



Knowledge Structures for a Domain-Specific Digital Library for Natural Resource Managers

Mathew Weaver¹
Timothy Tolle²
Marianne Lykke Nielsen³
Lois Delcambre⁴

¹ | Dept. of Computer Science & Engineering
OGI School of Science & Engineering
Oregon Health & Science University

² | Consultant, formerly with
Strategic Planning Group
Pacific Northwest Region
USDA Forest Service

³ | Department of Information Studies
Royal School of Library and Information Science

⁴ | Department of Computer Science
Portland State University

Outline

- Introduction
- Controlled Vocabularies
 - Identification and Evaluation Process
 - Motivating Issues
- Usability Tests
- Conclusion



Introduction

- Funded by the U.S. National Science Foundation as part of the Digital Government program
- In partnership with the U.S. Forest Service
- Research team covers a broad range of expertise
- Research and develop digital library of forest-related documents:
 - Environmental Impact Statements
 - Watershed Assessments
 - Scientific Reports
 - ...



Users



- Natural Resource Managers
- Domain specialists:
 - Biologists
 - Botanists
 - Climatologists
 - Geologists
 - Hydrologists
 - ...

Controlled Vocabularies: Identification



- Users identified major topics of interest:
 - Vegetation
 - Wildlife
 - Geography
 - Hydrology
 - Recreation
 - ...

Controlled Vocabularies: Identification



- Gathered existing vocabularies from various sources:
 - Textbooks and other reference books
 - Well-known organizations (USGS, Audobon Society, ...)
 - Existing datasets (geographic places)

Controlled Vocabularies: Identification



- Created new vocabularies based on:
 - Feedback from working groups of domain specialists
 - Review of keywords from existing documents

Controlled Vocabularies: Evaluation

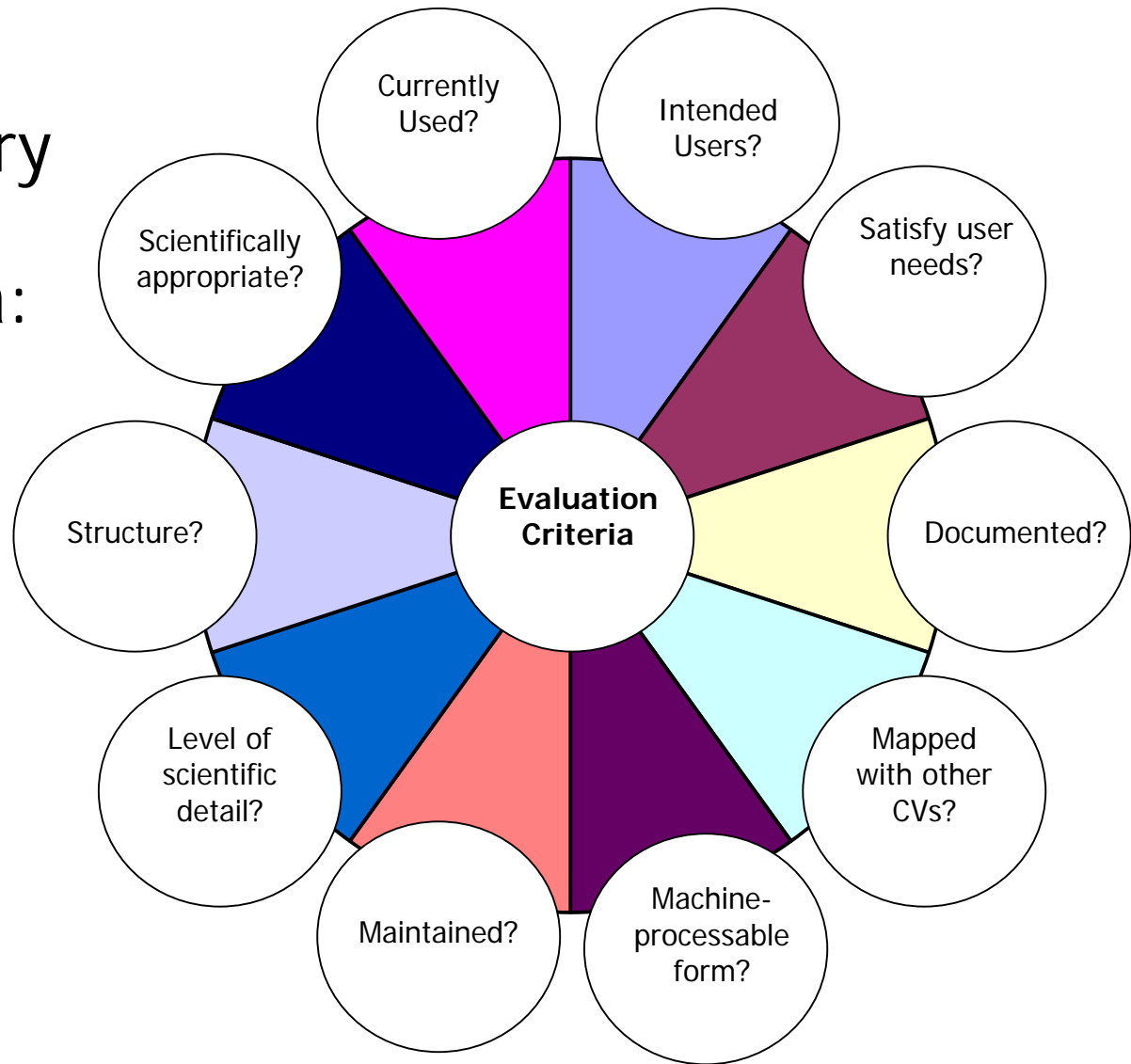


- Some topics, such as Vegetation, have numerous vocabularies available:
 - National Vegetation Classification Standard (FGDC)
 - International Plant Names Index (Royal Botanical Gardens, Harvard University Herbarium, Australian National Herbarium)
 - Integrated Taxonomic Information System (USDA, NOAA, Smithsonian, USGS, USEPA, NBII)
 - Garden Web glossary
 - National PLANTS Database (USDA, NPDC)
 - Forest Inventory and Analysis (USDA FS)
 - Pacific Northwest Ecoclass Codes for Seral and Potential Natural Communities

Controlled Vocabularies: Evaluation



- Evaluated each vocabulary based on various criteria:



Controlled Vocabularies: Evaluation



CV Name & Link	Currently Used?	Intended Users?	Satisfy User Needs?	Documented?	Mapped?	Format?
Nat'l Veg. Classification Standard	Yes	TNC, USGS, USFWS, USFS, BLM	High	High	No	Text
International Plant Names Index	Soon. Data download n/a	Scientists (botanical researchers)	Medium	High Collab. of (RBG, HUH, ANH)	Kew, Gray Card and APNI	Database
Integrated Taxonomic Information System	Yes	Scientists, researchers	High	High PLANTS database is ITIS plant data std.	Based on PLANTS, links to NBII search	ASCII delimited, XML
GardenWeb Botanical Glossary	Yes	Public (gardening community)	Low	N/A ("editors of Garden Web")	No	Database
National PLANTS Database	Yes	Researchers, public, government	High	High	No	ASCII delimited, Oracle DB [EBL1]

Controlled Vocabularies: Motivating Issues



- Maintain the vocabularies "as-is" – instead of forcing all of the terms into a single thesaurus or ontology
- Multiple occurrences of the same term
- Parallel hierarchies (codes, abbreviations, ...)
- Multiple aliases for the same term

Multiple Occurrences



- Experts from different domains use the same terms in different ways
- "temperature inversion" and "air pollution": fire fighters and climatologists both use these terms, but in different ways (different broader terms, ...)

Parallel Hierarchies



- Codes:
 - Hydrologic Unit Codes:
 - 1700010309: Quartz Creek
 - Plant Association Codes:
 - wheat grass/mulberry bush/white pine: CLD
- Abbreviations:
 - Oregon: OR
 - Environmental Impact Statement: EIA

Multiple Alias Terms



- Within a vocabulary, a single concept may be described using different words
- Example: within the Aquatic Habitat vocabulary, a partially submerged mooring may be interchangeably referred to as:
 - “mooring piling”
 - “dolphin”
 - “dead head”
- Within the vocabulary, these words are not siblings – because they describe the same thing
- Without the context of the vocabulary, these words (i.e. “dolphin”) refer to completely different things

Solution



- Multiple Occurrences:
 - let each vocabulary contain the term in all of the appropriate places
 - during indexing and searching, show the user all occurrences
- Parallel Hierarchies:
 - separate hierarchies related by “used for”
 - Combine codes and names as multiple alias terms
- Multiple alias terms:
 - Combine all aliases into one term: “mooring piling, dolphin, dead head”

Usability Test



- Will the user exploit multiple occurrences, parallel hierarchies, and multiple aliases during indexing?
- Will the user exploit multiple occurrences, parallel hierarchies, and multiple aliases during searching (interactive query expansion)?

Conclusions



- Include multiple vocabularies from various domains
- Represent each vocabulary as defined by the domain experts
- Exploit multiple occurrences, parallel hierarchies, and multiple alias terms to enhance indexing and searching