

BioPortal: A Web Portal to Biomedical Ontologies

Daniel L. Rubin,¹ Dilvan A. Moreira,¹ Pradip P. Kanjamala,² and Mark A. Musen¹

¹Stanford Center for Biomedical Informatics Research, Stanford University
MSOB X-215, 251 Campus Drive, Stanford, CA 94305 USA

²Division of Biomedical Informatics, Department of Health Science Research, Mayo Clinic
rubin@med.stanford.edu, dilvan@gmail.com, Kanjamala.Pradip@mayo.edu, musen@stanford.edu

Abstract

Ontologies are becoming critical to e-Science, but there are challenges for the community to find the ontologies relevant to their work, to declare mappings among related ontologies, and to provide feedback or critiques to ontology developers. We have created BioPortal, a Web portal to a virtual library of ontologies on the Semantic Web and a tool set enabling the community to access, critique, and improve ontologies. The BioPortal library contains over 50 ontologies from the biological and medical domains. In addition to a Web interface enabling researchers in cyberspace to locate these knowledge resources, BioPortal provides a suite of Web services, including ontology categorization, term search, graphical ontology visualization, and ontology version histories. As development of the knowledge resources in e-Science is increasingly collaborative, we are also creating novel tools in BioPortal to enable the community to create mappings between classes in related ontologies and to critique ontology content, providing feedback to ontology developers. Preliminary user experience with BioPortal has been extremely positive. BioPortal appears promising for unifying and disseminating ontology content on the Semantic Web, and it is providing tools needed by the research community to exploit these rich resources.

Motivation and Preliminary Results

Our objectives are to address the following key challenges for the e-Science community to access ontologies in cyberspace: (1) **Knowledge access and awareness:** The number of ontologies in e-Science has exploded. It is difficult for *consumers of ontologies* to find them and to compare the multitude of related ontologies. It is likewise challenging for *developers of ontologies* to stay abreast of all current ontology development efforts and to avoid duplication of effort. (2) **Search:** Ontologies are distributed throughout cyberspace, and the community needs the capability to search across all ontologies for those containing particular terms or types of terms. (3) **Diversity of Knowledge Representation (KR) Languages:** There are different KR languages (OBO Format, Protégé Frames, RDF, OWL), each requiring a different tool for perusing and searching ontology content. The community needs to access ontologies in a consistent way regardless of KR language. (4) **Overlapping content:** Many ontologies contain terms that

are related to terms in other ontologies (such as through synonymy). Methods are needed to make the linkages across ontologies explicit so that the community can unify biomedical knowledge. (5) **Community feedback:** Ontology development in e-Science is uncoordinated and fragmented, with only ad-hoc mechanisms available to users for critiquing ontologies or making suggestions for improvements.

To tackle these challenges, we are creating BioPortal (<http://bioportal.bioontology.org/>), a Web portal to a virtual library of e-Science ontologies. BioPortal provides a suite of ontology services enabling the community to access and exploit ontologies in a diversity of applications. BioPortal differs from related efforts such as the Open Biomedical Ontologies (OBO) Foundry, (<http://obo-foundry.org/>) and the Ontology Lookup Service (Cote et al. 2006). Beyond the ontology library functionality, BioPortal is providing the tools needed to reduce the fragmentation in ontology development and to enable community participation.

Ontology Access and Awareness

BioPortal collects and unifies the important medical and biological ontologies, and it provides a common access mechanism to all ontologies regardless of the KR language: all biological ontologies from OBO Foundry as well as key medical terminologies and ontologies. BioPortal also maintains prior versions of ontologies and ontology metadata such as the biomedical domain for an ontology (anatomy, physiology, etc), version number, provenance information, and indicators of ontology quality (whether the ontology adheres to design principles that promote interoperability). BioPortal keeps its ontology library current by pulling new versions of ontologies from the Web or from ontology authors who submit new releases.

The contents of the BioPortal ontology library can be viewed in two different ways: a catalog view of all ontologies in an alphabetical list, and a category view, which uses the category information in the ontology metadata to create a taxonomy of the ontologies (Figure 1). The ontologies can be browsed as an expandable tree or as a graph, showing the immediate local neighborhood (or all terms to the root) for selected ontology terms (Figure 2).

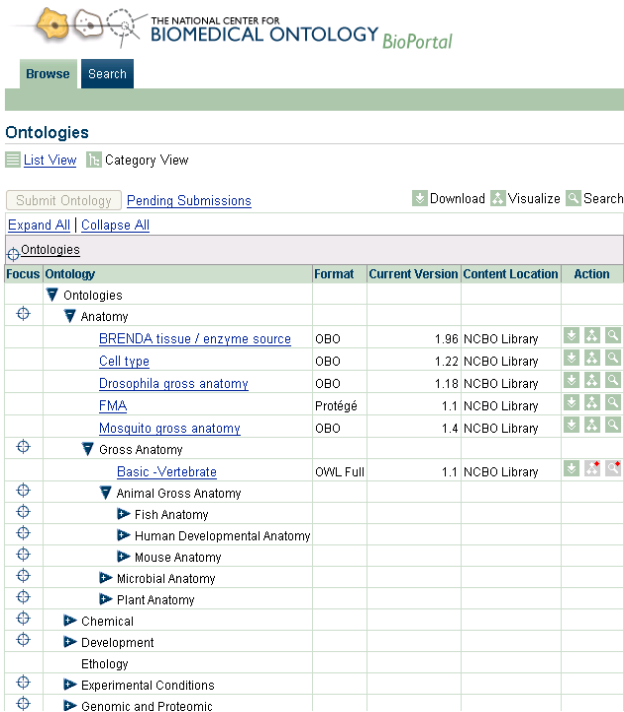


Figure 1. BioPortal ontology library displayed in the category view, showing a taxonomy of ontologies organized according to the type of the ontology.

BioPortal is built using a service-oriented architecture, encapsulating the breadth of its functionality as a set of services. In addition to a Web application layer for users, there is a Web services layer enabling agents in cyberspace to access the rich BioPortal ontology content.

Ontology Search

Users can search for terms within an ontology or across all BioPortal ontologies. Searches can be restricted to class names, properties or other attributes. Searches can also be based on exact matches or soundex. In addition, users can search ontology metadata to find particular types of ontologies. BioPortal contains a master index of all ontology content and metadata to streamline these searches.

Unification of KR Languages

Biomedical ontologies are created in a variety of KR languages, such as OBO Format, OWL, and Protege frames. BioPortal translates ontologies in the different KR languages into a common representation so that it can unify the diverse ontology content in cyberspace. This functionality is critical to search across all ontologies. To unify the diverse content, the BioPortal ontology library adopts a common representation of the terminological aspects of ontologies—the Common Terminological Services (CTS) model. Common terminological aspects include the repre-

sentation of terms (synonyms, terms in different language, lexical variations) and the relations between them.

The unification of ontologies in BioPortal also enables it to make all ontology content accessible to the Semantic Web—we are currently publishing all terms in BioPortal to the Web using Uniform Resource Identifiers (URI) so that diverse applications throughout cyberspace can refer to the breadth of biomedical entities in these ontologies.

Overlapping Content

There are many ontologies with similar entities; for example, there are several ontologies and terminologies for human anatomy. In e-Science, it is critical to be able to make statements about how entities in one ontology relate to those in another ontology. A biologist might want to say that “a white blood cell *hasPart* nucleus.” This statement would require a mapping from the “white blood cell” (Cell Type ontology) class through a *hasPart* relation to the “nucleus” (Gene Ontology) class. BioPortal will store mappings between classes in related ontologies so that users can make such statements about semantically-related classes. The mappings will also include metadata (such as the creator or type of automated mapping method) so that users of the mappings can filter them according to established trust relationships. The mappings will enable applications to bring together diverse data on the Semantic Web that was annotated with different, yet semantically-related ontology terms.

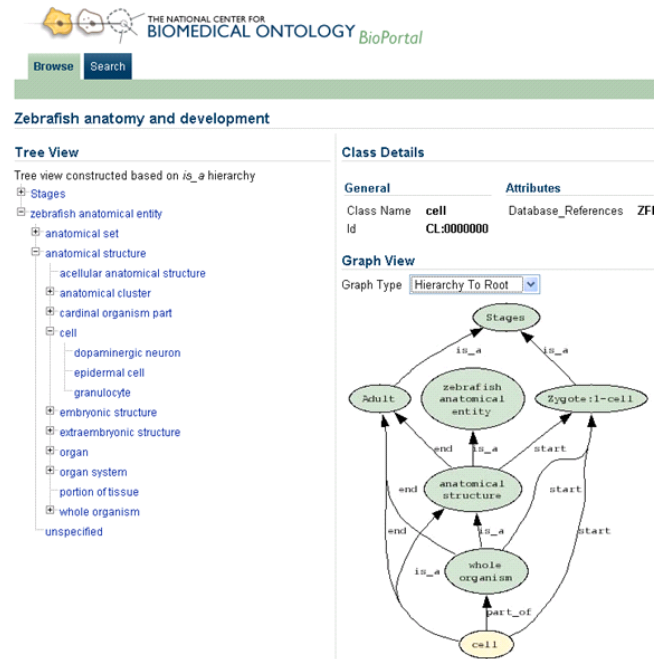


Figure 2. Ontology visualization in BioPortal. Ontologies are shown both as an expandable tree (left) as well as a local neighborhood graph (right; the selected class is highlighted in yellow).

Community Feedback

A particularly important role we see BioPortal fulfilling in the ontology community is helping it critique ontologies and record the changes needed to improve their quality. At present, community participation in ontology development is limited to email or face-to-face discussions which are disconnected from the ontologies themselves, making it challenging to track all the outstanding issues and to assess the status of each issue.

Developing Ontologies Cooperatively

We are implementing a BioPortal function that enables users to link structured comments to ontologies or to components of ontologies (such as individual classes), informing ontology developers and the community about potential problems and suggested resolutions. A tool called BioNotes has already been implemented (Figure 3). Users of ontologies commonly need new classes to be added to ontologies, or for existing terms to be renamed or moved in the ontology. BioNotes enables the community to provide this feedback in a structured manner directly linked to the classes in question.

Users can attach notes to ontology components as a com-

municative act or CA. CAs are based on speech act theory (Searly 1975) which define the functions of simply specified actions. BioNotes modifies the Change Ontology (Noy 2006) to add classes that represent the CAs defined for the FIPA-ACL language (FIPA 2002), such as Propose, Inform and Refuse. This modified ontology is used to qualify each note. For instance, a note can be added to a class and labeled `Propose->Add->Subclass`, to denote a proposal for a new class, another user can then add a note, labeled `Refuse`, to the first note to express his disagreement. Using this ontology, a lot of computer readable information can be easily attached to notes. Computers can use this information to help ontology developers to quickly localize feedback to the parts of the ontology that need to be addressed and to what users want changes.

In addition to the semantic information, users can add text and multimedia material (using HTML) to the body of a note, using an easy-to-use interactive web page editor, to argue about or explain their note. Using this editor, users can add figures, links and even movies to their notes. In fields highly dependent on images, like biology, radiology or anatomy, notes can become a multimedia documentation for an ontology.

The screenshot displays the BioPortal interface for the RadLex ontology. On the left, a 'Tree View' shows a hierarchical structure of classes, with 'foreign body' selected. The 'Class/Type Details' pane shows the class 'foreign body' (ID: di:RID5425) with attributes like 'di:Preferred_Name' and 'di:Comment'. The 'Marginal Notes' pane shows a note titled 'Plain radiograph of abdomen showing radio-opaque foreign body (a coin in the region of the gastric antrum)' with an associated image. The 'Graph View' pane shows a 'Local Neighborhood' graph with 'foreign body' at the center, connected to various subclasses like 'surgical implement', 'decoration', 'medical device', 'embolic material', 'other foreign body', 'bezoar', and 'missile'.

Figure 3. BioNotes: Marginal notes on ontologies. This screenshot shows feedback from the community on the RadLex ontology in BioPortal. The community can enter comments on ontology classes (left side of figure, icons to the right of the class name), and each marginal note can be viewed individually by selecting a class (details of note shown in upper right pane of figure). Here, a user made a marginal note to show an image illustrating the RadLex term, “foreign body.”

BioNotes also includes metadata on user feedback, such as the author of the comment, allowing the community to establish trust relationships and to prioritize suggestions and critiques. BioNotes is implemented as a client program running in the web browser (using java code compiled by GWT into javascript) and a set of Web Services provided by NCBO. This Web 2.0 application style (based on AJAX) offers a very interactive desktop feel to BioNotes, avoiding, as much as possible, the delays associated with web browser interaction (web browser waiting for a new page on each interaction).

BioNotes is still a prototype and feedback from our user community will shape its final UI and ontology. We believe that tools such as BioNotes will draw ontology users closer to ontology developers, improve communication, and ultimately enhance the quality of ontologies.

Evaluation

We have performed an initial evaluation of BioPortal, focusing on usability testing by a group of ontology users as well as through feedback from our user community. Two biomedical database curators independently used BioPortal to browse ontologies, search for terms, and to visualize the results, while assessing system performance, functionality, and the ease of use of the user interface. They agreed that BioPortal provides a user-friendly and intuitive experience for accessing, browsing, and searching diverse ontologies. The ability to visualize the local neighborhood of a search term was very helpful because a term one link removed from the search term is sometimes a better match for their needs than the original search term itself. Our user community has given us similar positive feedback. In fact, one community is adopting BioPortal as a platform for serving local medical terminologies in cancer centers.

ACKNOWLEDGEMENTS

This work was supported by the National Center for Biomedical Ontology, under roadmap-initiative grant U54 HG004028 from the National Institutes of Health. We thank all the talented members of NCBO for their contributions. We acknowledge and thank Daniel Armbrust, Harold Solbrig, Thomas Johnson, and Deepak Sharma for their contributions to the LexGrid project.

REFERENCES

Cote, RG, et al. 2006, *The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries*, BMC Bioinformatics, vol. 7, p. 97.

Noy, NF, et al. 2006, *A Framework for Ontology Evolution in Collaborative Environments*. 5th International Semantic Web Conference, Athens, GA.

Searle, JR. (1975), "A Taxonomy of Illocutionary Acts", in: Günderson, K. (ed.), *Language, Mind, and Knowledge*, Minneapolis, vol. 7

FIPA (2002), *Communicative Act Library Specification*, Foundation for Intelligent Physical Agents Online: <http://www.fipa.org/specs/fipa00037/SC00037J.html>