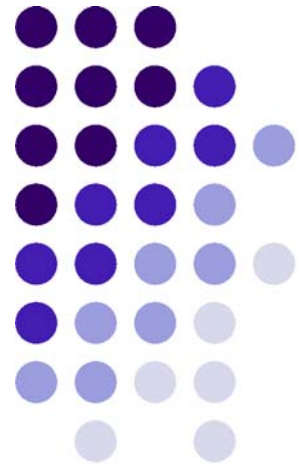# Enhancing domain-specific digital library with facilities for vocabulary exploration

Marianne Lykke Nielsen
Mathew J. Weaver
Lois Delcambre
Timothy Tolle

Outline

- Motivation
- Metadata++ model
- User evaluation
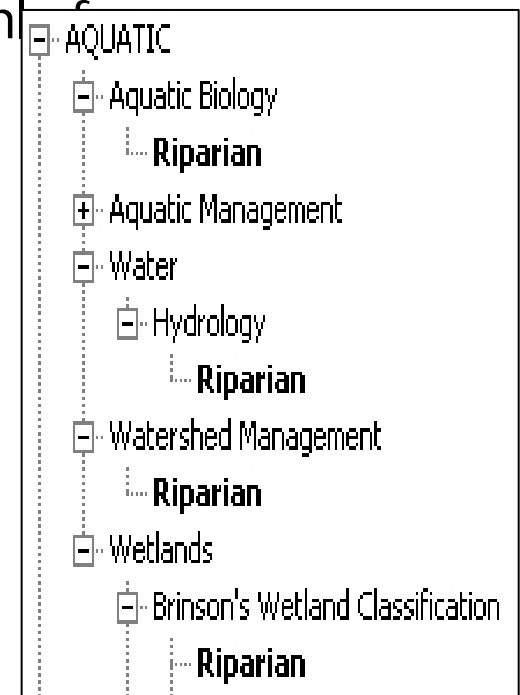- Findings
- Summing up

Motivation

- Domain-specific digital libraries are often multidisciplinary information environments that serve and involve a range of different disciplines

- Each discipline approaches the overall subject area from a specific perspective using its own special language and controlled vocabulary

- This provides indeterminism in indexing as well as searching

- System design should facilitate and encourage exploration and variety in choice of indexing as well as search terms

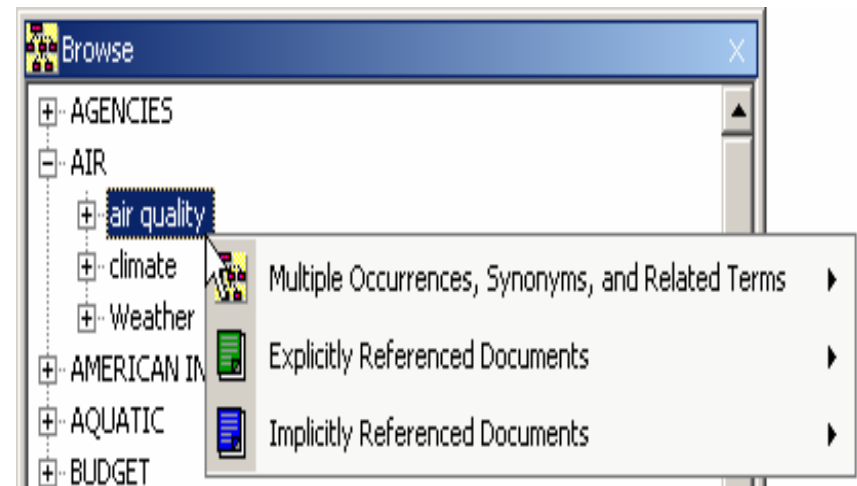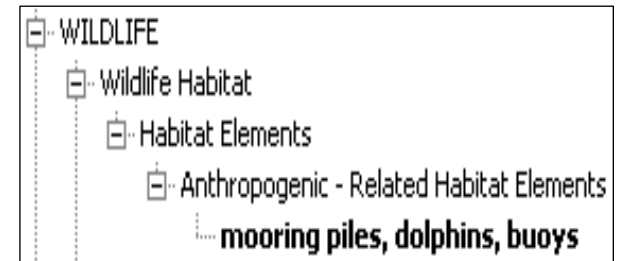- Variation in both operational and conceptual senses

Metadata++ model

- **Multiple controlled vocabularies**. We compile multiple controlled vocabularies in a hierarchical structure and allow users to index and search documents from the perspective of multiple controlled vocabularies. Each top level represent a vocabulary

- **Controlled vocabularies "as-provided"** . Only of adjustments to make the overall hierarchical structure meaningful

- **Multiple occurrences.** Polyhierarchical relationships and term meanings are not normalized. Terms may occur in multiple hierarchies and have several meanings and connotations

```
⊟ AQUATIC
   ⊟ Aquatic Biology
        ⊢ Riparian
   ⊞ Aquatic Management
   ⊟ Water
        ⊟ Hydrology
             ⊢ Riparian
   ⊟ Watershed Management
        ⊢ Riparian
   ⊟ Wetlands
        ⊟ Brinson's Wetland Classification
             ⊢ Riparian
```

Metadata++ model

- **Path-based definition.** Meaning and connotation are communicated by the term's hierarchical context

- **Polyterms.** Terms having same meaning and connotation within a controlled vocabulary is not normalized, but combined into a single polyterm


```
WILDLIFE
  Wildlife Habitat
    Habitat Elements
      Anthropogenic - Related Habitat Elements
        mooring piles, dolphins, buoys
```

- **Look-up related terms.** Right-click on term provides lists of related terms, true synonyms, and multiple occurrences


```
Browse                                              X
  AGENCIES                                          ▲
  AIR
    air quality
    climate          Multiple Occurrences, Synonyms, and Related Terms  ▶
    Weather
  AMERICAN IN        Explicitly Referenced Documents                    ▶
  AQUATIC
  BUDGET             Implicitly Referenced Documents                    ▶
```

Metadata++ model

- **Look-up related documents.** Right-click on term provides lists of manual indexed and automatic indexed documents
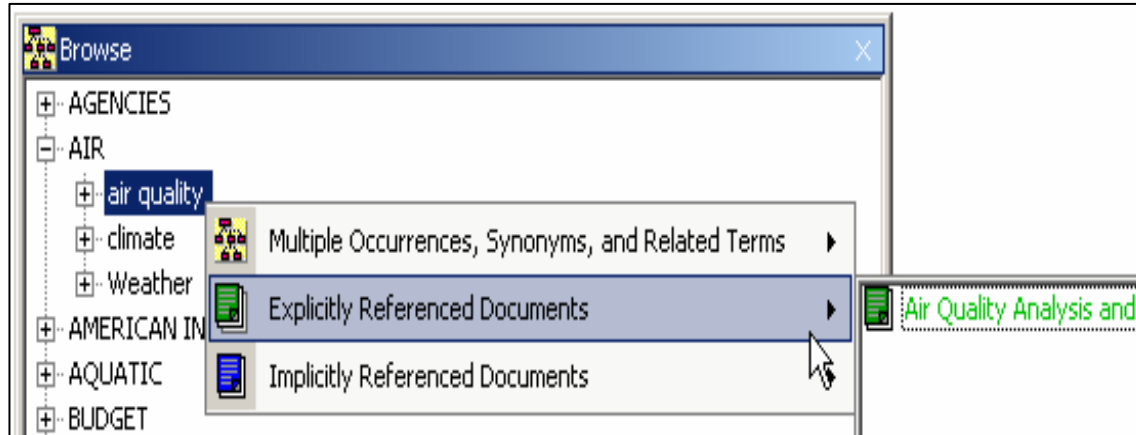
- **Query expansion.** Right click on search terms allow user to expand with descendants or related terms

User evaluation

- Usability test evaluated:
  - Users' understanding, comprehension and satisfaction:
    - Did users understand multiple vocabularies, multiple occurrences, and polyterms. Did they find the features useful?
  - User interaction and ease of use:
    - Did users explore the vocabularies, from an operational and conceptual perspective?
- Eight US Forest Service employees completed each two realistic information scenarios within a two-hour period
- Two search scenarios and two indexing scenarios
- Training session introducing the participants to the system
- Test sessions was attended by the participant, a moderator, and three observers
- Data collection: transcribed audio recording, written notes of observation, log files

Findings - searchers

- … show a high level of interaction with the system.

Actions (8 sessions)	Total	Average
Expand Term (Browsing Hierarchy)	92	11.5
Find Term (Searching hierarchy)	36	4.5
Right-click (Browse Related Terms)	36	4.5
Queries	167	20.9
Get All Descendants	8	1
View document keywords	25	3.1
Right-click (Browse documents)	352	44

- … show comprehension and appreciation of path-based terms and multiple occurrences.

- **Path-based definition**
  - "Visual reference"
  - "provides context"
  - "I do like how the path defines the term rather than a Google search which just comes up with a bunch of documents. Lot easier and better to sort through."
- **Multiple occurrences**
  - "Enlightening"
  - "Helped me think about it in various ways"
  - "Pointers to new viewpoints"
  - "Overwhelming" and "confusing"

Summing up

- Development of path-based thesaurus model that facilitate explorative searching across multiple vocabularies and multiple perspectives

- Evaluation of usability showed promising results -> model facilitates explorative searching

- Future work
  - Evaluation in healthcare domain
  - Comparison with traditional thesaurus model
  - Performance evaluation as well as evaluation of user interaction and satisfaction
  - Evaluation of semi-automated indexing based on path-based terms and document structures

Literature

Bates, M J (1986). Subject access in online catalogs: a design model. *Journal of The American Society for Information Science*, 37 (6), 357 - 376.

Borgman, C (1996). Why are online catalogs still hard to use? *Journal of The American Society for Information Science,* 47 (7). 493 – 503.

Jansen B J & Pooch, U (2001). A review of Web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technilogy,* 52 (3). 235 – 246.

Jansen, B J, Spink, A & Saracevic, T (1998). Reallife, real users, and real needs: a study and analysisof user queries on the Web. *Information Processing & Management,* 36 (2). 207 – 227.

Marchionini, G (1999). Augmenting library services: toward the sharium. *International Symposium on Digital Libraries 1999.* Available at: http://ils.unc.edu/~march/sharium/ISDL.pdf

Nielsen, M L (2004b). Task-based evaluation of associative thesaurus in real-life environment. *Proceedings of the ASIST 2004 Annual Meeting; "Managing and Enhancing Information: Cultures and Conflicts"*, Providence, Rhode Island, November 13 – 18. 437- 447.

Nielsen, M L, Delcambre, L, Tolle, T & Weaver, M (2005). Indexing and retrieval challenges in digital government systems – summary of an empirical research project. *Proceedings 2nd Scandinavian Workshop on eGov*. Copenhagen: CBS. 8 pages.

Sutcliffe, A. G. & Ennis, M. & Watkinson, S. J. (2000). Empirical studies of end-user information searching. *Journal of the American Society for Information Science*, 51, 13, 1211 – 1231.

Weaver, M J (2005). *Enhancing a domain-specific digital library with metadata based on hierarchical controled vocabularies*. Portland (OR): OGI School of Science & Engineering. Doctoral dissertation.113 pages.