

ONTO-ToolKit: enabling bio-ontology engineering via Galaxy

Erick Antezana, Aravind Venkatesan, Vladimir Mironov and Martin Kuiper

Department of Biology, Norwegian University of Science and Technology,
Høgskoleringen 5, N-7491 Trondheim, Norway.

venkatesan@nt.ntnu.no

Project page: <http://bitbucket.org/easr/onto-toolkit/wiki/Home>

Open Source License: [GNU General Public License](#)

<http://bitbucket.org/easr/onto-toolkit/src/tip/README>

Abstract

Biological data integration is a corner stone for systems biology approaches. Data integration is supported by a diverse series of tools, but still the lack of a consistent terminology to label these data presents significant hurdles, causing much of those biological data to remain disconnected or worse: to become misconnected. Bio-ontologies are being developed as a means to overcome those terminology issues. OBOF, RDF and OWL are among the most used ontology formats to capture terms and relationships in the Life Sciences. The Semantic Web also promises to support data integration and further exploitation of integrated resources via automated reasoning procedures.

Here we present ONTO-ToolKit, which is an extension to the existing PERL suite ONTO-PERL (<http://search.cpan.org/dist/ONTO-PERL/>), supporting the handling of OBO-formatted ontologies. ONTO-ToolKit is distributed as a set of Galaxy (<http://galaxy.psu.edu/>) tools. It provides not only a user friendly interface, via Galaxy, to manipulate OBO ontologies but also opens up the possibility to perform further biological (and ontological) analyses by using other tools provided within the Galaxy platform. Moreover, it provides some tools to translate OBO-formatted ontologies into Semantic Web formats such as RDF and OWL. Finally, it provides an interface (currently under development) to launch SPARQL queries from within Galaxy.

We present a couple of use cases to illustrate how the functionality of ONTO-PERL could be combined with the functionality of other tools from Galaxy. One of the use cases illustrates the functionality of ONTO-PERL to identify all the upstream terms (ancestors) of a particular Gene Ontology (GO) term to extend the information contained in the Gene Ontology Annotation (GOA) files. The other example illustrates how ONTO-ToolKit can be used to identify overlapping annotations for a given pair of proteins. Such an overlap would suggest that these proteins share a bio-molecular function, cellular localization or biological process.