

# **Traditional and digital libraries – a user perspective**

**Ragnar Nordlie  
Oslo University college**

# Physical vs. digital

- Traditional library values
  - Retrieval (“we have the answer”)
    - » Organization for retrieval (systems and metadata)
    - » Search assistance
  - Quality
    - » Currency of material
    - » Reliability of answers
  - Local relevance
  - Access (“seamlessness”)

# Physical vs. digital

- Will users access the digital library in the same way (or for the same purpose) as they do the physical?
- Will our methods and standards for describing the material be as conducive to the use of the documents in the digital as in the physical library?
- Can the digital collection be managed in the same way (or on the same principles) as the physical?

# Norwegian National library's digitization effort

- Full digitization of all collections
  - 1.5 mill.books
  - 50000 vols. newspapers
  - 2 mill. images
  - Norw. Broadcasting programming (radio & TV), complete since 1990 (10 years worth of programs each year)
  - 1.5 mill. periodicals
  - Music, ephemeral writing, posters & postcards..
- Time perspective 15 yrs, given financing
- Trial project 2007, 20000 volumes

# Challenges versus users

- Document representation
- Search
- Result presentation
- Granularity
- Availability

# Traditional library model

- Search interface based on users comprehending metadata
- Result presentation based on a bibliographic model (by author, by subject...)
- Description/search granularity on document level
- Accessibility (slow but) guaranteed

# User search behaviour

- Library last resort
- Google as system model
- Principle of least effort
- Little understanding of advanced search facilities
- High frequency of failed searches
- Tendency to formulate search on too general level, problems in adapting own terminology to system's

# Challenges for traditional knowledge organization

- User cannot express Own "problematic situation"
- Language is
  - Ambiguous
  - Infinite
  - under constant development
- Description rules (for cat./class.) are
  - incomplete
  - difficult to interpret
- Possible categories for description are
  - hard to define
  - of "infinite" amount
- Hierarchy vs. network



# NNL digitization project: Representation

- Full text searchable
- Scanning + OCR text generation
- Page image used for presentation, OCR text for searching
  - Vulnerable to scanner quality
  - Lacks xml and other markup

# Search interface

- Based on "Google window"
- Common access to all databases
- Retrieval vulnerable with respect to
  - quality of document representation (ocr)
  - Documents of very varying age, genre, language....
  - Collection size
  - Users' ability to choose search term

# Possible solutions to search problems

- (Automatic) combination with metadata?
  - E.g. Combination term / publication year to allow for spelling adjustment .....
- User modelling functionality
  - force / invite context description?
- Allow "folksonomic" description?

# Folksonomy example

- Tagcloud in LibraryThing for "Helge Ingstad" (archeologist who found norse viking settlements in Canada)
  - » History (8)
  - » Vikings (3)
  - » Archeology, non-fiction (2)
  - » Scandinavia, Swedish, seafaring, adventure, Canada, heathenry, america, explorers, native population, USA, documentary, unread, wishlist (1)

# Granularity

Should the DL index for searching and / or display parts of documents as well as full documents?

# INEX research results on retrieval and presentation of parts of documents

- when shown hitlist consisting of both relevant articles and article parts,
  - » Users choose full article as primary access point
  - » Users look at parts as often as they look at full articles but tend to end by looking at full article
  - » Users more easily and consistently identify relevance of full article.

# Result presentation

- According to presumed relevance
- Problems
  - Users need explanation of system reasoning about relevance
  - With so many and diverse documents, result list becomes impossible to navigate sensibly
  - Difficult to make connection between result list, which present text, and document, which is presented as image

# Accessibility

- Principle: free access to digital collection on same terms as for material in print
- Negotiation with rights management organizations
  - Author associations, publisher association, assoc.of translators, Literary rights management org.



# And then the fun begins

- 2.0 technologies
- Interactivity
- User as producer
- Communal peer review
- Open access
- .....