Information Retrieval – Current and Future Research

Norbert Fuhr
University of Duisburg-Essen
Germany

fuhrt@uni-duisburg.de
What is Information Retrieval?

“What Information Retrieval deals with uncertainty and vagueness in information systems”
(IR Specialist Group of German Informatics Society, 1991)

- **Uncertain** representations of the semantics of objects (text, images, …)
- **Vague** specifications of information needs (iterative querying)

1. **Area definition**
2. Global information access
3. Contextual retrieval
What to Retrieve?

“Retrieve that amount of knowledge which a user needs in a specific situation for solving his/her current problem” (Kuhlen 1991)

Consider specific user, situation and problem
→ contextual retrieval (part 3)

How to get this information
→ global information access (part 2)

Workshop “Challenges in Information Retrieval and Language Modeling”, 2002
http://ciir.cs.umass.edu/irchallenges/

1. Area definition  2. Global information access  3. Contextual retrieval
Global information access

“Satisfy human information needs through natural, efficient interaction with an automated system that leverages world-wide structured and unstructured data in any language.”

1. Area definition  2. Global information access  3. Contextual retrieval
Information access

- Information properties
  - media
  - structure
  - heterogeneity

- Access methods

1. Area definition   2. **Global information access**   3. Contextual retrieval
Information Media

- Text
- Facts
- 2D: graphics, images
- Speech
- Video
- 3D

**Open issues:** representation of the semantics of non-textual media

1. Area definition  2. Global information access  3. Contextual retrieval
Information structure

- Unstructured
- Semi-structured (XML)
- Fully structured
- Hyperlinked (Web)

Open issues: (regular) semi-structured, hyperlinked data (`hidden Web’)

1. Area definition  2. Global information access  3. Contextual retrieval
Heterogeneity

- Language: multilingual
- Media: multimedia
- Heterogeneous structures
- Heterogeneous services

1. Area definition  2. Global information access  3. Contextual retrieval
Heterogeneity(2)

Open issues:

- Standardization of non-trivial structures (e.g. Dublin Core) and services (e.g. XQuery text retrieval)
- Integration approaches based on uncertainty and vagueness

1. Area definition  
2. **Global information access**  
3. Contextual retrieval
Information Access Methods

Ad-hoc retrieval
One time queries (e.g. Web search)

Filtering/Routing
Constant search profile (e.g. Spam filtering)
Information Access (2):

- **Categorization/Clustering:**
  Group documents into predefined classes/ adaptive clusters

- **Topic Detection and Tracking:**
  Cluster news in stream
Information Access(3): Summarization

for browsing / survey on retrieval results

Sheffield Speech and Hearing Research Group - Speech ...
... We believe that statistical methods are well suited to this situation ...
Content/style
models for non-extractive summarization; Multi-document summarization; ...
www.dcs.shef.ac.uk/spandh/projects/s3l/ - 4k - Cached - Similar pages

Text Interpretation: Extracting Information
... state approximation and other shallow but effective sentence processing methods, and (2) the emergence of weak heuristic and statistical methods that help to ...
cslu.cse.ogi.edu/HLTsurvey/ch7node5.html - 11k - Cached - Similar pages

Citations: Nonparametrics: statistical methods based on ranks - ...
citeeer.nj.nec.com/context/224403/0 - 33k - Cached - Similar pages
Inform. Access(4): Question answering

Find text passage answering fact query

*Hoppe report 2000*
... Shooting rate: 1-2/sec; recording **length**: 700 msec; surveying speed: 0.8 m/sec on **River Rhine**, 1.4 m/sec on Main and Neckar; daily survey **length**: 20-25 km; ...
comp1.geol.unibas.ch/report2000/report_2_1.htm - 25k - Cached - Similar pages

*Bed Breakfast Rhine Hotel Im Malerwinkel Bacharach Middle River* ...
Bed Breakfast **Rhine** Hotel Im Malerwinkel Bacharach Middle **River** Valley of ...
Koblenz
Rhineland Bicycle Palatinate Rooms way **Length** Schlafzimmer rail ...
www.loreleytal.com/bacharach/im-malerwinkel/e/ - 8k - Cached - Similar pages

*[PDF]*202.1017 Fld Riwa river without
File Format: PDF/Adobe Acrobat - **View as HTML**
... m 3 /s **River** ID card name **Rhine** (Rhein, Rijn) origin Switzerland destination North
Sea character glacier **river** contents melt- and rainwater **length** 1,320 km ...
Information Access Methods

Open issues:

- Relevance of information access methods for applications?
- Combination of information access methods?

1. Area definition  2. Global information access  3. Contextual retrieval
Current IR Research

... focuses on models, methods and systems for information properties and access methods:

\[
\{\text{Media}\} \times \{\text{Structure}\} \times \{\text{Heterogeneity}\} \times \{\text{Access methods}\}
\]

1. Area definition
2. Global information access
3. Contextual retrieval
Contextual retrieval

“Combine search technologies and knowledge about query and user context into a single framework in order to provide the most appropriate answer for a user’s information needs.”

1. Area definition
2. Global information access
3. Contextual retrieval
Considering Context

1. Area definition    2. Global information access    3. Contextual retrieval
Time-dependence

- Batch retrieval
- Constant information needs (Filtering → adaptation)
- Interactive retrieval
- Personalization:
  - Preferences
  - Seen items
  - Evolving interests

1. Area definition  2. Global information access  3. Contextual retrieval
Interactive retrieval: Levels of search activities

1. **Move**: Low-level search function (e.g. type in search term, view retrieved document)
2. **Tactic**: several moves to further a search (e.g. broaden/narrow a query)
3. **Stratagem**: set of actions on a single domain (e.g. citation database, tables of contents of journals)
4. **Strategy**: complete plan for satisfying an information need (e.g. subject search, browse relevant journals, find referenced articles)

1. Area definition  
2. Global information access  
3. Contextual retrieval
Interactive Retrieval: Current Research

**Evaluation results:** quality differences between methods in batch retrieval vanish in interactive retrieval.

**Empirical studies:** information seeking as a sequence of interconnected but diverse searches.

**Specific methods for interactive retrieval required:**
- Information seeking: ‘berrypicking’
- Tactics & stratagems

1. Area definition
2. Global information access
3. Contextual retrieval
Work context

- Context-free
- Task-specific searches
- Workflow (application-specific)

1. Area definition
2. Global information access
3. Contextual retrieval
Workflow: Generic problem solving scheme

1. Problem understanding (Hypermedia system with introductory/survey articles)
2. Identification of possible solutions (Hierarchical hypermedia system)
3. Selection of optimum solution (Information retrieval system)

→ integrated systems required
Workflow example: Digital Library Life Cycle

Metalibrary

Discover

Re-Present

Retrieve

Interpret

Collate

Annotations, discussion threads

Personal/group library

1. Area definition
2. Global information access
3. Contextual retrieval
Social context

- Single user
- (Fixed) user groups
  - Collaborative information access
- (Open) communities

1. Area definition
2. Global information access
3. Contextual retrieval
Context dimensions

1. Area definition
2. Global information access
3. Contextual retrieval
Research on Contextual Retrieval

Currently very little research

- Lack of testbeds
- Bigger experimental effort
- More application-specific → generalization of results difficult

1. Area definition  2. Global information access  3. Contextual retrieval
Future Research

- Global information access
  - Media semantics
  - Exploiting structure
  - Heterogeneous structures and services

- Contextual retrieval
  - Consideration of time, social and work context
  - Major chance for improving IR quality
Conclusion

- Global information access
  - Focus of current research

- Contextual retrieval
  - Promises significant quality improvements
  - More research necessary
  - Requires close cooperation between research and industry