Improving Metadata Search Efficiency by Enabling Semantic Queries

Chad Berkley[1], Shawn Bowers[2], Matthew Jones[1], Mark Schildhauer[1]

[1] National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara
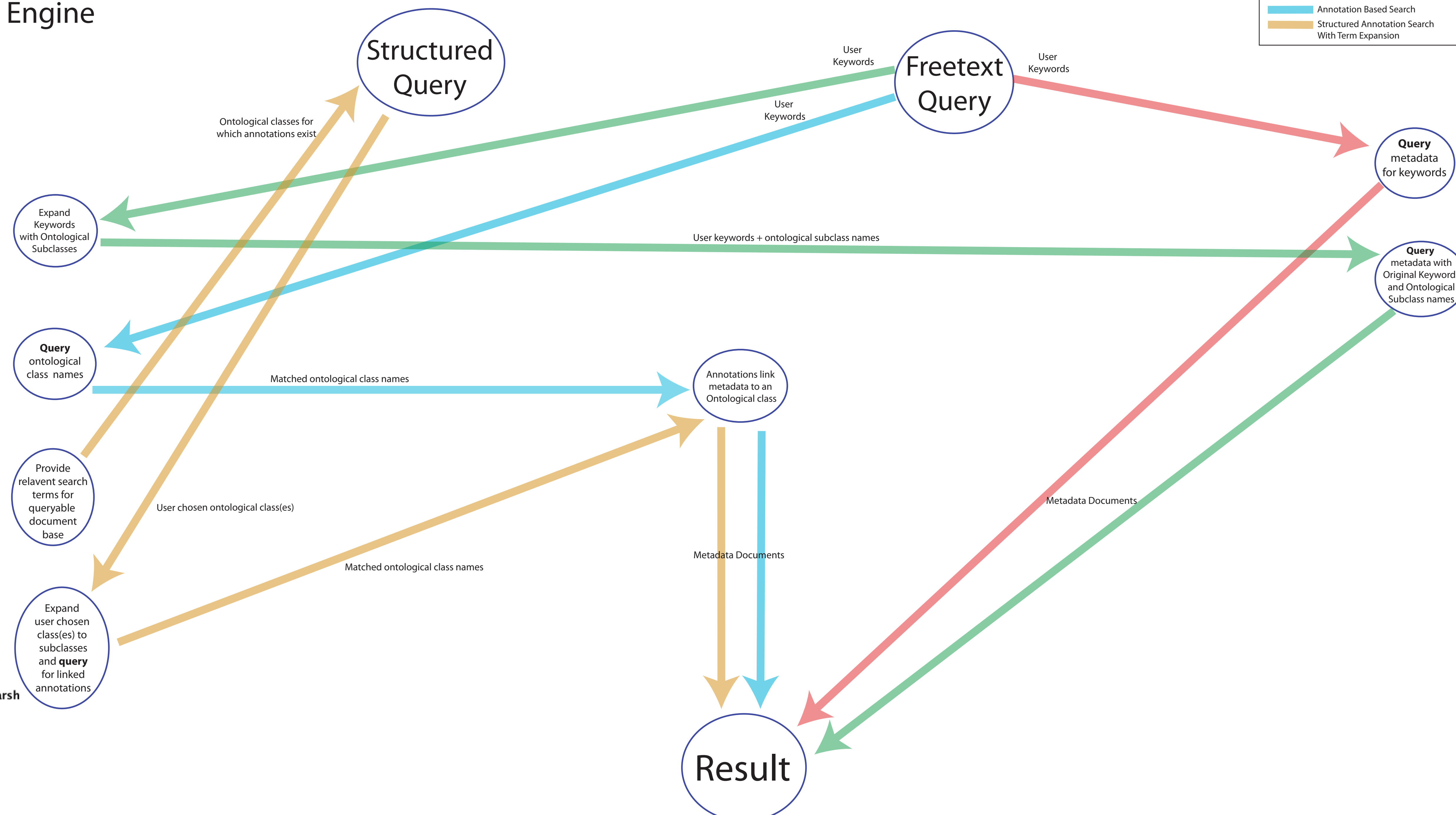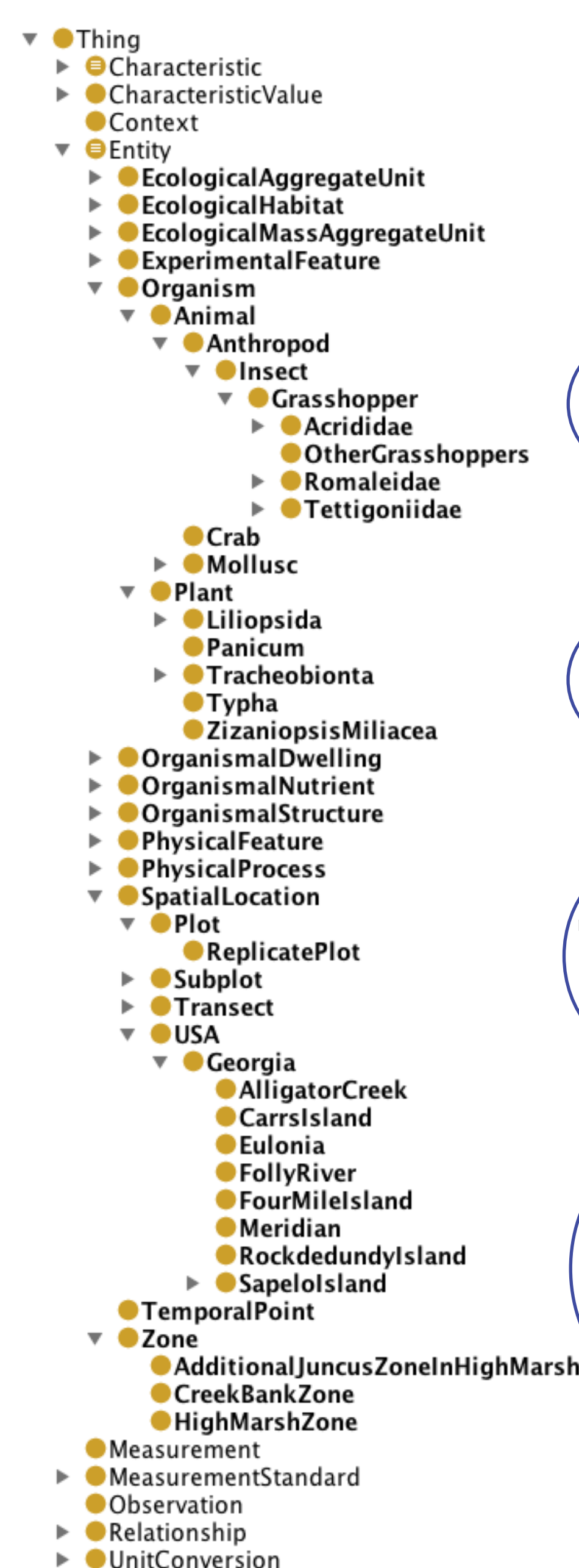
[2] University of California, Davis

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 relavence 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 persuing 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