



Improving Metadata Searches by Enabling Semantic Queries

Chad Berkley¹, Shawn Bowers², Matthew Jones¹, Mark Schildhauer¹

¹ National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara

² University of California, Davis

Rationale

- Increasing numbers of digitized ecological datasets make it critically important to improve techniques for more precisely locating and delivering relevant information from scientific searches.
- Ontology-enabled search can help to increase the number of relevant search results while decreasing the number of irrelevant results, allowing users to browse a smaller and more focused resultset.

Implementation Details

- Developed prototype of ontology and annotation extensions for Metacat and EML
- Implemented metadata keyword and structured search leveraging ontologies (term expansion)
- Created annotation and ontology management services based employing open-source approaches (Jena, Pellet, OWL-DL, etc.)
- Developed a small corpus of EML, semantic annotations and corresponding OWL ontologies

Future Goals

- Enhance structured search for multiple observation types
- Increase corpus of EML documents, annotations and ontologies
- Perform precision/recall studies to compare different search strategies
- Add dataset browsing and summarization using observation and measurement annotations
- Incorporate result ranking and explanation

Types of Searches Implemented

Basic Keyword Search - Search all metadata text fields for user-entered terms.

Ontological Term Expansion - User-entered keywords are expanded using subsumption hierarchies defined in an ontology, then each expanded term is run against the metadata document base. Terms can be combined with various boolean functions.

Annotation Based Searching - Metadata fields are annotated against the structure of an ontology. The annotation is searched for user-entered keywords instead of the metadata document itself. When a result is found, any metadata document linked via the annotation is returned.

Structured Annotation Search with Term Expansion - The user selects fields to search based on the types of observations and measurements within the datasets (given by the annotations). In the search, the user specifies what was observed and what characteristics and standards were measured of it.

Anatomy of the Searches

