Observing observations – an ontology-based approach for improving the reliability of biodiversity data

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

The Envirofi program aims to provide methods for gathering, managing and offering real-time environmental information on mobile devices and computers. The project seeks to develop methods for combining heterogeneous information from scattered sources and to enhance the assurance of data. A huge amount of biodiversity data is produced with different vocabularies and terminologies. This hinders data integration. We approach the problem from a semantic point of view by developing ontology models that enable the harmonising of biodiversity-related information in order to improve the quality of the data.

The number of observations of plants and animals is rising rapidly, and this has raised data quality into an important issue. Common to all observations is that organisms are referred to with a name. However, names are not reliable identifiers because they change over time, mainly due to redefinitions of taxon boundaries. As a result, even the regional species lists of the same country may have differing taxon names, which does not only cause confusion, but requires a time-consuming harmonisation of names when integrating data. Creating relations between taxon names allows users to use the names that they are the most familiar with. The metaontology TaxMeOn (Taxonomic Meta-Ontology) is applied for heterogeneous datasets allowing users to enter data and search information with their preferred taxon names. Common names with multiple languages including dialectical expressions are also supported.

Users are categorised on the basis of their a priori knowledge, i.e. novice nature hobbyists are provided interactive guidance for recognising organisms, while advanced hobbyists need more precise information about distribution, ecology and seasonal changes. The most typical misidentifications are eliminated by offering key characters for differentiating taxa that resemble each other. Researchers benefit as they can check e.g. details of biology and distribution of organisms in the field. The information provided by researchers is used as a validation for other users, and professionals can also corroborate occurrence data offered by other users if they wish.

The data validation is based on a comparison to other occurrence records in the same area, regional species lists whose names are cross-mapped with other lists and a comparison to the known phenology of an organism. Information about associations between species can be used for evaluating the reliability of an observation. For example, a presence of one species may be dependent of the presence of another species.

Ontologies enable gathering diverse information about nature into an interoperable form in which data can be analysed without harmonisation. Users can check how coherent their own observations are with advanced hobbyists and researchers, and data can be filtered according to its reliability.

The ENVIROFI - The Environmental Observation Web and its Service Applications within the Future Internet project is funded by the European Comission under the FP7 Future Internet Call.

URI
http://www.envirofi.eu/