Information Retrieval is for Everybody

FGIR Workshop
Hildesheim University, October 2006

Christian Wolff
Regensburg University, Media Computing
Overview

- Motivation
- Development phases in Information Retrieval
- Users and information literacy
- A storytelling example: A day in the life of the common IRS user
- Some observations and implications for IR research
- Conclusion
Motivation

- IRS usage commonplace in everyday life
- multimedia and mobile computing call for new IRS application (e.g. GIS, digital photography, ubiquitous network access)
- Web 2.0 challenges assumptions on user participation (tagging, user involvement)
- exploitation of
  - contextual information
  - new interaction techniques
Phases in IRS Development

1. IRS as instruments of information professionals in mediated communication contexts
2. IRS used by knowledge workers (“end users”)
3. web-based search: IRS used for (almost) anything (desktop setting)
4. “the digitization of the world picture” (Cerruzzi 2003): information systems and IRS are becoming ubiquitous, mobile, context-aware …

(phases overlap heavily!)
Users and information literacy

- ARD online survey 2005: 60% of the German population regularly use the internet, more than 90% in the group of the 14-19 years old
- almost 100% of pupils and students have private internet access (small studies with ~300 participants in 2005/2006)
- majority of children is familiar with basic internet and search engine concepts
- evident deficits in information literacy
  - linguistic phenomena
  - search strategies
  - query languages and operators
  - information quality judgments
  - knowledge about available resources
A storytelling example: IRS in everyday life

- In the morning: Collecting music, loading the iPod
  - different genre classifications
  - retrieval by example not viable for many tasks (humming, singing, whistling)
  - problems of describing and classifying music
  - limited portability of licenses
  - no similarity search (see MusicFinder)
  - problem of information management
Example: ID3-Tagging a Music Library

![Windows Media Player interface](image)

This screenshot of the Windows Media Player interface demonstrates ID3-tagging capabilities. The interface allows users to add metadata to music files, such as artist names and album titles. The screenshot shows the Medienbibliothek section, where users can select and tag music by various composers and artists, including Mozart, W.A. Mozart, and Wolfgang Amadeus Mozart. This feature is particularly useful for organizing and searching music libraries, making information retrieval more accessible to everyone.
A storytelling example: IRS in everyday life

- In the morning

- At work:
  - searching for literature
    - heterogeneous tools and information
    - different retrieval models and data source characteristics
    - complex criteria for availability of sources
  - personal information management
    - limitations of traditional file system structures (monohierarchical)
    - traditional fulltext retrieval becoming available via desktop search engines (classic VSM paradigm)
    - advanced techniques (aspect/context/situation awareness, embedded retrieval) missing
# Matrix of typically available literature access systems

<table>
<thead>
<tr>
<th></th>
<th>Finding</th>
<th>Accessing / Ordering</th>
</tr>
</thead>
<tbody>
<tr>
<td>books (monographs,</td>
<td>local OPAC, regional catalogs, book sellers</td>
<td>local OPAC, regional catalogs, book sellers</td>
</tr>
<tr>
<td>edited books)</td>
<td>(„Amazon“)</td>
<td>(„Amazon“)</td>
</tr>
<tr>
<td>papers (journals,</td>
<td>Integrated DBs (Scopus, Web of science),</td>
<td>electronic journal library, publishers‘ dls,</td>
</tr>
<tr>
<td>conferences)</td>
<td>science DBs (INSPEC,</td>
<td>integrated portals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(vascoda, io-port, ...)</td>
</tr>
<tr>
<td>grey literature</td>
<td>Citeseer, Google Scholar, integrated</td>
<td>directly online</td>
</tr>
<tr>
<td></td>
<td>portals</td>
<td></td>
</tr>
<tr>
<td>primary research</td>
<td>Google, specialised dbs</td>
<td>online? (registration?)</td>
</tr>
<tr>
<td>data, software, ...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A storytelling example: IRS in everyday life

- **In the morning**
- **At work**
- **In the afternoon**
  - online shopping
    - great number of platforms and meta platforms
  - „a colleague‘s visit“: taking (digital) pictures
    - manyfold media
    - no content metadata on media production time
"a colleague's visit": no descriptive metadata
A storytelling example: IRS in everyday life

- In the morning
- At work
- In the afternoon
- In the evening
  - some time for tagging and picture sorting, querying images in Flickr
  - (media convergence: interactive TV and web-based IS)
What does a typical folder with holiday pictures look like?

<table>
<thead>
<tr>
<th>Date</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-06-07</td>
<td>DSCN0476.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0477.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0478.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0443.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0444.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0447.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0448.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0449.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0450.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0451.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0452.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0453.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0454.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0455.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0456.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0457.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0458.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0459.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0460.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0461.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0462.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0463.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0464.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0465.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0466.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0467.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0468.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0469.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0470.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0471.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0472.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0473.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0474.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0475.JPG</td>
</tr>
<tr>
<td>2006-06-07</td>
<td>DSCN0509.JPG</td>
</tr>
</tbody>
</table>

Information Retrieval is for Everybody

Wolff, Media Computing, Regensburg, 13
„Abendstimmung“

We found 1,005 photos about Abendstimmung.

View: Most relevant • Most recent • Most interesting

From ecker
From Schrom
From pixelpet
From boris_gass

From ginqium
From status6
From ecker
From hillie1947

From Digitaler...
From Juza
From iBlogging
From ecker

From
From
A storytelling example: IRS in everyday life

- In the morning
- At work
- In the afternoon
- In the evening

At night
  - desktop index is being updated
  - image analysis compares pictures and suggests tags
  - user interaction data are evaluated for further search
  - (the user sleeps)
Some theses

- **everybody** is an IRS user
- the **amount of data** (files, information knowledge) is growing rapidly (e.g. > 500,000 items in my Google desktop index)
- **all media** are subject to IR processes - in the media production chain not just **search** is relevant, but also the **descriptive** steps (indexing, tagging)
- **economic, organizational** as well as **temporal** criteria influence effectiveness measures (as compared to traditional effectiveness)
- the potential of **social software** is still quite unclear
- **text** and **concept related query paradigm** will prevail for a while and for most media
- **media convergence** will gain importance
Some questions ...

- new evaluation criteria needed?
  - user satisfaction
  - task completion

- how do we
  - describe
  - analyse

situatio, context, interaction history etc.?

- how do we fuse different data sources like
  - declarative knowledge (on the user)
  - sensor data
  - ...

Information Retrieval is for Everybody
Conclusion

- demand for
  - user-related research
  - information literacy in everyday life
    - taking IL beyond the library user / student paradigm
    - the digital divide as a problem
      - not of access to technology
      - but as of a lack of training
  - models
    - information in everyday life
    - of future I(R)S
    - modeling context/situation/experience