BOWiki – A collaborative environment for editing biomedical ontologies

Hoehndorf R[1,2,3], Kelso J[3], Visagie J[3]
[1]*Institute for Informatics, University of Leipzig, Germany*
[2]*Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Germany*
[3]*Departement of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Germany*

Background As biomedical ontologies such as the Gene Ontology[1] increase not only in size but also in importance, quality assurance of biomedical ontologies becomes an important issue. We identify eight basic criteria on which the quality of biomedical ontologies can be evaluated: correctness, consistency, completeness, expressiveness, definitions, stability, simplicity and ontological foundedness. Many of these criteria are at least partially social criteria or can be improved by including the user community of the ontology in the curation process. Consistency and correctness are measures for how appropriate the ontology captures the knowledge or whether there are contradictions either within the ontology itself (consistency) or to the scientific theory of the domain (correctness). Completeness refers to the question how much and how fast knowledge of a domain is covered. These, together with definitions (which are created, maintained and accepted in a community), and simplicity (which must be measured against the background of the users) are social criteria and in order to judge the quality of an ontology using the criteria we mentioned, a reference to a community is necessary. Preferably, the community is not only asked for input, but is directly involved in a collaborative effort of creation and maintenance of the ontology.
A wiki [4] is an approach for collaborative editing and maintenance of a website. Anyone visiting a site is allowed to add, delete or change all content in an easy and quick way. A *semantic* wiki is a wiki which has a formal representation model of the knowledge contained within. Most of the existing semantic wikis are based on the formalisms OWL[5] and RDF[6]. We investigate methods for implementing semantic wikis for use in ontology curation in biomedical applications.

Methods We extend the semantic model to make it suitable for curation of biomedical ontologies in the following ways: enhancement of the user-interface; inclusion of a content-rating system; support for versioning and individual tracks; replacement of the knowledge representation formalism (currently on OWL) by a formalism supporting non-monotonic reasoning; foundation in a top-level ontology.

The user-interface of a wiki is unlike the user-interface used in the curation of biomedical ontologies. It is text- or content-centered, while the curation of biomedical ontologies uses structure-centered views and tools. Since the underlying structure of current semantic wikis is based on RDF or OWL, representing the knowledge in a graph-like format is possible. For more complex representation formalisms such as those based on first order logic, conceptual graphs or similar visualizations can be used.

Additionally, we include a content rating system not only for the tradiational, textual content of the wiki, but also for its semantic or formal structure. The content-rating is usable by all users of the wiki. Rating is possible for every piece of information, including ratings themselves. It is possible to extract layers of the ontology above a certain rating in order to obtain version of the ontology with different confidence values.

In science, it may be the case that there are different, incompatible views on an area of knowledge. Therefore, we include the possibility for the supporters of this minority view to adjust the ratings for themselves in a way which reflects their view on the domain. They may then choose to share their views with other users. This possibility is provided in addition to a global, standard view on the ontology.

The logic formalism provided by OWL and RDF, used as the foundation of most semantic wikis, has the drawback of allowing only monotonic inference. While most human reasoning is monotonic, some inferences we make are non-monotonic. We investigate the use of Situation Theory And Situation Semantics[3] (STASS) as the underlying formalism. STASS provides not only the possibility for monotonic and non-monotonic reasoning, but also for the representation of context and granularity.
Finally, we plan to use this wiki in the curation of ontologies in medicine and biomedicine, but semantic wikis do not provide the facilities to reference existing, upper categories of these domains. Therefore, the extensions of a semantic wiki which we will provide will use a biomedical core ontology based on the top-level ontology GFO[3] and allow the users of the wiki to reference concepts of this top-level.

Results We believe that our work, when finished, can contribute to keeping large ontologies correct and complete, without much effort by a few curators. Furthermore, it allows for the annotation of gene products, for example their functions, by the people which find out about these, requiring minimal effort on their side. This will lead to a more complete resource of which everybody can benefit[7]. Furthermore, it makes ontology and information extraction by literature mining amenable. Currently, these methods generate huge amounts of data with very low confidence. Due to the collaborative and open nature of a wiki, the data generated by these extraction methods can be reviewed not only by a few people, but by a large community. Overall, we believe that the community-centered approach, as in the wiki which we are developing, will ultimately lead to more recent, more correct, more complete and easier understandable knowledge in biomedical ontologies.
We started the implementation of an ontology curation framework based on the principles described here. Our work is based on a Semantic Media Wiki. As a first step, this wiki is extended by the concepts of the top-level ontology GFO. From this, a biomedical core-ontology is derived in this wiki. The first use-case is an ontological model of functions.

References

[1] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1):25–29, May 2000.
[2] Keith Devlin. *Logic and Information*. Cambridge University Press, 1991.
[3] Barbara Heller and Heinrich Herre. *Ontological categories in GOL.* Axiomathes, 14(1):57–76, 2004.
[4] Bo Leuf, Ward Cunningham: *The Wiki way: quick collaboration on the Web.* Addison-Wesley, London 2001.
[5] Deborah L. McGuinness and Frank van Harmelen. *OWL Web Ontology Language.* http://www.w3.org/TR/2004/REC-owl-features-20040210/
[6] Frank Manola and Eric Miller. *RDF Primer*. http://www.w3.org/TR/2004/REC-rdf-primer-20040210/, 2004.
[7] Kai Wang. *Gene-function wiki would let biologists pool worldwide resources.* Nature 439, 2006.