

Radiation of beetles into cyberspace – two case studies of modelling taxonomic information

Laurenne, N., Tuominen, J. and Hyvönen, E.

The amount of biological information has increased during the last decades. Enormous amounts of information is hidden in museum collections and literature, and according to the current trend, this is being digitized and released for public use. However, the availability of this information does not guarantee its usability because biological vocabulary may be loaded by semantic changes over time. For example, scientific names of organisms undergo numerous changes since the publication of their original taxon description. Changes in taxon names are the biggest obstacles for data integration and they impede interoperability of information systems and human communication. Also, searching information may be tricky if the meaning of the name has changed into another one.

In taxonomy, there is no single phylogenetic tree or classification of organisms but multiple ones which overlap with each other and change over time. Typically a name combination alters as a result of a species being shifted into a new genus, which in turn redefines the limits of the old and new genus and re-organises the classification. Often redefinitions of taxon boundaries are taken into account only by taxonomists, but the wider audience encounters them in the form of changed taxon names and new classifications of taxa.

We present two case studies in which taxonomical information of two beetle groups have been modelled according to the TaxMeOn metaontology. The first case study shows how the relations of cerambycid beetles are cross-mapped between species lists and the second one describes how an elateroid beetle genus has been reclassified several times in the last two hundred years. These two cases differ from each other as the former is based on the limited information that is available in species lists and the second one is based on taxonomic literature and the interpretation of it by several researchers.

We provide examples of how challenges are encountered when modelling intuitively self-evident information and applying it at a practical level of name management. The use cases show the misleading nature of scientific names of organisms and raise the importance of linking information on the basis of their meanings when querying, indexing and combining data in a digitized world.

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. Co-operation with Jyrki Muona and Hans Silfverberg, 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://onki.fi/>