Ontology Views for Collaborative Ontology Development

Stuart Aitken and Jonathan Bard

Informatics, University of Edinburgh

This project will implement tools to support the collaborative development of biological ontologies by expert biologists. Biological ontologies are typically developed by loosely-organised groups of experts who participate in the ontology development process remotely[1]. Current technological support for bio-ontology development relies on stand-alone ontology editors such as OBOedit, COBrA and Protégé[2] for creating new ontology versions, and private email, email lists and/or Wikis for the distribution of ontology files and discussions about their contents. The resulting exchanges are often far from enlightening, and ineffective in their use of the available storage, versioning and visualisation techniques.

The wide success of the Gene Ontology (GO) has shown the important role a shared conceptual view of gene products can have on data interpretation. Subjects such as anatomy, phenotype and clinical studies are either active areas for ontology development where consensus has yet to be reached, or have ontologies whose organisation needs to be kept under review. At the very least, any ontology that is in use will be curated and subject to minor extensions and alterations. In all cases, effective collaborative ontology development is essential to progress and appropriate tools are needed.

The creation and exploitation of links across ontologies and anatomies is as yet relatively unexplored, an exception being the XSPAN project (*http://www.xspan.org*) that made links between the developmental anatomies of model species and the cell-type ontology. Creating such links requires access to multiple ontologies, and this requires computational access to a set of ontology files (i.e. ontologies of different domains) and support for searching over their contents. The current generation of ontology tools are ineffective as they allow only a single file to be edited, and then be saved on the users machine, or possibly checked back into a database for archival. Essential sharing and exchange features are missing.

To address the problems of collaborative development and resource integration, we propose making a web-accessible system (a Grid portal) that maintains a single centrally-held XML document containing the ontology for which all users in a specified group will have access and editing rights. The system will provide each user with their own view of the ontology which will be created on-the-fly from the single source 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, all of which we assume will be in the Resource Description Framework (RDF) or its extension, the Web Ontology Language (OWL). The search function will be capable of operating over all ontologies in the database thus providing the basis of resource integration. We also consider the archival of other resources that link ontologies such as RDF models of the development of model organisms as proposed by Bard (2007), and will consider SBML models. The project will build on an existing Grid-based implementation of an ontology management server and graphical interface components developed on BBSRC grant BB/D006473/1, and a key aim will be to make the GUI easy for non-computer scientists to use.

Programme and Methodology

Thus far the term *view* has been used informally, in the sense of a perspective or a viewpoint that the end user will take of the ontology. More technically, *database* and *XML views* can be distinguished from *ontology views*. These views are not visualisations, although visualisations are important and inter-related (see below). 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

---

[1] The Gene Ontology is a notable exception as its curation is carried out by specialised staff, following defined procedures, using in-house tools.

[2] OBOEdit: *http://oboedit.org* ; COBrA: *http://www.xspan.org* ; Protégé: *http://protege.stanford.edu*

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 term 'ontology segmentation' has been coined to describe the extraction of a subset of an ontology (Seidenberg and Rector, 2006). An *ontology view* must account for the knowledge representation scheme used to define the ontology: it is necessary to know which elements are classes, instances, relations or quantifiers as these must be treated appropriately. The terms and relations should also be connected – in contrast with *database views* which need not result in a set of connected elements. The motivation for computing *ontology views* is similar to that for *database views*: extracting a smaller subgraph of terms from larger structure should improve the efficiency of automated querying and reasoning, as well as aid human comprehension and discussion.

The proposed system 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 efficiently as automata, thus avoiding a potentially exponential increase in the size of the rewritten query. *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 that we will implement. 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. Annotations are important to Open Biological Ontologies as they capture all the definitional text, synonym, subset and cross-reference information that cannot be expressed in the graphical structure of the ontology. The OBO 'subset' mechanisms and the Gene Ontology's 'GO-slims' are examples of *annotation views* in current use. We propose to use query-based and link traversal-based (Noy, 2004) methods to generate *ontology views* that will aid human comprehension of the extracted ontology fragments, and so help users understand the impact of their concept definitions, and those of others.

Updates and deletions to ontologies will be handled at the XML document level, where changes will be tagged on a per-user and timestamp basis. OWL 1.1 will be used in order to exploit the associated XML schema on which XML methods depend. XML views corresponding to user and version will implement an efficient multi-user version management system.

Turning again to the user's perspective, 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-1* and *user-2* 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 system will house multiple ontologies, RDF documents and other XML documents such as SMBL models and will allow searching across all stored ontologies/models/documents in order to extract references (e.g. URIs) that can be used to further extend and cross-link the various resources. The system will support the generation of computed *is-a* links by DL classifiers, and the visualisation of the extended ontology.

The system will be implemented using GridSphere[3], an open-source framework for developing web-accessible Grid portals. The key advantage of this is that users have access to tools, but do not need to install, configure or maintain them as the portal runs in a web browser and the software is installed centrally on the server by the developers. From our experience of developing for plug-in architectures, we conclude that, for biologists to be able to use such tools, it is crucial that they be installed, configured and maintained for them. This portal framework is also compatible with the OGSA-DAI-based system we wish to reuse, and hence the advantages of the existing Grid-based solution are retained.

---

[3] http://www.gridsphere.org/

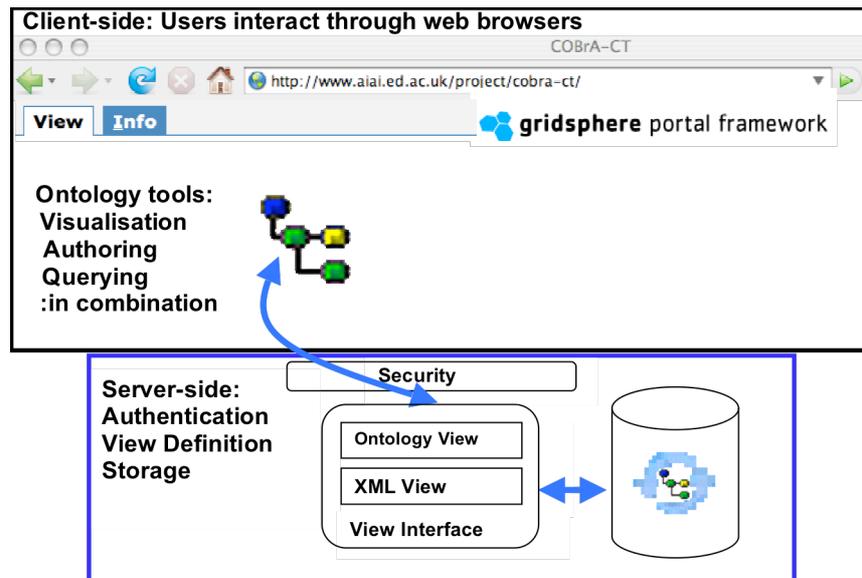# Ontology Views for Collaborative Ontology Development

Fig 1: Schematic of the portal architecture: Ontologies are viewed and edited by graphical components in the GridSphere portal which runs in the user's web browser. The ontologies are held in a central XML database, operated within the Grid framework.

Case Studies

The ontology portal will be evaluated by case studies:

1. An external study evaluating the tools in the on-going development of the Ontology for Biomedical Investigations (OBI) will carried out in collaboration with the EBI and partners (see letter of support).

2. A study on the introduction of spatial and topological concepts into the developmental mouse anatomy providing support for discussion and review of alternative conceptualisations will be carried out with the Human Genetics Unit of the MRC and collaborating institutes (see letter of support).

3. The on-going work of Bard (2007) on the creation of developmental process linkage graphs will provide an in-house test for the search and resource linkage functions.

The design of a Common Anatomy Reference Ontology (CARO) is currently being debated by an international group of anatomists for the major model systems (Haendel *et al.,* 2007). We are discussing with Dr M Haendel (Oregon USA) whether our system might provide a vehicle for this work.

Related Work

The Biomedical Informatics Research Network (BIRN) has tools and resources that we plan to incorporate. In particular, the lexicon (BIRNLex) can be immediately added to our resource database as it is made available in OWL. We will explore the possibility of obtaining computational access to the BIRN mouse tissue lineage data[4] (itself derived from earlier work by Bard and Aitken) which may be simplified by the use of the GridSphere architecture. We have strong links to the National Center for Biomedical Ontology (*http://www.bioontology.org*) which provides the core portal for bio-ontologies. Aitken actively contributed to the NCBO work of implementing the OWL representation for OBO, and designed our tools to be compatible with the resulting standard.

The NCBO tools (e.g. OBOEdit and Amigo) do not support collaboration – user activity is limited to exploring the resources created previously – but they do support search over ontologies. However, the search results must be copied into some other tool, and so there is no integration between searching and authoring.

---

[4] http://www.nbirn.net/tools/mouse_tissue_lineage_hierarchy/index.shtm

The Protégé 4 editor (*http://protege.stanford.edu*) has recently been released and, once the service and client APIs are available, we would aim for compatibility with them, and endeavour to port GUI components (but this is outside the scope of the current project).

Collaboration Opportunities

As identified above, BIRN and NCBO are related initiatives and the sharing of methods, tools and resources is planned (or indeed on-going). Collaboration with the EBI and MRC-HGU is part of the work plan.

References

Aitken, J.S., Korf, R., Webber, B.L. and Bard, J.B.L. (2004) COBrA: A Bio-Ontology Editor. *Bioinformatics,* 21(6):825-826.

Aitken, J.S. (2005) Formalising concepts of species, sex and developmental stage in anatomical ontologies. *Bioinformatics* 21(11):2773-2779.

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* (*in press*)

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.