Using the Semantic Web to Integrate Ecoinformatics Resources

Cynthia Sims Parr csparr@umbc.edu
Andriy Parafiynyk andr1@umbc.edu
Joel Sachs jsachs@umbc.edu
Rong Pan pan.rong@umbc.edu
Lushan Han lushan1@umbc.edu
Li Ding ding.li@umbc.edu
Dept. of Computer Science and Electrical Engineering
University of Maryland Baltimore County
Baltimore, MD 21250 USA

Tim Finin
Dept. of Computer Science and Electrical Engineering
University of Maryland Baltimore County
Baltimore, MD 21250 USA
finin@umbc.edu

David Wang
Dept. of Computer Science
University of Maryland
College Park, MD, 20742 USA
Tw7@cs.umd.edu

Abstract

We demonstrate an end-to-end use case of the semantic web's utility for synthesizing ecological and environmental data. **ELVIS** (the Ecosystem Location Visualization and Information System) is a suite of tools for constructing food webs for a given location. ELVIS functionality is exposed as a collection of web services, and all input and output data is expressed in OWL, thereby enabling its integration with other semantic web resources. In particular, we describe using a Triple Shop application to answer SPARQL queries from a collection of semantic web documents.

1. Introduction

SPIRE (Semantic Prototypes in Research Ecoinformatics - http://spire.umbc.edu) is a distributed, interdisciplinary research project tasked with building semantic web prototypes for invasive species science.

Our demonstration focuses on ELVIS (the Ecosystem Location Visualization Information System), a suite of tools motivated by the belief that food web structure plays a role in the success or failure of potential species invasions. Because very few ecosystems have been the subject of empirical food web studies, response teams are typically unable to get quick answers to questions like "what are likely prey and predator species of the invader in the new environment?" The ELVIS tools seek to fill this gap.

All data that we produce is expressed in OWL via a collection of ecological and evolutionary ontologies. This, together with our service-oriented architecture, enables much flexibility in integrating with other semantic web applications

In particular, we use our SPARQL query engine Triple Shop, which allows a user to specify SPARQL queries over arbitrary collections of semantic web documents, together with a reasoning level. Thus, we are able to integrate and reason over diverse ecoinformatics data in response to ad-hoc queries.

1.1 Related Work

Previous work on data integration in ecological informatics includes online data repositories [2] and workflow [4] ontologies. Individual food web researchers maintain and share their own digital data archives, in individualized data formats, though more accessible standardized archives are beginning to emerge [1]. There are good databases on invasive species (e.g. http://www.issg.org/) but they are not automatically integrated with information about non-invasive species with which they interact. To our knowledge, there does not exist web-based support for modeling an invasive species.

The Joseki SPARQLer (http://www.sparql.org/query.html) service was the model for our Triple Shop. We have added a number of reasoning capabilities, from simple subsumption to OWL.

2. ELVIS

The task of providing food web information for a user-specified location breaks into two distinct problems: constructing a species list for a given location; and constructing a food web from a given species list (and habitat information). Our demonstration focuses on the latter of these problems.

2.1 The Food Web Constructor

The Food Web Constructor (FWC) uses empirically known food web links to predict food web links not yet recorded. A user can choose which food web studies to use for prediction or exclude from 257 datasets we compiled from previously digitized literature. Taxonomic and phylogenetic distances are then used to weight evidence supporting or failing to support links between the inputted taxa.

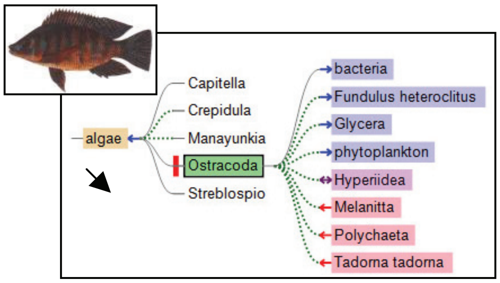

Figure 1. Nile Tilapia, an invader in Florida ecosystems, is predicted to eat algae and have no potential predators. Organisms predicted to be nearby in the food web (to the right of algae) could be impacted by or mediate the introduction of this competitor.

Taxa can be entered several different ways: simple text lists, XML files, or food web number. In this latter case we seek to reconstruct feeding links based on the rest of the database and can therefore assess the success rate of the different algorithms or model parameters.

Each suspected link is reported, together with references to supporting evidence. Summary statistics of the food web are also reported.

2.2 Evidence Provider

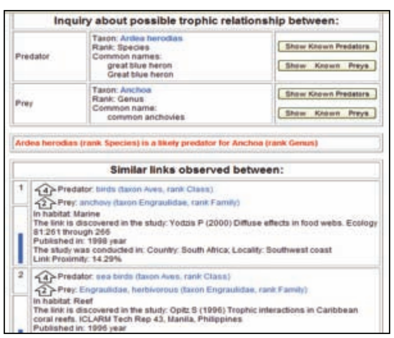

Figure 2. The Evidence Provider shows the evidence for a predicted trophic link.

The Evidence Provider allows a user to drill down on a specific link to see the evidence for and against it. This includes actual observed links, the study in which they were published, and the relationship between the species in the observed link and the predicted link.

3. The SPIRE Triple Shop

The Triple Shop is a SPARQL query engine that builds on the Joseki web service. Users specify SPARQL queries over arbitrary collections of semantic web documents, and a level of reasoning (from none to OWL-Full). Originally developed as a component of our Swoogle search engine [3], the Triple Shop is now a stand-alone application.

3.1 Using the Triple Shop to Integrate Food Web and Natural History Data

The Triple Shop illustrates the potential of the semantic web to support rapid querying of distributed scientific databases for a variety of scenarios. For example, Figure 3 shows a query that selects data from two ontologies - ETHAN (which represents evolutionary trees and natural history information) and SPIREEcoConcepts (which represents food web relationships) – and from the ELVIS database to determine known predator-prey relationships among an invader and a specific group of native species in a particular habitat, as reported in previous studies. (The ETHAN ontology itself represents an integration of taxonomic and phylogentic information from

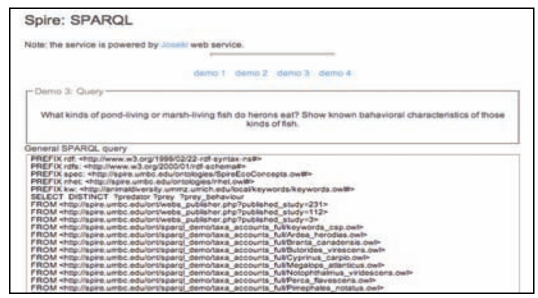

Figure 3. A Triple Shop query synthesizes data from multiple sources.

4. Acknowledgements

This research was supported by NSF ITR 0326460 and matching funds received from USGS Nat. Bio. Information Infrastructure.

References

[1] Dunne, J. A. The network structure of food webs. In: Ecological Networks: Linking Structure to Dynamics in Food Webs, eds. Pascual, M. and Dunne, J. A. Oxford University Press, 2005.pp. 27-86.

[2] Jones , M. B.; Berkley, C.; Bojilova, J.; Schildhauer, M. P. 2001. Managing scientific metadata. IEEE Internet Computing. Vol: 5(5). Pages 59-68.

[3] Li Ding *et al.*, "Swoogle: A Search and Metadata Engine for the Semantic Web", Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management, November 2004.

[4] Ludaescher, B. et al. 2004. Scientific workflow management and the Kepler system. Concurrency and Computation: Practice and Experience.