

Ontology Views for Collaborative Ontology Creation: The BioSphere Portal

Stuart Aitken and Jonathan Bard
{s.aitken, j.bard}@ed.ac.uk
Informatics, University of Edinburgh,
Appleton Tower, Edinburgh EH8 9LE, UK.

The BioSphere portal is being designed to support the collaborative creation of biological ontologies by expert biologists. While existing web-based portals allow users to view bio-ontologies, and to search over them, support for ontology development still relies on stand-alone ontology editors for creating new ontology versions, and private email, email lists and/or Wikis for the distribution of ontology files and discussions about their contents. These approaches are ineffective in their use of the available storage, versioning and visualisation techniques. They also discourage the adoption of new methods such as the web ontology language (OWL), as users must learn new tools (or plugins) to cope with the new syntax, before they can even begin to appreciate pros and cons of OWL semantics and reasoning.

To better support collaborative development and ontology integration, we are implementing a Grid-based portal that provides storage, visualisation and editing functions for all community members to the centrally-held ontology that they are constructing. The user community is typically a loosely-organised group of expert biologists who participate in the ontology development process remotely. The system will provide each user with their own view of the ontology which will be created on-the-fly from the single source OWL/XML document that is, in fact, shared by the entire group. For each user, the system will automatically manage the edits they make and support versioning. That is, after editing their view of the ontology, the modifications will be saved in the central source document as updates or deletions in an efficient manner, so maintaining full versioning. The use of an XML database will allow search over the contents, thus providing the basis of resource integration.

Ontology Views: Informally, the user's *view* can be thought of as a perspective or a viewpoint on the shared ontology. More technically, *database* and *XML views* can be distinguished from *ontology views*. A *database view* is generated by a query over the source database to produce a result (a table) that is appropriately organised for the task in hand. The definition of an *ontology view* is less well established. It has been claimed that an ontology view should result in a new (smaller) ontology derived from, but independent of, the source ontology (Bhatt, 2004). Alternatively, the *ontology view* might be the set of terms and definitions within a certain radius of a selected term, a set which does not itself constitute an ontology but a connected sub-graph (Noy, 2004; Voltz, 2003).

The BioSphere portal will use views in both senses: *XML views* will be used to extract user and version-specific fragments of the single OWL/XML document that is the shared ontology source. To this end, we will use the methods of Buneman (2001, 2002) and Fan (2004, 2007) - who has solved the problem of rewriting XPath queries. *Ontology views* will be derived once the XML view has been specified, and will be represented in data structures that the end user can manipulate through graphical components. These views will include the *logic view*, where the visualised structures are oriented around the classes and their definitions; and the *annotation view*, where the visualised structures are based on the annotations made to terms.

These views will aid human comprehension of the extracted ontology fragments, and so help users understand the impact of their concept definitions, and those of others. The system will allow each user to see (a) which nodes/links (concepts and definitions) in their view of the ontology are and are not shared with others, (b) the common core of the ontology (agreed classes and links) and (c) the ontology from other users' perspectives. The system will provide appropriate visualisations of the views as annotated graphs and trees. It will also be possible to specify combinations of views, e.g. definitions on which *User-A* and *User-B* disagree. These features will help users to understand the conceptualisations of others, and so assist the process of reaching consensus. A simple notification system will alert users to changes made by others.

The BioSphere Portal: Use of the portal framework allows user groups and security options to be configured easily, thus allowing editing as well as browsing functions to be provided to a dispersed user group. Portlets can use new methods and formats without burdening the user with unnecessary details e.g. of syntax. This allows more significant problems to be addressed by users, which, in the case of ontologies, include the implications at the semantic level of OWL-based representation and reasoning.

The BioSphere portlets are being implemented using GridSphere 3.1, OGSA-DAI 3 and eXist 1.2, all of which have been successfully integrated within tomcat 5.5.26. GUI components are being implemented using MVC and standard Java/JavaScript models, while XML rewriting components will be steps in OGSA-DAI workflows. Thus existing frameworks are being reused, and novel reusable components created.

Acknowledgements: BBSRC grant BB/F015976/1

Ontology Views for Collaborative Ontology Creation – Aitken and Bard

References

- Aitken, S., Chen, Y. and Bard J. OBO Explorer: An Editor for Open Biomedical Ontologies in OWL (2008) *Bioinformatics*, 24(3):443-444.
- Aitken, S., Chen, Y., Webber, B., Fan, W. and Bard, J. (2007) Managing the transition from OBO to OWL: The COBrA-CT bio-ontology tools. *Proc. UK e-Science All-Hands Meeting 2007*.
- Bard, J.B.L. (2005) Anatomics: the intersection of anatomy and bioinformatics. *J. Anatomy*. 206 (1) :1-16.
- Bard, J.B.L., Rhee, S.Y. and Ashburner, M. (2005) An ontology for cell types. *Genome Biol.* 6(2):R21
- Bard J. (2007) Systems developmental biology: the use of ontologies in annotating models and in identifying gene function within and across species. *Mamm Genome*. Epub PMID: 17566825.
- Bhatt, M. Taniar, D. and Dillon, T. A Distributed Approach to Sub-Ontology Extraction *Proc. AINA*:636-641.
- Buneman, P., Davidson, S., Fan, W., Hara, C. and Tan, W. (2001) Keys for XML. *Proc. WWW 10*:201-210.
- Buneman, P., Khanna, S., Tajima, K. and Tan, W. (2002) Archiving scientific data. *Proc. SIGMOD*:1-12.
- Fan, W., Chan, C.Y. and Garofalakis, M. (2004) Secure XML Querying with Security Views *Proc. SIGMOD* :587-598.
- Fan, W., Geerts, F., Jia, X. and Kementsietsidi, A. (2007) Rewriting Regular XPath Queries on XML Views *Proc. ICDE* :666-675.
- Haendel, M.A *et al.* (2007) CARO – The Common Anatomy Reference Ontology. In *Anatomy Ontologies for Bioinformatics* (Eds.) Burger, A., Davidson, D. and Baldock, D.
- Moreira, D.A. and Musen, M.A. (2007) OBO to OWL: A Protégé OWL Tab to Read/Save OBO Ontologies *Bioinformatics* doi:10.1093/bioinformatics/btm258
- Noy, N., Musen, M. (2004) Specifying Ontology Views by Traversal *Proc. ISWC*:713-725.
- Seidenberg, J. and Rector, (2006) A. Web Ontology Segmentation: Analysis Classification and Use *Proc. WWW*:13-22.
- Voltz, R. *et al.* (2003) Views for Light-weight Web Ontologies *Proc. SAC*:1168 – 1173.