Modeling and Publishing Biological Names and Classifications on the Semantic Web

Nina Laurenne, Jouni Tuominen, Mikko Koho and Eero Hyvönen
Semantic Computing Research Group (SeCo)
Aalto University, Department of Media Technology
P.O. Box 15500, FI-00076 Aalto, Finland
University of Helsinki, Department of Computer Science

Abstract
Periodic changes characterize the scientific naming system. As a result, the biggest challenge lies in ascertaining the actual meaning of names, when multiple taxonomic concepts are associated with them. This makes it hard to integrate biological information from different sources, such as publications, online databases, and museum collections, and search for it. On the Semantic Web, the problem can be approached by representing taxa, checklists, and their relations as ontologies that are decipherable for machines. Our goal is to establish a centralized ontology repository of biological names and classifications in Finland.

We have developed an ontology model for biological taxon names and the ONKI Ontology Service for publishing it as a service for humans and machines. The aim is to achieve a practical and maintainable name system for researchers, environmental authorities, and amateurs for finding biological names to use, for indexing content correctly and cost-efficiently using ontology services, and for laying out a foundation for making heterogeneous biological content interoperable in applications.

The model consists of three parts on the basis of the elaborateness of taxonomic information and the needs of the users. The parts are maintainable independently, but associations between them are possible. (1) Scientific names and taxonomic concepts are treated separately and detailed taxonomic information can be associated with them. The processes of systematic research that are relevant to changes of taxon names or concepts such as descriptions of new taxa, splitting and lumping of taxa are conceptualized. (2) Checklist-type information is supported; names occurring in different checklists, but referring to the same taxon can be linked. (3) Also, vernacular names in multiple languages including dialects are supported. The ontology model covers temporal dimensions, which make taxon names traceable to reveal conflicting taxonomies and competing views. The possibility to connect imprecise taxonomic knowledge to precise information allows for versatile and flexible name management for users with different needs. Results of queries do not only return a currently valid/accepted name but lead the user to the source of the information. Ontology-based queries enable the retrieval of relevant contradictory information, which is an important feature for scientists.

We have two use cases regarding beetles to demonstrate the usage of the name ontology. (1) The cerambycid beetle names of five Finnish checklists from the years 1936-2010 are linked. The ontology is applied to the observational data approximately from the same period. The geographic information on the observational data is then disambiguated using Finnish Spatio-temporal Ontology (SAPO). By applying the name ontology and SAPO we can explore the distribution of the cerambycid beetles in the time-scale without extensive data harmonisation. (2) The other use case is nine genera of Afro-tropical eucnemid beetles and their chaotic classification. The model will be tested with the pilot group, in which splitting and lumping are common and names have changed for various reasons. For example, at least eight taxonomic concepts are associated with the genus

Pterotarsus. This data is challenging as it includes study results, mistakes, and various nomenclatural changes.

Acknowledgements
This work is part of the national FinnONTO program and is funded by the Finnish Funding Agency for Technology and Innovation (Tekes) and a consortium of 38 public organizations and companies. Fruitful co-operation with Jyrki Muona, Hans Silfverberg, Hannu Saarenmaa, Leo Junikka, and the Finnish Museum of Natural History is acknowledged.

URLs
Semantic Computing Research Group: http://www.seco.tkk.fi/
National Semantic Web Ontology Project in Finland (FinnONTO), 2003-2012: http://www.seco.tkk.fi/projects/finnonto/
Biological Ontologies and Vocabularies: http://www.seco.tkk.fi/ontologies/biology/
ONKI Ontology Service: http://www.onki.fi
Finnish Spatio-temporal Ontology SAPO: http://www.seco.tkk.fi/ontologies/sapo/