Collaborative Protégé: Enabling Community-based Authoring of Ontologies

[Demonstration]

Tania Tudorache   Natalya F. Noy   Mark A. Musen

Stanford Center for Biomedical Informatics Research, Stanford University
Stanford, California 94305
{tudorache, noy, musen}@stanford.edu

ABSTRACT

Ontologies are becoming so large in their coverage that no single person or a small group of people can develop them effectively and ontology development becomes a community-based enterprise. We present Collaborative Protégé—an extension of the Protégé ontology editor that we have designed specifically to support the collaboration process for a community of users. During the ontology-development process, Collaborative Protégé allows users to hold discussions about the ontology components and changes using typed annotations; it tracks the change history of the ontology entities; it provides a chat and search functionality. Users edit simultaneously an ontology stored in a common repository. All changes made by a user are seen immediately by other users. Collaborative Protégé is open source and distributed with the full installation of Protégé.

1. BACKGROUND

Recent developments are dramatically changing the way that scientists are building ontologies. First, as ontologies are becoming commonplace within many scientific domains, such as biomedicine, they are being developed collaboratively by increasingly large groups of scientists. Second, ontologies are becoming so large in their coverage (e.g., NCI Thesaurus with 80K concepts) that no one user or small group of people can develop them effectively. Hence, organizations such as the NCI Center for Bioinformatics "outsource" some of their ontology development to the scientific community at large. Third, in the last one or two years, many users have become quite familiar and comfortable with the concept of user-contributed content, both in their personal and professional lives (cf. Web 2.0). Thus, domain experts need tools that would support collaborative ontology development and would include collaboration as an integral part of the ontology development itself.

2. SYSTEM DESCRIPTION

Our laboratory has developed Protégé—a widely used open-source ontology and knowledge base editor [2]. At the time of this writing, Protégé has more than 100,000 registered users. Users can build ontologies in Protégé using different representation formalism ranging from *Frames*, to *RDF(S)* and *OWL*, and store them in file or database back-ends.

Protégé can be run as a standalone application, or in a client–server setting. In the client–server mode, the ontologies are stored on a central Protégé server. Users access the ontologies on the server to browse and edit them through desktop or web Protégé clients. All the changes that are made by a user in an editing session are immediately reflected on the ontology stored in the repository, meaning that also other users who are working on the same ontology will immediately see the changes performed by others. This mode of collaborative editing is different than the source code control model (e.g. CSV and SVN), in which users have to check out a copy of the ontology, make their changes, then resolve possible conflicts, and check in a new version of the ontology.

We have developed Collaborative Protégé[1] as an extension to the client–server Protégé. Collaborative Protégé [3] enables users who develop an ontology collaboratively to hold discussions, chat, annotate ontology components and changes—all as an integral part of the ontology-development process. The key feature of Collaborative Protégé is the ability to create annotations. In this context, *annotations* are typed comments (e.g. example, proposal, question, etc.) attached to ontology components, or to the descriptions of ontology changes, or to other annotations. We define the structure of the annotations in the *Changes and Annotations ontology* (CHAO) [1]. CHAO contains classes that define the annotation types (e.g. *Comment*, *Proposal*) and classes for describing different types of changes that a user can do in an editing session (e.g. *Class_Created*). Instances of the annotation types will represent actual annotations that users are making on ontology components or changes. To each domain ontology stored in the Protégé server repository, we associate a CHAO knowledge base that contains the instances of annotations and changes corresponding the to domain ontology. Several domain ontologies can share the same CHAO knowledge base.

3. USER INTERFACE

The user interface of Collaborative Protégé (Figure 1) is implemented as a graphical extension of Protégé. Panel A in Figure 1 shows the class tree, panel B shows the selected class information (in this case Gene_Product)—just like in the original Protégé user interface, while panel C displays the *collaborative tabs*. Each of the collaborative tabs supports one of the several collaboration features. For example, in the *Annotations* tab, the user can add comments to ontology components; in the *Changes* tab, the user may see the change history of the selected class and comment on a change; in the *Search* tab, the user can search all annotations on different criteria; in the *Chat* tab, the user may discuss with other online users, and so on.

An alternative user interface for Collaborative Protégé is Web-Protégé[2] – a web-based interface to the Protégé back-end that supports light-weight editing as well as some of the collaboration features.

---

[1]http://protege.stanford.edu/doc/collab-protege/
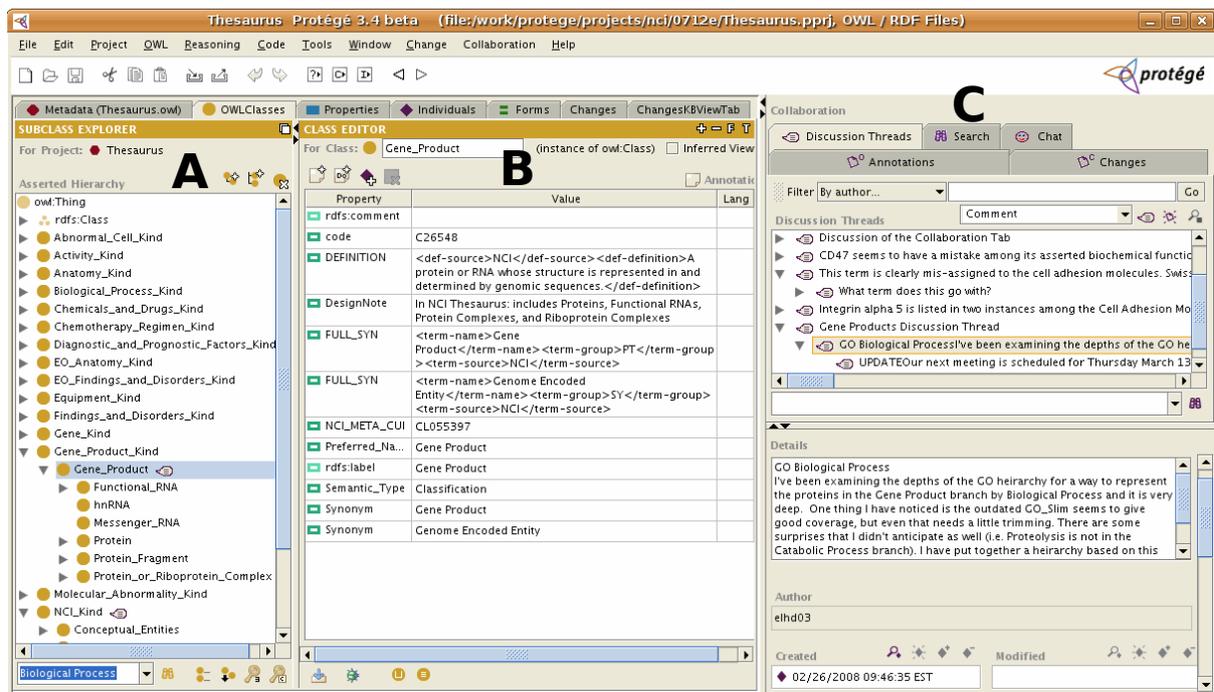[2]http://bmir-protege-dev1.stanford.edu/webprotege/

Figure 1: The Collaborative Protégé extension used with the NCI Thesaurus. Panel A shows the class tree; panel B displays the description of the selected class `Gene_Product`**; and panel C shows the discussion among users about this class.**

4. KEY FEATURES

In this demonstration, we will highlight some key features of Collaborative Protégé.

Annotation of ontology components and changes. Users are able to create notes or annotations of different types associated to ontology components (e.g. comment, example, proposal, etc.) or change events (e.g. the creation of a class). Collaborative Protégé also provides the ability to create discussion threads attached to the ontology as a whole.

Change tracking. Collaborative Protégé records all actions made by a user and stores them in a structured log as instances of the `Change` class of the CHAO ontology. The change log contains information about who and when made a change in the ontology. Because the changes are also stored as instances of the CHAO ontology, they can also be annotated to record design rationale, or other information.

Filtering and Search. For projects with large sets of annotations, search and filtering are crucial features. Users of Collaborative Protégé can filter and search in each of the collaboration panels using different criteria, such as author, date, annotation text, and annotation type.

Chat. Users connected at the same time to a Protégé server can discuss and exchange live messages. One feature that sets Collaborative Protégé chat functionality apart from other chat clients is the support for sending links to entities in the ontology (e.g. a class). The user who receives a chat message containing an entity link, can simply click on that link, and she will see the definition for that entity.

Extensibility and API access. We designed Collaborative Protégé in such a way that it is very easy to extend the annotation types available in the user interface. It only requires to add a subclass of `Annotation` in the CHAO ontology and the tool will be able to handle the new type of annotation. Collaborative Protégé is built using a plug-in architecture, meaning that other groups can very easily add their own custom collaboration panel. The extensibility is made possible by exposing all the collaboration information (e.g. annotations and changes) through API calls, so they can be easily used and integrated in other applications.

5. CONCLUSIONS

We have presented Collaborative Protégé – an extension of the Protégé editor that supports the collaborative ontology development process. Some of the main features of the tool include the possibility to annotate ontology components and changes, change tracking, search and filtering of annotations, a chat functionality and simultaneous editing. Collaborative Protégé is open source and distributed with the Protégé installation.

Acknowledgments

6. REFERENCES

[1] N. F. Noy, A. Chugh, W. Liu, and M. A. Musen. A framework for ontology evolution in collaborative environments. In *5th ISWC*, volume LNCS 4273, Athens, GA, 2006. Springer.

[2] Protégé. `http://protege.stanford.edu/`.

[3] T. Tudorache, N. F. Noy, and M. A. Musen. Supporting collaborative ontology development in Protégé. In proc. of *the 7th ISWC 2008*, Karlsruhe, Germany, 2008.