

Demo Proposal:

NCIEditTab: Protege 3.4 Plugin for Editing and Maintaining NCI Thesaurus

Gilberto Fragoso, Sherri de Coronado, NCI Center for Bioinformatics

The NCI Thesaurus (NCIt) is a biomedical terminology with about 80,000 concepts. It supplies a majority of the controlled terminology for the Cancer Bioinformatics Grid (caBIG) as well as special terminology for FDA, CDISC and other organizations and other parts of the cancer research community. It is mission critical. NCIt is published monthly, and edited by a group of about 15 domain expert editors who need to work simultaneously. Editing of the terminology is conducted with the NCIEditTab plugin in the Protégé environment. The terminology is published to the LexBIG terminology server (see “LexBIG/EVS” in [1]), accessed directly through the LexBIG API or through web browsers [2], or downloaded by end users in several formats [3].

The NCIEditTab is a package of plugins for Protege 3.4 for supporting terminology editing and maintenance at NCI, publicly available for download at [4]. It provides for workflow, customized editing, unique code generation, history recording, and batch editing and loading. The NCIEditTab environment also includes tabs for Lucene-based queries, reporting, and classification explanations.

To support the NCI’s workflow the NCIEditTab uses Protégé in client server mode with a database backend for concurrent editing. The NCIEditTab was designed with a user interface for domain expert editors simpler than the native OWL interface in Protégé (Figure 1). In addition, it provides a large number of edit checks and enforces a number of business rules to improve the quality of the edited vocabulary. It also supports the change of ontologies over time, controlling those changes with notions of splitting and merging, retirement, change ontologies and history. Meta level data such as who and when are captured for auditing purposes, in both a change ontology and a history plugin.

PROMPT is used for reviewing changes from baseline to baseline and accepting or rejecting those changes, while additional plugins to create concept history from user history have been added. The workflow tab enables a workflow manager to create worklists, to enable editors to suggest concepts that need to be edited, and to enable retrieval of change proposals from the BiomedGT wiki [5], based on the Semantic Media Wiki. Finally classification services have been added to assist debugging the classification results. This explanation plugin interacts with Pellet, and by assisting debugging it facilitates the creation of consistent terminologies. These plugins are packaged up for easy installation. Instructions are available with the download at [4]. The NCIEditTab will be demonstrated at the Protégé Conference.

[1] https://cabig-kc.nci.nih.gov/Vocab/KC/index.php/Main_Page

[2] <http://bioportal.nci.nih.gov/> based on the NCBO Bioportal

[3] Current formats include OWL, Ontylog, and flat file, at ftp://ftp1.nci.nih.gov/pub/cacore/EVS/NCI_Thesaurus/

[4] https://gforge.nci.nih.gov/frs/?group_id=174

[5] http://biomedgt.nci.nih.gov/wiki/index.php/Main_Page

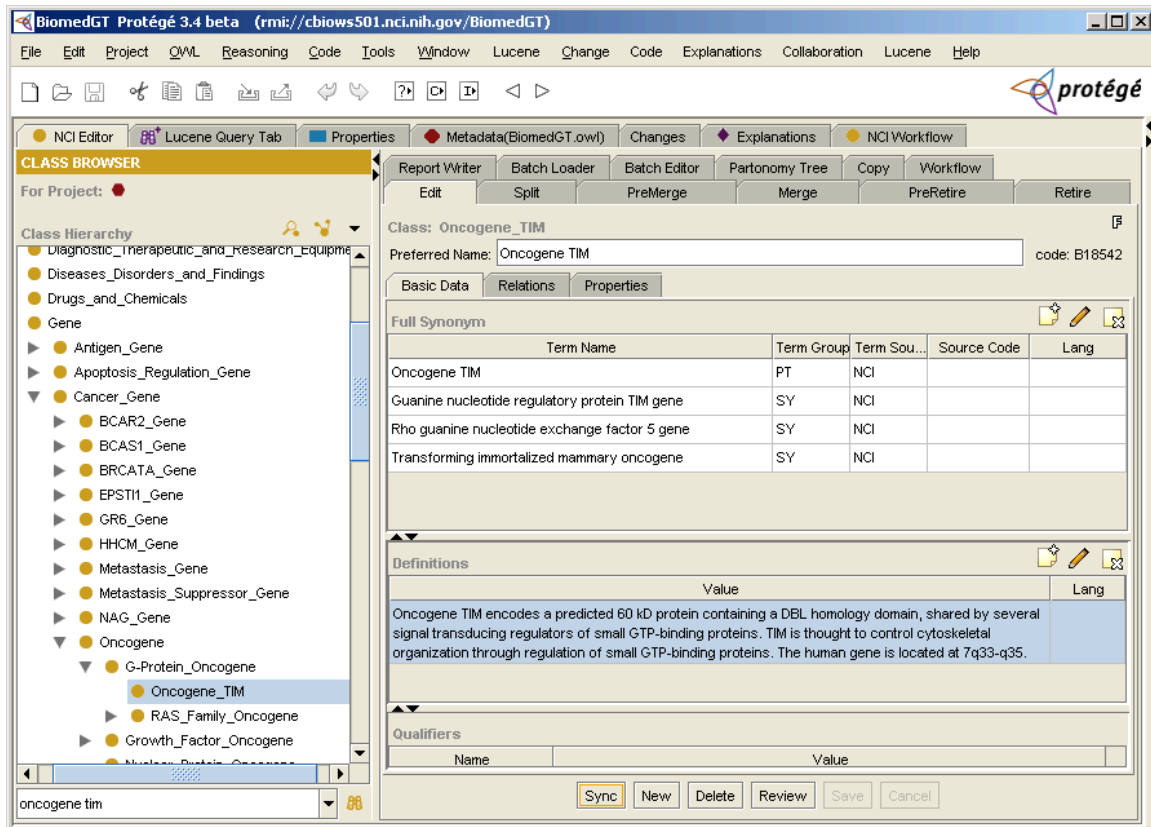


Figure 1. The basic editing subtab of the NCIEditTab, showing the Basic Data sub-tab which supports the editing of terms and definitions.