

MULTI-SOURCE ONTOLOGY-BASED MAIZE PHENOTYPE SEARCH ENGINE

Jason Green
Dr. Chi-Ren Shyu, Thesis Supervisor

ABSTRACT

In the midst of this genomics era, major plant genome databases are collecting massive amounts of heterogeneous information, including sequence data, gene product information, as well as images and descriptions of mutant phenotypes. While basic browsing and search capabilities are available to allow researchers to query and peruse the names and attributes of stored data, advanced search mechanisms that can take advantage of textual descriptions of various types of stored data are nonexistent. Furthermore, though much time and effort have been afforded to the development of plant-related ontologies, the knowledge embedded in these ontologies remains largely unused in available plant search mechanisms. Addressing both of these issues, we have developed a unique search engine for the phenotypes in MaizeGDB. This advanced search mechanism exploits the content and structure of available domain ontologies for the purposes of query enrichment, with currently both the Plant Ontology and Gene Ontology being utilized. The search engine also has the flexibility to integrate various text description sources to aid the user in retrieving desired phenotype information. This framework can be generalized to any domain with a domain-specific ontology or to sets of text sources that are heterogeneous and interconnected.