



Improving Metadata Search Efficiency 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

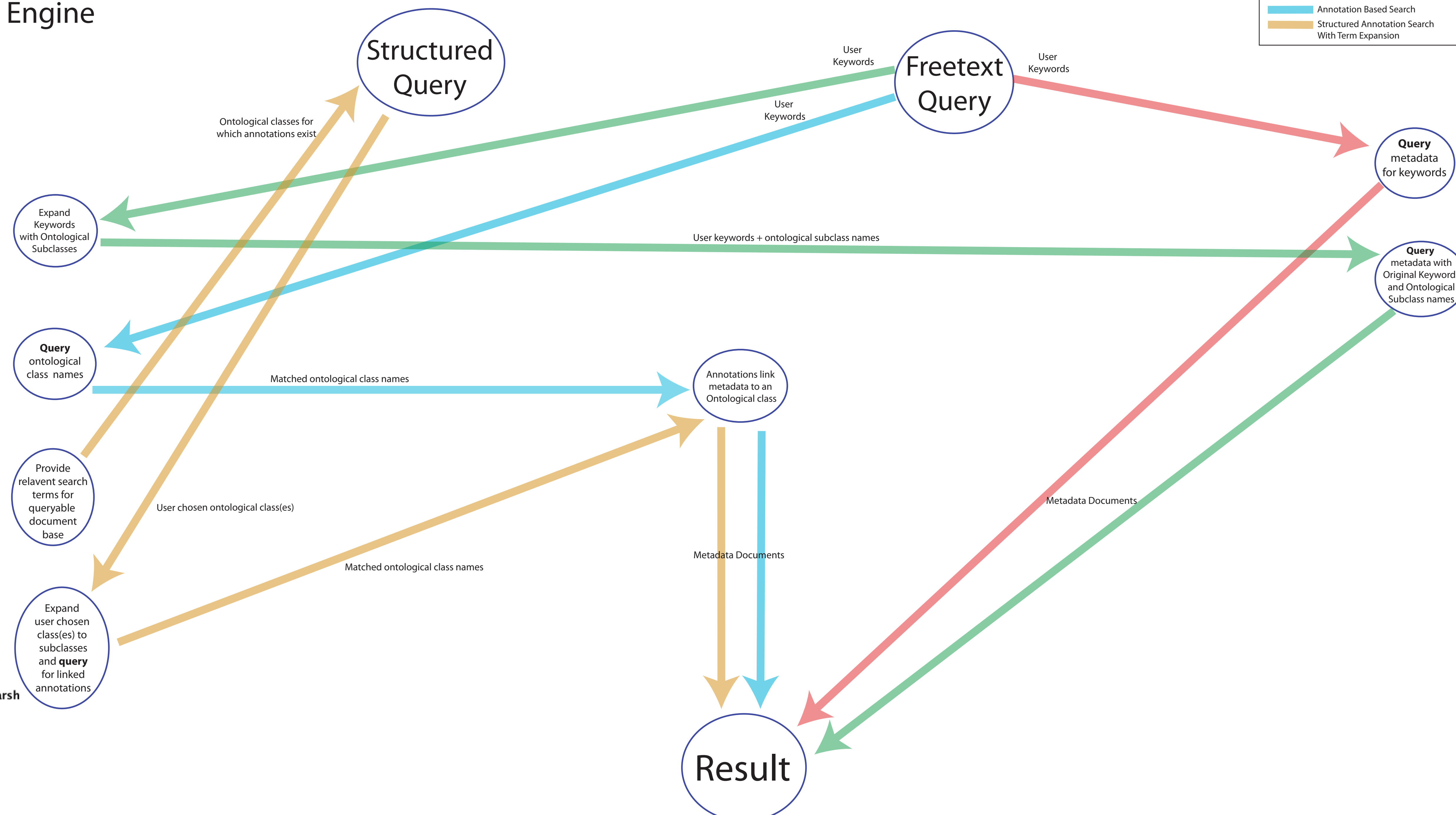
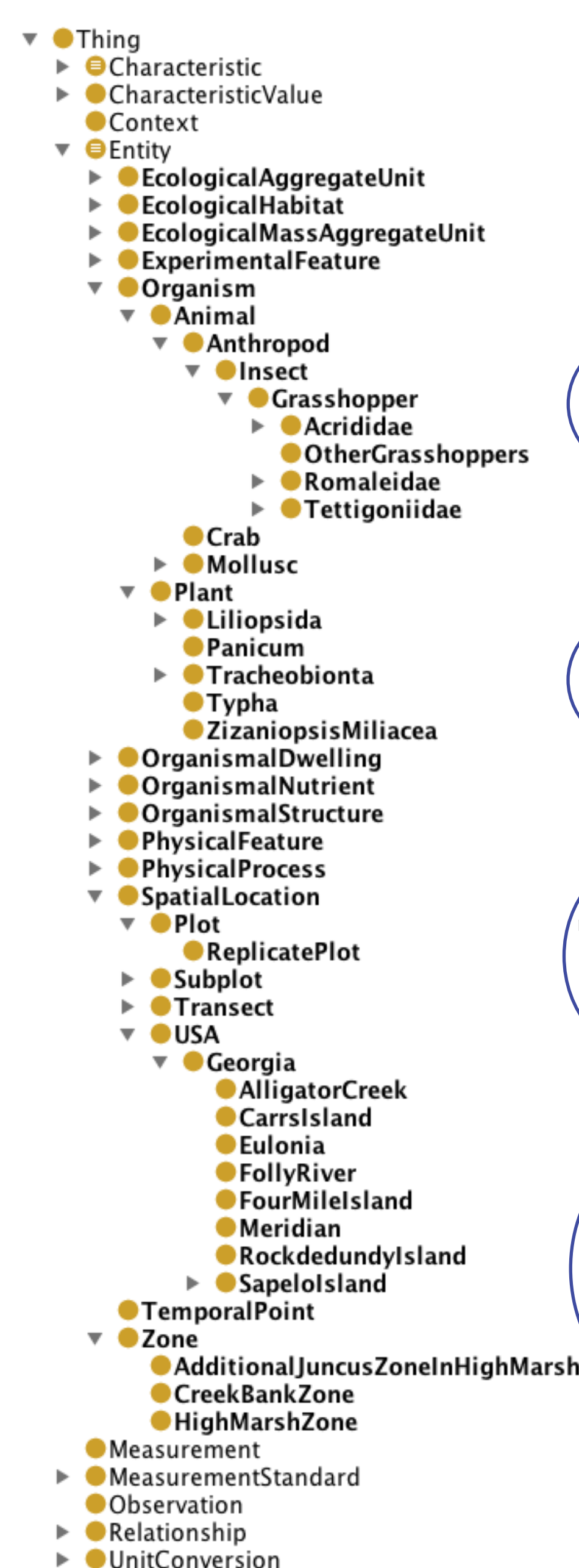
- Increasing numbers of digitized ecological datasets make it critically important to improve techniques for more precisely locating and delivering relevant information from scientific searches.
- Semantic search features increase the relevance of search results while decreasing false positive matches allowing users to browse a smaller, more relevant resultset.
- Structured browsing of semantically enabled datasets yields a more precise result which does not allow the user to choose browse parameters that do not yield results.
- Ontologies allow datasets to share a common vocabulary for more uniform searching, while not constraining scientific expression.
- Free tools make these techniques available to anyone who is interested in pursuing them.

Types of Searches

- Basic Keyword Search** - Search all metadata text fields for user entered terms.
- Ontological Term Expansion** - User entered keywords are expanded using a controlled vocabulary 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** - This search allows the user to narrow down search results by browsing ontological categories relevant to ecological datasets. As each category is chosen, a new resultset is displayed. Only categories that yield results are displayed to the user.

Anatomy of the Searches

Ontological Reasoning Engine



Metacat Metadata Repository

Biocomplexity Data Search

Home

Data Set Citation

Penning's SC. 2003. Fall 2000 plant monitoring survey -- biomass calculated from shoot height and flowering status of plants in permanent plots at GCE sampling sites 1-10.

KNB Data Registry: [um:lsid:knb.ecoinformatics.org:knb-lter-gce/92.26](http://knb.ecoinformatics.org/knb-lter-gce/92.26/knb-lter-gce/92.26/knb) ([http://knb.ecoinformatics.org/knb-lter-gce/92.26](http://knb.ecoinformatics.org/knb-lter-gce/92.26/knb))

Data Tables, Images, and Other Entities:

Metadata download: [Ecological Metadata Language \(EML\) File](#)

Data Table: [PLT-GCEM-0303a_3_0.TXT \(View Metadata | Download File\)](#)

Alternate Identifier: [PLT-GCEM-0303a.3.0](#)

Online Distribution Info:

Download File: http://gce-lter.marsci.uga.edu/lter/asp/db/dataset_details.asp?accession=PLT-GCEM-0303a

Data Set Owner(s):

Organization: **Georgia Coastal Ecosystems LTER Project**

Address: Dept. of Marine Sciences,
University of Georgia,
Athens, Georgia 30602-3636 USA

Email Address: gcelter@uga.edu

Web Address: <http://gce-lter.marsci.uga.edu/lter/>

Individual: **Dr. Steven Penning's**

Organization: University of Houston
Department of Biology and Biochemistry,
University of Houston,
Houston, Texas 77204-5513 USA

Email Address: spennings@uh.edu

Web Address: <http://www.bchs.uh.edu/~steve/>

Metadata Provider(s):

Organization: **Georgia Coastal Ecosystems LTER Project**

Address: Dept. of Marine Sciences,
University of Georgia,
Athens, Georgia 30602-3636 USA

Email Address: gcelter@uga.edu

Web Address: <http://gce-lter.marsci.uga.edu/lter/>

Associated Party:

Individual: **Wade Sheldon**

Organization: University of Georgia