

Lukasz Ziemba - University of Florida

Ontology-based Information Extraction in a Digital Library for Water Conservation

The Ph.D. project is a digital library management system in the domain of water conservation. The ontology was developed in order to manage a set of interrelated library keywords (ontology concepts) comprising a water conservation taxonomy, as well as various types of library resources, including publications, technical reports, models, data sets, authors, organizations and more. Each concept provides a number of terms and definitions in multiple languages, ontology-assisted navigation and search allow for viewing related concepts and their associated resources.

Information extraction techniques are being investigated to facilitate searching, navigation and adding of new information. The techniques include statistical methods, data mining and natural language processing (NLP).

Clustering techniques facilitate discovery of new concepts from the collection of publications, NLP techniques suggest concepts relevant to a new publication being added to the library. The digital library supports multiple languages and follows the Open Archives Initiative (OAI) standard for sharing metadata with other digital repositories. A web-based interface for the digital library is available at <http://library.conservefloridawater.org/>. The ontology is implemented using the Lyra ontology management system (OMS) developed at the University of Florida, in order to provide database management functions (secondary storage, query processing using ontology reasoners, authoring tools) to support development and application of large scale ontologies.