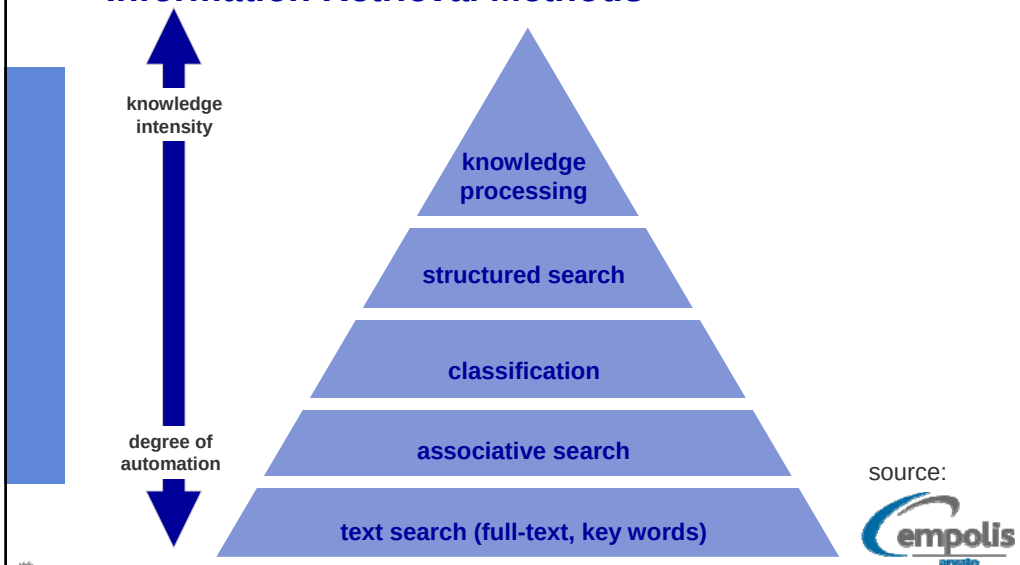
Kinds of Meta-data and Meta-Knowledge

- Textindex
  - full-text: words contained in text documents
  - key words: manually assigned to documents
- structured meta-data
  - attributes and their values
- classification systems / taxonomies
  - (hierarchy of) categories
- thesaurus
  - terms and their relations
- semantic nets / ontologies
  - concepts and relations

meta-data

meta-knowledge: knowledge organisation

Information Retrieval Methods



source: empolis arvato