

Publishing Biological Classifications as SKOS Vocabulary Services on the Semantic Web

Jouni Tuominen, Matias Frosterus, Nina Laurenne 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

Taxon names may refer to more than one taxon and a taxon may have multiple names, which makes information retrieval and data integration problematic. On the Semantic Web, taxon names with unambiguous URIs can be collected into controlled vocabularies or ontologies, which enable the sharing of information in an interoperable way. For example, the observational data of birds can be annotated using these vocabularies. The vocabularies may contain relations between taxa, which can be used for further enhancing information retrieval. For instance, a user interested in the ecology of carnivores is possibly interested in the ecology of cats, too.

We have used the SKOS (Simple Knowledge Organization System) data model to represent a taxonomic hierarchy in RDF (Resource Description Framework). The basic unit of the SKOS model is a concept, which is used for representing taxa that are ordered into a single classification. The vocabulary contains information of the taxa, e.g. their taxonomic ranks, scientific names, and common names.

The taxonomic hierarchy is modeled by using the hierarchical *skos:broader* relation. For example, the genus *Felis* is included in the subfamily Felinae. The preferred scientific and common names of the taxa are represented with the property *skos:prefLabel* and alternative names with *skos:altLabel*. The authorship information of a taxon is defined with the property *skos:note*, and the property *rdf:type* is used to indicate the taxonomic rank. We have extended the SKOS data model by introducing the property *creator*, which states the organization that has created the data, and *linkToWikipedia*, which provides the user with additional information about a taxon in Wikipedia. Taxa are referred to by using unique URIs that point to the location of the information describing the object on the web.

The model is demonstrated with the worldwide checklists of mammals (4,629 species) and birds (9,300 species). These checklists are extensive and contain the vernacular names in Finnish, Swedish and English, making them useful for a wide audience.

Once a taxonomic checklist has been represented in SKOS, it can be published instantly in the ONKI Ontology Service. The ONKI service provides a SKOS vocabulary browser for the human user and ready-to-use web widgets, and application interfaces (API) for applications. These components enable browsing, querying and visualizing of vocabularies, thus supporting use cases such as content indexing, taxon name disambiguation, searching, and query expansion.

The ONKI SKOS browser consists of three main components: 1) taxon name search with semantic autocomplete, 2) hierarchy and 3) properties of taxa. When typing text to the search field, a query is performed to match the taxon names. The result list shows the matching names that can be selected for further examination. When a name is selected, the classification is visualized, and the properties are shown.

The taxonomic data can be maintained and edited with standard tools supporting the SKOS data model, such as the ontology editor Protegé 4 with the SKOSED plugin and the SAHA metadata editor.

At the moment, new taxonomic checklists (over 80,000 species) of the Finnish Museum of Natural History are being published in ONKI. These vocabularies are integrated with other ontologies using the national Semantic Web ontology infrastructure FinnONTO.

Acknowledgements

This project is part of the national FinnONTO program funded by the Finnish Funding Agency for Technology and Innovation (Tekes) and a consortium of 38 public organizations and companies.

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/>

SAHA – Browser-based Semantic Annotation Tool: <http://www.seco.tkk.fi/services/saha/>