

NCI PROTÉGÉ

1.3

Workflow Manager's Guide



NATIONAL[®]
CANCER
INSTITUTE

Center for Bioinformatics

CONTENTS

About This Guide	1
Purpose	1
Audience	1
Topics Covered	2
Text Conventions Used	2
Credits and Resources	3
Application Support Contacts	4
Chapter 1	
Workflow Management	5
Overview of BiomedGT	5
About BiomedGT	5
<i>Purpose</i>	5
<i>Audience</i>	6
About the BiomedGT Wiki	6
Overview of Internal NCI Workflow	8
Primary Goals	8
Workflow Roles	8
Workflow Stages	9
About the NCI Workflow Tab	10
Workflow Procedures	12
Importing Packages from the BiomedGT Wiki	12
Assigning (or Reassigning) a Workflow Item	12
Chapter 2	
Baseline Comparison	13
Overview of Baseline Comparison	13
About the Comparison Process	13
About the Prompt Plug-in	15
Pre-Comparison Procedures	15
Notifying Protégé Users of Database Shutdown	15

Starting the Remote Protégé Client	15
Comparison Procedures	17
Enabling the Prompt Tab	17
Starting the Comparison	18
Tools for Interpreting the Comparison Results	19
<i>About the Table View Subtab</i>	20
.....	22
<i>About the Tree View Subtab</i>	22
<i>About the Users Subtab</i>	24
Accepting and Rejecting Changes	26
<i>Accepting a Single Change</i>	26
<i>Accepting All Changes</i>	26
<i>Rejecting a Single Change</i>	26
<i>Rejecting All Changes</i>	27
Saving the Concept History	27
Disabling the Prompt Tab	27
Preparatory Procedures for the Next Comparison	28
Managing Baseline Versions	28
<i>About the Naming Convention for a Baseline Version</i>	28
<i>Changing the Baseline Version Number for the Next Baseline Build</i>	29
Managing the Annotation Ontology	30
<i>About the Annotation Ontology</i>	30
<i>Cleaning Up the Annotation Ontology</i>	30
Exporting a Project File	31
Saving the Inferred Terminology File (Publication Baseline Only)	32
Deleting Extraneous Information	33
<i>About the Accepts Files and Journal File</i>	33
<i>Deleting the Accepts Files and Journal File</i>	33
Post-Comparison Procedures	34

Chapter 3

Classification	37
Overview	37
About the Classification Process	37
About the Pellet Reasoner	38
Running the Pellet Reasoner in Normal Client Mode	38
Pre-Classification Procedures	39
Notifying Protégé Users of Database Shutdown	39
Starting the Remote Protégé Client	39
Classification Procedures	41

Enabling the OWL Classes Tab	41
Starting the Reasoner	42
Asserting (Accepting or Rejecting) Current Classification Results	42
<i>Scenario 1: Accepting a Suggestion</i>	43
<i>Scenario 2: Rejecting a Suggestion</i>	44
Asserting (Accepting or Rejecting) Results from the Previous Classification	44
Saving the Current Classification Results as a Report	45
Exiting Protégé Without Saving the Project	46
Post-Classification Procedures	46
Index	49

ABOUT THIS GUIDE

This section introduces you to the *NCI Protégé 1.2 Workflow Manager's Guide*.

Topics in this Section

- *Purpose* on this page
- *Audience* on this page
- *Topics Covered* on page 2
- *Text Conventions Used* on page 2
- *Credits and Resources* on page 3
- *Application Support Contacts* on page 4

Purpose

This guide is a reference for workflow managers using Protégé at the National Cancer Institute (NCI). It explains how to perform basic workflow management tasks, run a Prompt database comparison, and run the Pellet reasoner.

Audience

This guide is designed for NCI workflow managers who already have a solid understanding of the NCI Protégé semantic model and NCI Thesaurus editing processes and procedures.

Note: This guide is not intended for editors (modelers) who do not perform management tasks.

Topics Covered

This section provides an overview of the topics included in this guide.

- *Chapter 1* provides an overview of BiomedGT terminology and the workflow processes and procedures used by external collaborators and NCI terminology curators.
- *Chapter 2* explains how to use the Prompt comparison tool to compare two versions of the NCI Protégé database. It also includes a reference section that explains how to interpret the results of the comparison.
- *Chapter 3* explains how to classify the NCI Protégé database, accept or reject proposed changes, and prepare and save a classification report for distribution.

Text Conventions Used

This section explains conventions used in this guide. The various typefaces represent interface components, keyboard shortcuts, toolbar buttons, dialog box options, and text that you type.

Convention	Description	Example
Bold	Highlights names of interface elements such as option buttons, check boxes, drop-down menus, menu commands, command buttons, or icons.	Click Search .
<u>URL</u>	Indicates a Web address.	http://domain.com
Text in SMALL CAPS	Indicates a keyboard shortcut.	Press ENTER.
Text in SMALL CAPS + Text in SMALL CAPS	Indicates keys that are pressed simultaneously.	Press SHIFT + CTRL.
<i>Italics</i>	Highlights newly introduced terms; also highlights references to other documents, sections, figures, and tables.	This is known as <i>cloning</i> . See <i>Figure 4.5</i> .
<i>Italic monospaced type</i>	Represents text that you type.	In the New Subset text box, enter <i>Proprietary Proteins</i> .
Note:	Highlights information of particular importance	Note: This concept is used throughout the document.
{ }	Surrounds replaceable items.	Replace {last name, first name} with the Principal Investigator's name.

Credits and Resources

The following people contributed to the development of this guide.

<i>NCI Protégé Development and Management Teams</i>		
<i>Development</i>	<i>Quality Assurance</i>	<i>Project and Product Management</i>
Robert Dionne ²	Maria-Elena Hernandez ⁶	Frank Hartel ¹
Timothy Redmond ⁵	Steven Hunter ⁷	Gilberto Fragoso ¹
Tania Tudorache ⁵	Tracy Safran ³	Sherri De Coronado ¹
Natasha Noy ⁵		Laura Roth ²
Kim Ong ⁴		Carol Creech ²
Iris Guo ⁴		
Chris Callendar ⁶		
<i>Documentation</i>	<i>Systems and Application Support</i>	<i>User Community Resources</i>
Edward VanArsdall ²	Tracy Safran ³	John Bradsher ²
	Robert Wynne ²	Liz Hahn-Dantona ²
		Nicole Thomas ²
¹ National Cancer Institute Center for Bioinformatics (NCICB)	² Lockheed Martin	³ Science Application International Corporation (SAIC)
⁴ Northrup Grumman	⁵ Stanford Medical Informatics (SMI)	⁶ University of Victoria
⁷ Ekagra Software Technologies		

Application Support Contacts

To obtain general information about NCI Protégé, receive support, or report a bug, contact NCICB Application Support.

Support Option	Additional Information
For NCI Protégé Users	
GForge	<p>If you are an NCI Protégé user, post all bug reports and requests for enhancements on GForge using the following URL: https://gforge.nci.nih.gov/tracker/?group_id=174</p> <p>For instructions on reporting bugs, see the <i>NCI Protégé 1.2 Editor's Guide</i>.</p>
Administrative tasks for NCI Protégé	<p>EVS Operations Support can provide the following services:</p> <ul style="list-style-type: none"> • Loading a database • Creating a user account • Creating or configuring a Protégé project • Installing or reinstalling the Protégé client on a user's machine <p>Note: If you need any of these services, send your request to the e-mail address under the For General Users section (below).</p>
For General Users	
E-mail	<p>Write to ncicb@pop.nci.nih.gov and include the following information:</p> <ul style="list-style-type: none"> • Your contact information, including your phone number • The name of the application that you are using • The URL (for Web-based applications) • A description of the problem and the steps required to recreate it • The text of any error messages you have received.
Web	<p>NCICB support: http://ncicbsupport.nci.nih.gov/sw/</p> <p>NCI Protégé online support forums: https://gforge.nci.nih.gov/forum/?group_id=174</p>
Telephone	<p>Local: 301-451-4384 Toll-free: 888-478-4423</p> <p>Telephone support is available Monday through Friday, 8 a.m. - 8 p.m. Eastern time, excluding government holidays.</p>

CHAPTER 1

WORKFLOW MANAGEMENT

This chapter provides an overview of BiomedGT and the processes and procedures used by external collaborators and NCI terminology curators. It also provides procedural information that is specific to workflow managers.

Topics in this Chapter

- [Overview of BiomedGT](#) on this page
- [Overview of Internal NCI Workflow](#) on page 8
- [Workflow Procedures](#) on page 12

Overview of BiomedGT

The following topics introduce you to the BiomedGT terminology and the BiomedGT wiki:

- [About BiomedGT](#)
- [About the BiomedGT Wiki](#)

About BiomedGT

Purpose

Biomedical terminologies are typically large, standalone terminologies that cover many domains in one structure and use a centralized editing model. Recognizing that this type of terminology no longer adequately supports the needs of the biomedical research community, the NCI Center for Bioinformatics (NCICB) and the cancer Biomedical Informatics Grid (caBIG™) have launched the Biomedical Grid Terminology (BiomedGT).

BiomedGT is an open, collaboratively developed terminology for translational research. This new terminology builds on the strengths of the NCI Thesaurus, including concept orientation, description logic, and public accessibility. While the initial content of

BiomedGT is based on the NCI Thesaurus, the terminology has been restructured to facilitate open content development. The goal is to evolve BiomedGT into a set of smaller federated sub-terminologies called *domain namespaces*, which cover specific areas of biomedical knowledge. As experts in the relevant research communities stabilize the content of the various domain namespaces, new content will be incorporated into the description logic-based, federated BiomedGT by the NCICB EVS and volunteer editors.

BiomedGT is intended to be the first of a new generation of biomedical terminologies that are created using an open content development approach. Such an approach empowers the wider biomedical research community to directly participate in extending and refining the terminology upon which they depend. The approach also enables maximum community participation without requiring that contributors learn complex terminology curation software and without sacrificing the quality assurance standards needed to manage biomedical terminologies.

Audience

BiomedGT is intended to address the needs of biomedical researchers who work in both basic and clinical research. Researchers can acquire content by

- downloading it in multiple formats such as OWL and RRF;
- using Web-enabled applications; or
- using Application Programming Interfaces (APIs).

About the BiomedGT Wiki

The BiomedGT wiki (<http://biomedgt.org>) is a Web-based, collaborative terminology authoring tool being developed by the NCICB and Apelon, Inc. *Figure 1.1* shows the home page of the wiki.

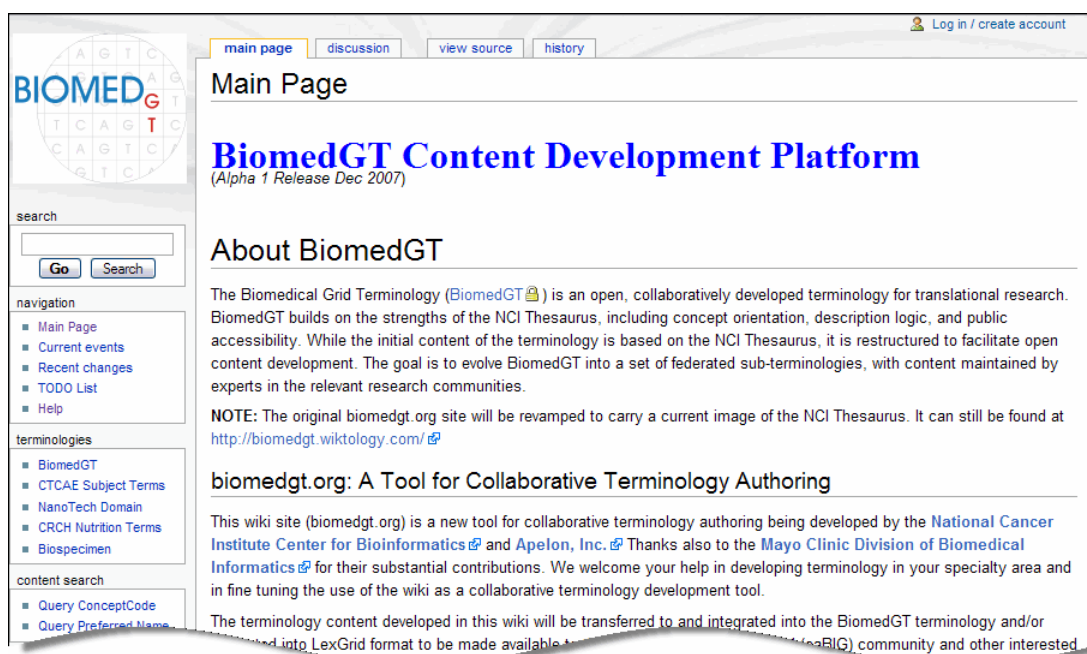


Figure 1.1 BiomedGT Home page

The BiomedGT wiki pulls terminology data from a read-only database and displays it on a Web page. This facilitates the process described in Table 1.1.

Stage	Action	Responsibility
1.	Review the read-only, published BiomedGT terminology.	Outside collaborator
2.	Using the provided Web form, submit a <i>structured</i> or <i>unstructured</i> proposal for any of the following reasons: <ul style="list-style-type: none"> • To suggest changes to existing concepts (structured) • To suggest the addition of new concepts (structured) • To direct NCI curators to a specific wiki discussion page by providing the URL (unstructured). Structured proposals are also called <i>packages</i> because they often include batches of concepts. Unstructured proposals typically refer to a wiki discussion page that might discuss specifics about a concept and then make a recommendation.	Outside collaborator
3.	Import packages and incorporate them into the vocabulary as needed. This stage has its own internal workflow processes and procedures as described in Overview of Internal NCI Workflow on page 8.	NCI BiomedGT curators
4.	Communicate with submitting collaborators to clarify suggestions or explain the outcome of decisions regarding proposals.	NCI BiomedGT curators
5.	Publish updated read-only database and display it on the Web page.	BiomedGT administrator

Table 1.1 Workflow process: BiomedGT to Protégé

For more information about BiomedGT, see the following online resources:

- BiomedGT project on GForge:
<http://gforge.nci.nih.gov/projects/biomedgt/>
- EVS section of the NCICB web site:
http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/vocabulary
- BiomedGT wiki site:
<http://biomedgt.org>

Overview of Internal NCI Workflow

As biomedical research community experts increasingly use the BiomedGT wiki for collaborative terminology development, NCI curators need a managed, organized workflow process. This topic discusses the process established for the NCI.

Primary Goals

Following are the primary goals of the BiomedGT/Protégé workflow:

- To provide capability that supports workflow manager and modeler roles, including assignments, tracking, auditing, and reporting
- To provide a means for extracting new terms from the BiomedGT wiki and incorporating them into assignments for NCI Protégé modelers
- To provide reports that help management assess modeling progress.

Workflow Roles

The internal workflow for NCI Protégé consists of two primary roles:

- **Manager**

A manager oversees the coordination and distribution of work assignments. The assignments may originate from the BiomedGT wiki or from other members of the editing team.

The manager performs the following procedures:

- Imports packages from the BiomedGT wiki
- Creates assignments for editors
- Tracks assignments
- Incorporates completed assignments into the NCI Thesaurus.

Note: The manager may also perform routine editing.

- **Modeler (Editor)**

A modeler is an NCI editor who performs routine editing tasks. Modelers cannot import packages, create assignments, or delete assignments. However, they can make *suggestions*, which are similar to assignments but which do not become assignments until a manager has reviewed them.

The modeler performs the following tasks:

- Receives assignments and accepts or rejects them
- Incorporates new or changed terminology into the NCI Thesaurus using the NCI Protégé work environment
- Maintains responsibility for the assigned concept, including any follow-up communication to the collaborator who submitted the concept for consideration
- Marks assignments as complete and submits them to the manager.

Workflow Stages

Table 1.2 provides an overview of the internal workflow stages and the responsible role for each stage.

Note: Each submitted item is called a *workflow item* until it has been assigned to a modeler. It then becomes an *assignment*.

Stage	Action	Responsibility
1.	Log into Protégé.	Manager
2.	Using the NCI Workflow tab, import packages from the BiomedGT wiki into Protégé.	Manager
3.	Create workflow items with a status of CREATED.	System
4.	Review each new workflow item and assigns it to a modeler.	Manager
5.	Change the status of the workflow item to ASSIGNED.	System
6.	Log into Protégé.	Modeler
7.	Check the NCI Workflow tab for assignments.	Modeler
8.	Accept an assignment.	Modeler
9.	Change the status of the assignment to ACCEPTED.	System
10.	Examine the proposed and existing (if any) concepts.	Modeler
11.	Perform one of the following tasks: <ul style="list-style-type: none"> Edit and save the existing concept (if any); if completed, mark the assignment as COMPLETED. Add notes to the assignment regarding the proposed concept and existing concept (if any); if completed, mark the assignment as COMPLETED. Reject the assignment. 	Modeler
12.	Change the status of the assignment to UPDATED (if not marked COMPLETED), COMPLETED, or REJECTED, depending on the action in step 11.	System
13.	Review completed and rejected assignments; address rejected assignments according to the nature of the proposal.	Manager

Table 1.2 BiomedGT wiki process for curating terminology

Note: In addition to work items generated by BiomedGT wiki imports, the manager also handles work items that have been submitted internally by modelers. The system assigns a status of SUGGESTED to internally submitted work items.

About the NCI Workflow Tab

The NCI Workflow tab is the major workflow component used by both managers and modelers. The tab layout is a table grid format, with each row representing a workflow item or assignment (*Figure 1.2*).

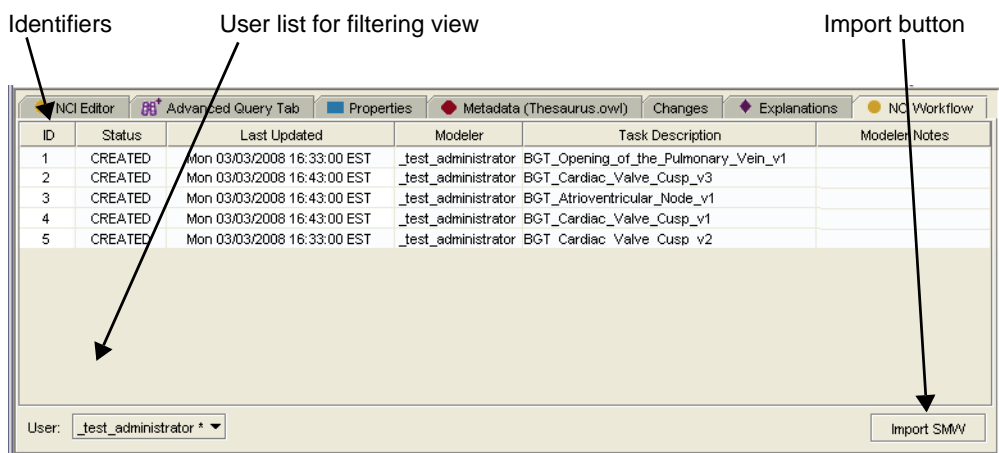


Figure 1.2 NCI Workflow tab

Protégé determines what the NCI Workflow tab displays based on the privileges of the user who is currently logged in:

- Managers can see all assigned and unassigned workflow items and can identify the assignee and status of each assigned item. They can also filter the assignments by user.
- Modelers see only workflow items that have been assigned to them.

Table 1.3 describes the fields that identify each workflow item and assignment.

Field	Description
ID	Shows a sequential number that is assigned to each workflow item. Once a number is assigned to an item, it remains the same, even if the same item is re-imported.
Status	Shows the current status of a workflow item or assignment: <ul style="list-style-type: none"> • CREATED: The manager has imported a workflow item but has not yet assigned it. • SUGGESTED: A modeler has created a workflow item that is relevant to an existing concept. • ASSIGNED: The manager has assigned a workflow item to a modeler but the modeler has not yet accepted it. • ACCEPTED: The modeler has accepted the assignment. • UPDATED: The modeler saved a change to a class or added a note but did not mark the assignment as COMPLETED. • COMPLETED: The modeler has completed the assignment.

Table 1.3 NCI Workflow tab fields

Field	Description
Last Updated	Shows the weekday, date, and time in the following format: Fri 03/07/2008 16:34:00 EST
Modeler	Initially shows the name of the user (manager) who imported the workflow item. Once assigned, this field displays the name of the assignee.
Task Description	For structured proposals, shows the name of a concept or a URL pointing to a wiki discussion page. Concept names are preceded by the letters BGT to note that the concept has been imported from the BiomedGT wiki.
Modeler Notes	Shows the most recent note added to a workflow item or assignment, followed by an ellipsis (...). If you hover the mouse pointer over the date and time, pop-up text shows the note text starting with the most recent entry.

Table 1.3 NCI Workflow tab fields (Continued)

The grid area of the NCI Workflow tab supports right-click commands. Table 1.4 describes the available commands. Some commands may appear as unavailable depending on your user privileges.

Command	Description
Edit Concept	Displays the proposed concept and existing concept (if any) on the Workflow subtab. You can then edit, add notes to, or reject the concept.
Accept Assignment	Enables you to accept an assignment.
Modify Assignment	Opens a window in which you can add notes and view previous notes.
Complete Assignment	Enables you to change the status of an assignment to COMPLETED.
Delete Assignment	Enables you to delete an assignment, depending on your user privileges and the status of the assignment: <ul style="list-style-type: none"> Managers can only delete assignments with a status of CREATED, SUGGESTED, REJECTED, or COMPLETED. Modelers can only delete SUGGESTED assignments.
Cancel	Closes the shortcut menu.

Table 1.4 NCI Workflow tab - right-click commands

Workflow Procedures

This topic details the main procedures that NCI workflow managers typically perform. Most of the procedures originate from the NCI Workflow tab.

Note: For more information about workflow manager procedures in relation to the overall NCI internal workflow process, see [Overview of Internal NCI Workflow](#) on page 8.

Importing Packages from the BiomedGT Wiki

To assign new workflow items from outside collaborators, you first need to import a package from the BiomedGT wiki. The NCI Workflow tab provides a button for this purpose.

To import a package from the BiomedGT wiki, follow these steps:

1. Log in to Protégé.
2. Click the **NCI Workflow** tab.
3. Click the **Import SMW** button in the lower right corner of the tab.

One of two messages appear, depending on whether packages are available on the wiki:

- If packages are available, a confirmation message appears. The message lists the ID numbers of any new assignments.
- If no packages are available, a message indicates that there is nothing available to import.

Note: For each newly imported item, the **Modeler** column shows the name of the user who imported that item.

Assigning (or Reassigning) a Workflow Item

After importing items from the BiomedGT wiki, you need to assign those items to modelers.

To assign a workflow item to a modeler, follow these steps:

1. Click the **NCI Workflow** tab.
2. Locate the item that you want to assign.

Tip: If you are assigning an item that you imported, the items shows your name in the **Modeler** column. To see only items with your name, select your name from the **User** drop-down list in the lower left corner of the **NCI Workflow** tab.

3. In the **Modeler** column, select a different name from the drop-down list.

Note: These steps also apply to reassigning a workflow item to a different user.

CHAPTER 2

BASELINE COMPARISON

This chapter explains how to use the Prompt comparison tool to compare two versions of the NCI Protégé database. It also includes reference topics that explain how to interpret the results of the comparison.

Topics in this Chapter

- [Overview of Baseline Comparison](#) on this page
- [Pre-Comparison Procedures](#) on page 15
- [Comparison Procedures](#) on page 17
- [Preparatory Procedures for the Next Comparison](#) on page 28
- [Post-Comparison Procedures](#) on page 34

Overview of Baseline Comparison

As a workflow manager, you will need to regularly compare two versions of the Protégé database, examine the results, and resolve any conflicts. This topic introduces you to the process and the tool that you will use.

About the Comparison Process

An ontology maintained in Protégé is known as a *project* and is stored as a file with a `.pprj` extension. During each editing cycle, editors use the master Protégé project, which is maintained by Operations staff. This master project represents the current state of the database.

Each time you run a baseline comparison, you export a copy of the master project file and use it as a baseline during the next comparison cycle. This enables you to identify changes to the database and accept or reject them.

As editors make changes to the database, their changes are stored in a file called an *annotation ontology* (also called a change ontology). This file serves as an audit trail of

changes as they take place. During a comparison, Prompt uses this file to determine changes between the current project file and the most recent baseline file. Once you accept or reject the changes in the annotation ontology, you use a menu command to clean up the ontology, then close Protégé and delete specified files from the host machine to remove the record of changes that you have already reviewed.

Table 2.1 summarizes the process stages for a baseline comparison.

Stage	Action	Responsibility
<i>Pre-Comparison Procedures</i> on page 15		
1.	Send e-mail or chat to alert modelers that the NCI Protégé database will be shut down and that they need to log out.	Manager
2.	Log out of the database.	Modeler
3.	Log into the Protégé host machine.	Manager
4.	Shut down the Protégé server.	Manager
5.	Shut down the Explanation server	Manager
6.	Start the Protégé client and connect to the BiomedGT project.	Manager
<i>Comparison Procedures</i> on page 17		
7.	Configure the Protégé client to display the Prompt tab.	Manager
8.	Using Prompt, compare the current version of the database project (.prj file) against the most recent baseline project.	Manager
9.	Accept or reject changes.	Manager
10.	Save the concept history by exporting a copy of the current master database.	Manager
11.	Disable the Prompt tab.	Manager
<i>Preparatory Procedures for the Next Comparison</i> on page 28		
12.	Change the baseline version shown on the Metadata tab.	Manager
13.	Clean up the ontology and archive the current version.	Manager
14.	Export a copy of the current project file for use as a baseline for the next database comparison.	Manager
15.	(Publication baseline only) Start the Explanation server, run the Pellet reasoner, and save the inferred terminology file.	Manager
16.	Exit the remote Protégé client without saving the project.	Manager
17.	Delete all remaining extraneous files from the host machine.	Manager
<i>Post-Comparison Procedures</i> on page 34		
18.	Start the Explanation server (or restart if it was not previously shut down).	Manager
19.	Start the Protégé server and verify that it is running.	Manager
20.	Verify that the Explanation tab is communicating with the Explanation server.	Manager
21.	Log off of the Protégé host computer.	Manager

Table 2.1 Database comparison process

About the Prompt Plug-in

In NCI Protégé, the Prompt tab operates the Prompt plug-in, which enables you to compare two different versions of the same ontology, examine the results, and accept or reject changes. When you run a comparison, the current ontology is considered the new version, and the external file that you specify is treated as the older version. This helps you to establish a version control system.

Prompt also has other capabilities that are currently not in use at the NCI. For more information about Prompt, visit <http://protege.cim3.net/cgi-bin/wiki.pl?Prompt>.

Pre-Comparison Procedures

Before you run a baseline comparison, be sure that you complete the following procedures:

- [Notifying Protégé Users of Database Shutdown](#) on this page
- [Starting the Remote Protégé Client](#) on this page

Notifying Protégé Users of Database Shutdown

Before running a baseline comparison, let all users know that the database will be shut down. This gives them a chance to save their work and log off.

To notify users of a database shutdown, follow these steps:

1. Using e-mail or chat, announce that the database will be shut down.
2. Log out of the Protégé client on your own machine.

Starting the Remote Protégé Client

Note: In the following instructions, the current NCI Protégé directory is represented as `Protege.Server-x.y.z`. When typing commands, substitute the directory name with the current Protégé version number (**Example:** `Protégé.Server-1.2.1.21`).

To start the Protégé client on the remote server, follow these steps:

1. Log in to the Protégé host computer:
 - a. Open Reflection X.
 - b. Enter your user name and password.
2. Shut down the Protégé server:
 - a. For **BiomedGT**, type the following command:

```
cd /usr/local/protege/Protege.Server-x.y.z
```

or

For the **NCI Thesaurus**, type the following command:

```
cd /usr/local/protegeqa/NCIThesaurusGroup/Protege.Server-x.y.z
```

- b. Press ENTER.
 - c. For **BiomedGT**, type the following command:

```
./shutdown_protege_server.sh localhost:2633
```

or

For the **NCI Thesaurus**, type the following command:

```
./shutdown_protege_server.sh localhost:2822
```
 - d. Press ENTER.
3. Shut down the Explanation server:
 - a. For **BiomedGT**, type the following command:

```
cd /usr/local/protege/Explanation.Server-x.y.z
```

or

For the **NCI Thesaurus**, type the following command:

```
cd /usr/local/protegeqa/NCIThesaurusGroup/Explanation.Server-x.y.z
```
 - a. Press ENTER.
 - b. Type the following command:

```
./stop_explanation_server.sh
```
 - c. Press ENTER.
4. Start the remote Protégé client:
 - a. For **BiomedGT**, type the following command:

```
cd /usr/local/protege/Protege.Client-x.y.z
```

or

For the **NCI Thesaurus**, type the following command:

```
cd /usr/local/protegeqa/NCIThesaurusGroup/Protege.Client-x.y.z
```
 - b. Press ENTER.
 - c. For **BiomedGT**, type the following command:

```
./run_protege_BGT.sh
```

or

For the **NCI Thesaurus**, type the following command:

```
./run_protege_NCIT.sh
```

- d. Press ENTER.
- e. When prompted to select a project, do either of the following:
 - Select **BiomedGT**.
 - Select **Thesaurus**.

Comparison Procedures

After you start the remote Protégé client, perform the following procedures in the following order:

- [Enabling the Prompt Tab](#) on page 17
- [Starting the Comparison](#) on page 18
- [Tools for Interpreting the Comparison Results](#) on page 19
- [Accepting and Rejecting Changes](#) on page 26
- [Saving the Concept History](#) on page 27
- [Disabling the Prompt Tab](#) on page 27

Enabling the Prompt Tab

Use the Prompt tab to compare the current Protégé project file to a specified baseline file (usually the most recent version of the database). This tab is not enabled by default.

To enable the Prompt tab, follow these steps:

1. Open the Configure File window using the following menu command:
Project > Configure...
2. In the **Visible** column, check the box to the left of the **Prompt** widget.
3. Click **OK** to close the Configure File window.

The Prompt tab now appears on the far right ([Figure 2.1](#)).

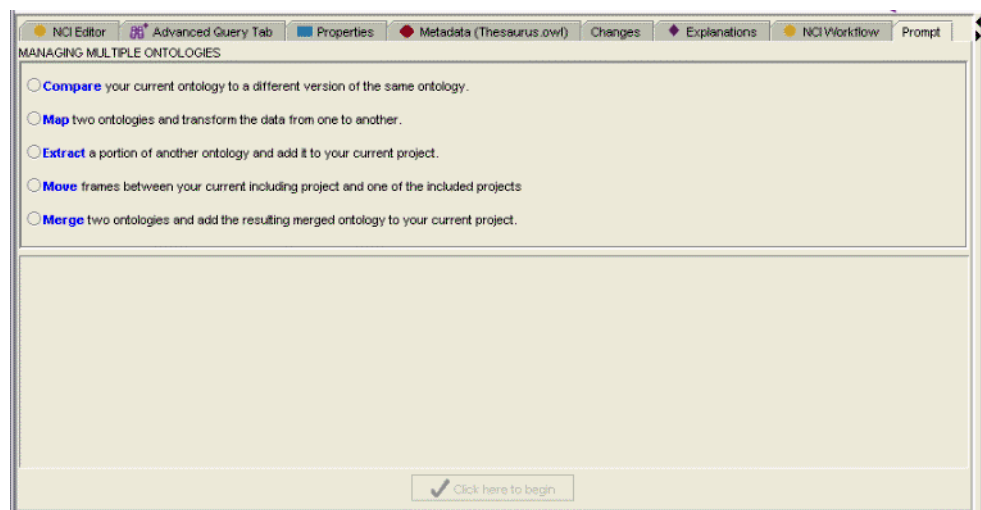


Figure 2.1 Prompt tab


Starting the Comparison

Prerequisite procedure: [Enabling the Prompt Tab](#) on page 17

To initiate a baseline comparison using the Prompt tab, follow these steps:

1. Click the **Prompt** tab.
2. Select the following option (if it is not already selected):
Compare your current ontology to a different version of the same ontology.

New text entry fields and check boxes appear in the bottom half of the tab.

3. In the field labeled **Choose the version to compare with the current project**, do one of the following:
 - Enter the path and name of the file that represents the most recent baseline ([Figure 2.2](#)); or
 - Select a file by clicking the **Browse for File** button  on the right, just above the field ([Figure 2.2](#)).
4. Leave the remaining settings as they are:
 - **Display changes for included frames** is checked by default.
 - The field labeled **Select a slot containing a concept ID** is empty.

Note: Although you can theoretically use this field and its accompanying check box to add a slot and specify that a comparison be run only on a specific concept, the feature does not currently work as expected.

5. At the bottom of the tab, click the button labeled **Click here to begin** ([Figure 2.2](#)).

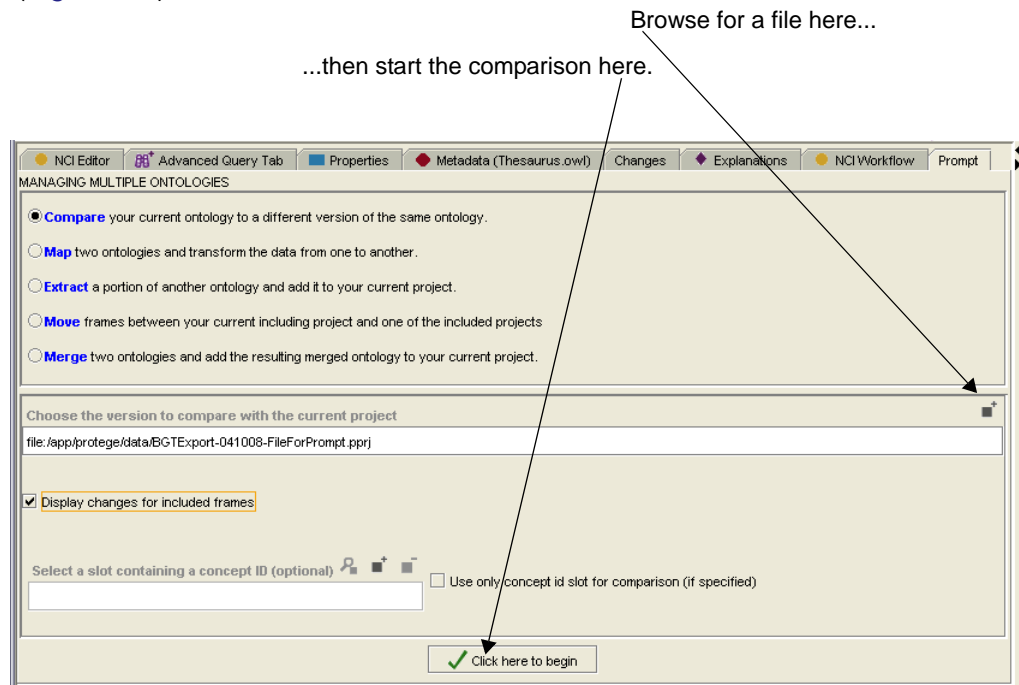


Figure 2.2 Comparing two versions of the database

A small status window opens in the foreground. When the comparison finishes running, the window disappears and a set of subtabs appears. The prominent Tree View subtab shows the two ontologies in a single tree hierarchy (*Figure 2.3*).

For more information about reading the subtab display, see *Tools for Interpreting the Comparison Results* on page 19.

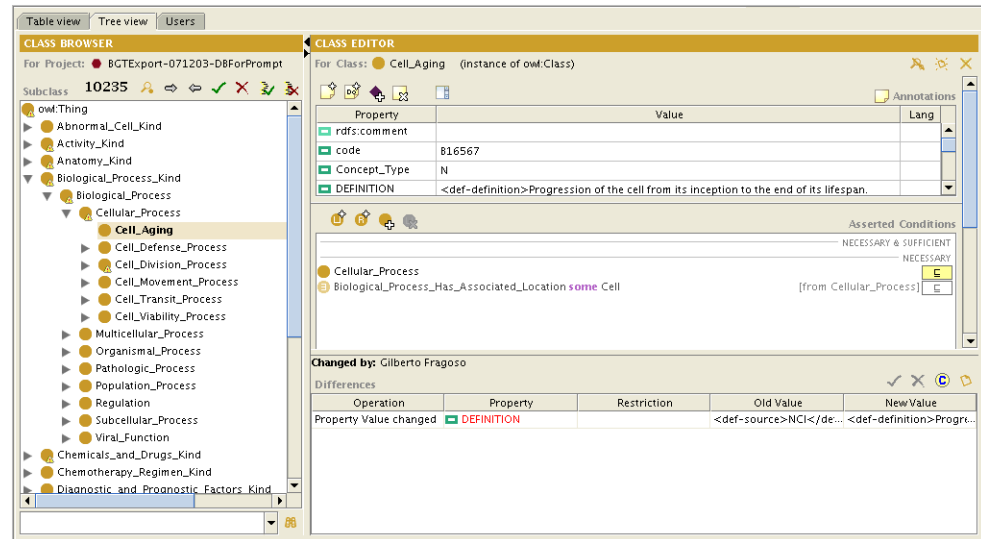


Figure 2.3 Prompt comparison results - Tree View subtab

Tools for Interpreting the Comparison Results

This topic explains the three Prompt subtabs used for interpreting database comparison results. The subtabs are discussed in the order that they appear on the interface:

- [About the Table View Subtab](#) on page 20
- [About the Tree View Subtab](#) on page 22
- [About the Users Subtab](#) on page 24

About the Table View Subtab

The Table View subtab (*Figure 2.4*) shows all of the frames in the two compared ontologies and lists their similarities and differences.

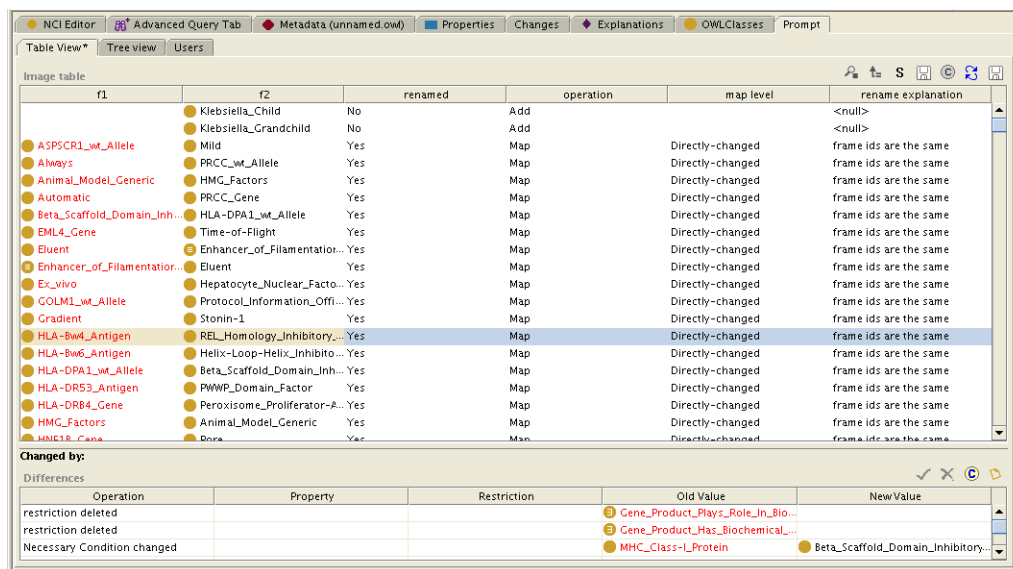


Figure 2.4 Table View subtab

Table 2.2 describes the upper and lower panels of the Table View subtab.

Panel/Column	Description
Image Table	
f1	Represents the frames in the original ontology.
f2	Represents the frames in the changed ontology.
renamed	Indicates whether a class was renamed (Yes or No).
operation	Indicates the nature of the change: <ul style="list-style-type: none"> Add: A new class was added to the ontology. Map: The relationship between the class and certain other classes was changed.
map level	Indicates the type of change: <ul style="list-style-type: none"> Directly-changed: A direct change to the class that does not affect any other class. Isomorphic change: A change to the child class that causes a change in its parent class.
rename explanation	Provides an explanation for any classes that have been renamed; otherwise displays the following message: <i>frame name and type are the same.</i>

Table 2.2 Table View subtab - panel descriptions

Panel/Column	Description
Differences	
Operation	Explains the nature of the change (for example, <i>Property Value changed</i>).
Property	Describes which property was changed (if any).
Restriction	Describes which restriction was changed (if any).
Old Value	Shows the value before the change.
New Value	Shows the new value resulting from the change.

Table 2.2 Table View subtab - panel descriptions (Continued)

Table 2.3 describes the buttons used in each Table View panel.










Button	Description
Image table	
View Frames 	Opens a window in which you can view annotations, asserted conditions, and disjoints. The window is essentially a standalone version of the Class Editor.
Back-references for Frames 	Opens a window that shows references (frames, slots, and facets) to the selected class and enables you to drill down to further references.
Show Sources 	Opens a window in which you can view Source classes, Source slots, and Source instances.
Save to File 	Saves the change as a tab-delimited .diff file. Note: This is <i>not</i> the button used to save the NCI concept history.
Change Statistics 	Opens a window showing a summary of change statistics, including additions, deletions, splits, merges, direct changes, and total changes.
Refresh Table 	Refreshes the Image table panel.
Save in NCI DB Format 	Saves the concept history and stores it in a database table. For more information, see Saving the Concept History on page 27.
Differences	
Accept/Reject Diff 	Enable you to accept or reject differences that appear in the Differences panel. This panel shows mainly property changes.
Compare Changes 	Opens a Compare Versions window in which you can examine properties, asserted conditions, and disjoints for both the original class the changed class.

Table 2.3 Table View subtab - button descriptions


Button	Description
View Annotations 	Opens a Change annotations window showing the following change information: <ul style="list-style-type: none"> Action: Type of change Description: Nature of the change Author: Person who made the change Created: Date and time that the change was made

Table 2.3 Table View subtab - button descriptions (Continued)

About the Tree View Subtab

The Tree View subtab (*Figure 2.3* on page 19) is the prominent subtab immediately after the comparison process finishes running. This subtab shows the two compared ontologies in a single hierarchy.

The Tree View subtab displays three panels:

- The Class Browser panel displays the combined hierarchy. The number appearing just above the hierarchy displays the total number of changed classes since the previous baseline. Changed classes are shown in bold black text, and newly added classes are shown in blue, underlined text (*Figure 2.5*). Deleted classes are shown in red, strikethrough text.

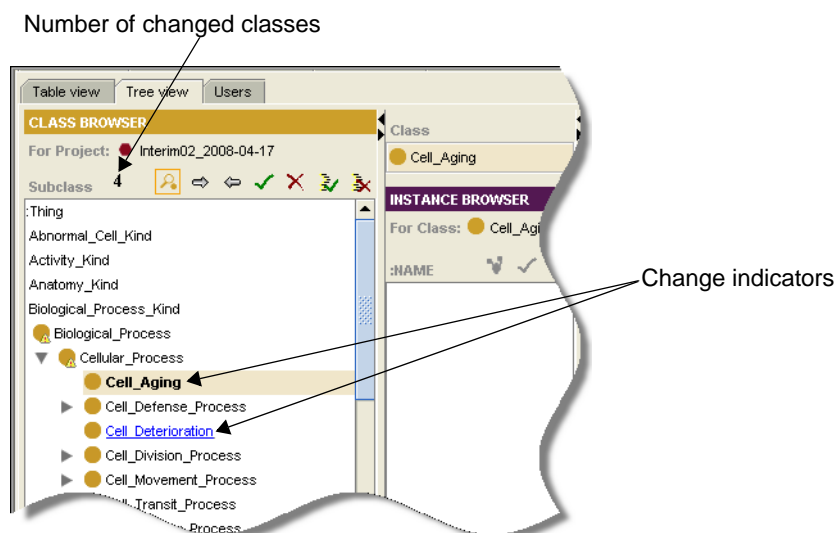


Figure 2.5 Class Browser - change indicators

- The Instance Browser panel shows the class that is currently selected in the Class Browser. Using the buttons on this panel, you can view references to an instance, accept an instance, or reject an instance.
- The Class Editor panel enables you to edit the selected instance.

To control the viewing space on the Tree View subtab, use the Collapse and Expand buttons at the edge of each panel (*Figure 2.6*).

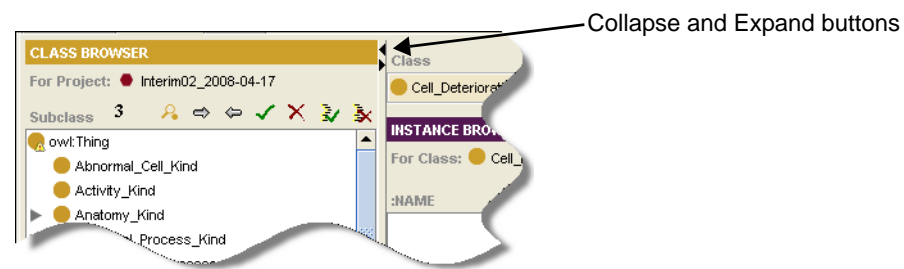


Figure 2.6 Collapse and Expand buttons

Table 2.4 describes the buttons used in each of the Tree View subtab panels.










Button	Description
Class Browser	
View Class 	Opens the Class Editor in a new window where you can perform all of the functions of the Class Editor panel (view, edit, and add notes to the selected class).
Next/Previous Change 	Used to browse forward and backward through changes in the ontology, one change at a time.
Accept/Reject Change 	Used to accept or reject the change that is currently selected in the class hierarchy.
Accept/Reject All Changes 	Used to accept or reject <i>all</i> changes shown for the current comparison.
Instance Browser	
View References to Instance 	Opens a window that shows references (frames, slots, and facets) to the selected instance.
Accept/Reject Instance Change 	Used to accept or reject a change to the selected instance.
Class Editor	
Show/Hide, Create, or Delete Class Notes 	Enables you to create class notes, show and hide them, and delete them.
Accept/Reject Diff 	Enable you to accept or reject differences that appear in the Differences panel. This panel shows mainly property changes.
Compare Changes 	Opens a Compare Versions window in which you can examine properties, asserted conditions, and disjoints for both the original class the changed class.

Table 2.4 Tree View subtab - button descriptions for each panel


Button	Description
View Annotations 	Opens a Change annotations window showing the following change information: <ul style="list-style-type: none"> • Action: Type of change • Description: Nature of the change • Author: Person who made the change • Created: Date and time that the change was made

Table 2.4 Tree View subtab - button descriptions for each panel (Continued)

About the Users Subtab

The Users subtab lists each user who made changes to the ontology. The Users with changes panel on the left has four columns showing the following information:

- User shows the user's name.
- Changed shows the number of changes for each user.
- Conflicts shows the number of conflicts caused by a user's changes.
- Conflicts with shows the name of the user with conflicting changes.

The first of the two Ontology component panels on the right shows a list of changes made by a selected user. The second panel shows conflicts (if any) between the selected user's changes and another user's changes.

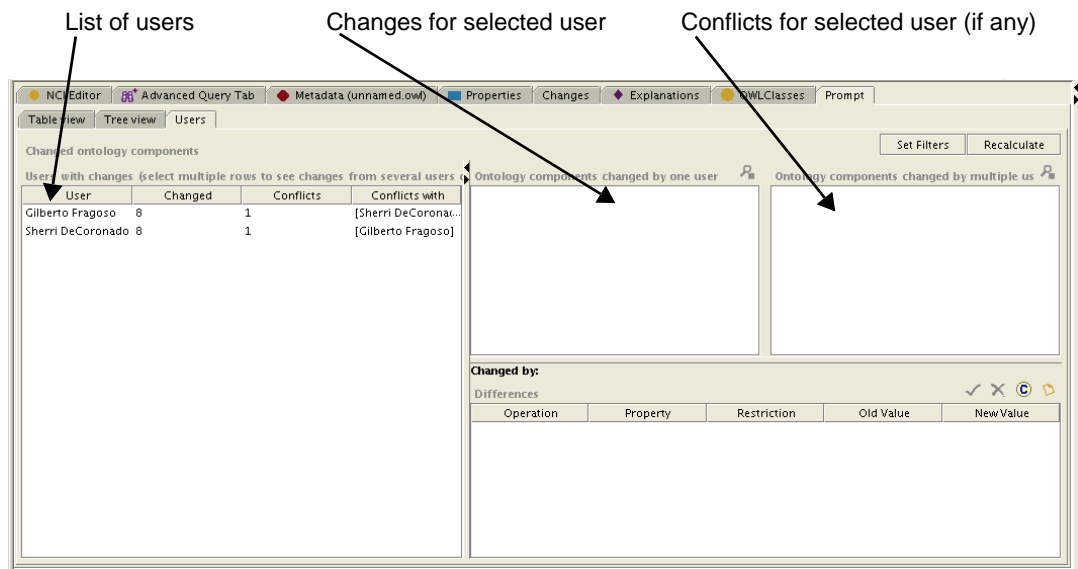


Figure 2.7 Users subtab

Table 2.5 describes the buttons used on the Users subtab.


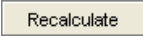




Button	Description
Set Filters 	Opens a window used to control what the panel displays: classes, properties, individuals, and anonymous ontology components.
Recalculate 	Updates the total number of changes. This is useful when you have accepted individual changes and want to update the total.
Ontology components panel	
View Changed Ontology Component 	Opens a window showing annotations for the item that is currently selected in either of the Ontology component panes. Select an item, then click this button to view its annotations.
Differences panel	
Accept/Reject Diff 	Enable you to accept or reject differences that appear in the Differences panel. This panel shows mainly property changes.
Compare Changes 	Opens a Compare Versions window in which you can examine properties, asserted conditions, and disjoints for both the original class the changed class.
View Annotations 	Opens a Change annotations window showing the following change information: <ul style="list-style-type: none"> • Action: Type of change • Description: Nature of the change • Author: Person who made the change • Created: Date and time that the change was made

Table 2.5 Users subtab - Button descriptions


Accepting and Rejecting Changes

Although you can accept or reject changes from various panels in the three Prompt subtabs, the following procedures focus on the toolbar found on the Tree View subtab.

Accepting a Single Change

To accept a single change, follow these steps:

1. Select the **Tree View** tab.
2. In the Class Browser, navigate to the change that you want to accept.


Tip: To identify changes in the hierarchy, look for a small triangular symbol with an exclamation mark inside it.
3. Select the change.
4. Using the toolbar at the top of the Class Browser, click the **Accept** button .

The name of the changed class no longer appears in bold or underlined text.

Accepting All Changes

When you run the Accept All command, you accept all changes identified in the current comparison. If you have already accepted or rejected individual changes, running this command accepts all of the *remaining* changes.

To accept all changes, follow these steps:


1. Select the **Tree View** tab.
2. Using the toolbar at the top of the Class Browser, click the **Accept All** button .

The names of the changed classes no longer appear in bold or underlined text.

Rejecting a Single Change

To reject a single change, follow these steps:

1. Select the **Tree View** tab.
2. In the Class Browser, navigate to the change that you want to reject.

Tip: To identify changes in the hierarchy, look for a small triangular symbol with an exclamation mark inside it.
3. Select the change.
4. Using the toolbar at the top of the Class Browser, click the **Reject** button .

The name of the changed class no longer appears in bold or underlined text.

Rejecting All Changes

When you run the Reject All command, you reject all changes identified in the current comparison. If you have already accepted or rejected individual changes, running this command accepts all of the *remaining* changes.

To reject all changes, follow these steps:

1. Click the **Tree View** tab.
2. Using the toolbar at the top of the Class Browser, click the **Reject All** button



The names of the changed classes no longer appear in bold or underlined text.

Saving the Concept History

The *concept history* provides a record of the edit actions (creation, merge, split, modification, and retirement) performed on a concept throughout its inclusion in the terminology. The history includes the dates on which each edit action occurred. The concept history is saved in a database format and is intended for use in the publication of the terminology to an online server such as LexBig.

To save the concept history, follow these steps:

1. Click the **Table View** subtab.
2. In the Image table panel (upper panel), locate the toolbar on the far right.

Note: Before you complete the next step, note that there are two **Save** buttons in the toolbar. Although the correct button is on the far right, it could be moved in future releases. To ensure that you use the correct button, hover the mouse pointer over the button and make sure that the popup tool tip reads *Save in NCI DB Format*.

3. Click the second of the two **Save** buttons—the one on the far right.

The concept history is saved to a database table.

Disabling the Prompt Tab

After completing a database comparison, disable the Prompt tab to ensure that it does not appear in the main Protégé project.

To disable the Prompt tab, follow these steps:

1. Select the following menu command: **Project > Configure...**

The Configure File window opens.

2. In the **Visible** column, clear the check box to the left of the **Prompt** widget.
3. Click **OK** to close the Configure File window.

The Prompt tab is no longer visible.

Preparatory Procedures for the Next Comparison

After you run a comparison, review the results, and accept or reject changes, perform the following procedures to prepare for the next comparison.

- [Managing Baseline Versions](#) on this page
- [Managing the Annotation Ontology](#) on page 30
- [Exporting a Project File](#) on page 31
- [Deleting Extraneous Information](#) on page 33

Managing Baseline Versions

The following topics explain the naming convention for a baseline version and provide steps for changing the version number:

- [About the Naming Convention for a Baseline Version](#) on this page
- [Changing the Baseline Version Number for the Next Baseline Build](#) on page 29

About the Naming Convention for a Baseline Version

When you export the current baseline file for use during the next comparison, you use a specified convention for the `owl:versionInfo` property shown on the Protégé Metadata tab. You also use a specified naming convention for the exported file.

Specifying the owl:versionInfo Property Value

Assume that today is Friday, April 4, 2008, and you have just run a comparison for the BiomedGT project. This is your first comparison in April 2008. You now need to change the `owl:versionInfo` property displayed on the Metadata tab to express the value of the *next* baseline build.

The value of the `owl:versionInfo` property follows this format:

- year of the current baseline build: *yy*
- a period: `.`
- month of the current baseline build: *mm*
- a lowercase letter representing the *next* baseline build; using the first example, if the current build is letter *a*, then the next build will be letter *b*.

Using the above example, if you just ran your first comparison of the month on April 4, 2008, then the `owl:versionInfo` property will show a letter *b* to represent the *second* build of the month: `08.04b`.

Naming the Exported File

When you export the project file, use the following string for the file name:

- project name: BiomedGT or Thesaurus
- an underscore: _
- the year (yy), month (mm), and day (dd) of the *current* baseline build: yymmdd
- an underscore: _
- the version number from the Metadata tab.

Using the April 4, 2008 example, you have just run the first build of the month. Thus, when you export the file, you give it the following name:

```
BiomedGT_080404_08.04a
```

For more information about the naming convention for the `owl:versionInfo` property, see [Specifying the owl:versionInfo Property Value](#) on page 28.

Changing the Baseline Version Number for the Next Baseline Build

Prerequisite topic: [About the Naming Convention for a Baseline Version](#) on page 28

For baseline version tracking, change the value of the `owl:versionInfo` property on the Protégé Metadata tab to show the current year, the current month, and the baseline version of the *next* baseline build.

To change the value of the `owl:versionInfo` property, follow these steps:

1. Click the **Metadata** tab.

The panel on the right shows two columns: **Property** and **Value**. The first listed property, `owl:versionInfo`, shows the baseline version number.

2. Double-click the value column for `owl:versionInfo`.
3. Change the value to reflect the next baseline version.

For example, if you just ran your second comparison on May 11, 2008, and plan to run a third comparison next week, the value would `yy.mm.c`, or `08.05c`.

4. Click outside of the **Value** column to deselect the property.

Managing the Annotation Ontology

When you first log in to Protégé and select a project, you may notice a project named *Annotation_BiomedGT* (*do not edit*). This is the *annotation ontology*. This topic explains the purpose of the annotation ontology and explains how to clean up the ontology.

About the Annotation Ontology

The *annotation ontology* (also called a change ontology) that tracks all changes and shows who made them. While using Protégé, you can view the current annotation ontology using the Changes tab.

The annotation ontology file is accompanied by two other files with identical names but different extensions (.rdf and .rdfs). The file names are prefixed with the word *annotation*, an underscore, and the project name.

For example, for a project called *BiomedGT*, you might see the following files:

- annotation_BiomedGT.pprj
- annotation_BiomedGT.rdf
- annotation_BiomedGT.rdfs

Prompt reads the annotation ontology to determine changes in the current baseline. You run a cleanup procedure to clear this information so that Protégé will start a new version and use the new version during the next comparison.

Cleaning Up the Annotation Ontology

NCI Protégé includes a utility for cleaning up the annotation ontology. The utility is available from the Change menu.

To clean up the annotation ontology, follow these steps:

1. Select the following menu command: **Change > Clean Up Changes Ontology**

The Clean Up Changes Ontology window opens (*Figure 2.8*).

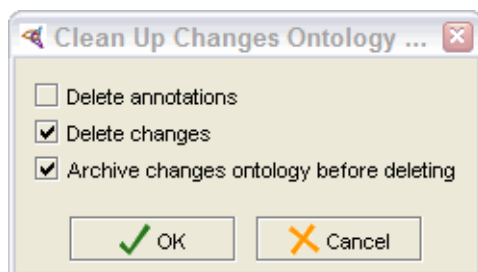


Figure 2.8 Clean-up Changes Ontology window

2. Check the appropriate boxes:
 - a. **Leave the Delete annotations** box unchecked.

Caution: Checking this box deletes all annotations in the ontology. Leave it unchecked to ensure that running comments from the Collaborative Protégé plug-in are retained from baseline to baseline.

- b. Check the **Delete changes** box if it is not already checked.

This step deletes all changes in the ontology.

- c. Check the **Archive changes ontology before deleting** box if it is not already checked.

This step creates an archive of the current version of the ontology and saves the archive in a subdirectory of the folder where the annotation ontology is stored. The subdirectory has the following name:

```
annotation_prjName.parc
```

`prjName` represents the name of the main project.

This `parc` subdirectory creates a subfolder for each archive of the annotation ontology. Each folder is named according to the date and time:

```
YYYY.MM.DD_HH.MM.SS
```

A folder created on June 28, 2008, at 10:45 a.m. would be called
2008.06.28_10.44.22.

3. Click **OK** to start the cleanup.

Once the cleanup is finished, a confirmation message appears.

4. Click **Close** to close the message window.

5. To save the cleared change ontology, select the following menu command:

Change > Save Changes Ontology

Exporting a Project File

Prerequisite topic: [About the Naming Convention for a Baseline Version](#) on page 28

At the end of each baseline comparison cycle, you export a copy of the current master database by converting it to an OWL file project. You then use the exported file as a baseline for comparison during the next cycle. This enables you to identify changes that have been made between cycles and accept or reject them.

To export the project file, follow these steps:

1. Select the following menu command: **File > Convert Project to Format**

The Select Format window opens.

2. Select **OWL / RDF Files** from the list, then click **OK**.

The OWL / RDF Files window opens. In the **OWL file name or URL** field, note that the OWL file is given the same name as the project (`.pproj`) file.

3. Ensure that the name shown in the **OWL file name or URL** field is in the following format:

```
{project name}_yymmdda.owl
```

where the letter `a` represents the first build of the month, `b` represents the second, `c` represents the third, and so on.

4. If the file name does not follow the established convention, change it in the **Project** field, then verify the change in the **OWL file name or URL** field.
5. Leave the Language set to **RDF/XML-ABBREV**.
6. Click **OK**.
7. Exit the Protégé client without saving the project.

Saving the Inferred Terminology File (Publication Baseline Only)

After exporting the project file, you need to run the reasoner and save the inferred hierarchy.

Note: If you have shut down the explanation server before running the comparison, ensure that you restart it before performing this task.


To save the inferred terminology file, follow these steps:

1. Start the reasoner from the OWL Classes tab by selecting the following menu command:

OWL > Classify taxonomy

For more information about running the reasoner, see *Chapter 3, Classification*, on page 37.

2. After the reasoner runs, click the **Save** button on the Inferred Hierarchy panel to save the inferred hierarchy.

Note: If you do not see the inferred hierarchy panel, click the button on the far right immediately beneath the **Subclass Explorer > Asserted Hierarchy** panel .

3. Save the file using the naming convention described in [About the Naming Convention for a Baseline Version](#) on page 28.

Be sure to append the word *Inferred* to the name (for example, BiomedGT-Inferred_080404a).

Deleting Extraneous Information

During a baseline comparison, Prompt identifies changes by using information stored both in the Protégé master database and in external files on the host computer. After running a comparison and addressing all of the changes, you then clean up the database and delete specific external files in preparation for the next comparison.

About the Accepts Files and Journal File

After running a Prompt comparison, you need to close Protégé and delete several files. These include a file called `Thesaurus.pjrn`, also known as a *journal* file, and any files that match the project name but have the word *accepts* appended to the name. Usually there are three accepts files with the extensions `.pprj`, `.pont`, and `.pins`.

For example, for a project called BiomedGT, you might see the following files:

- `BiomedGT-accepts.pprj`
- `BiomedGT-accepts.pont`
- `BiomedGT-accepts.pins`

Deleting the Accepts Files and Journal File

Prerequisite topic: [About the Accepts Files and Journal File](#) on page 33

Caution: Ensure that Protégé is closed before deleting the files from the host machine.

At this point, delete the Accepts files and Journal file from the host machine. Except for the journal file with the `.pjrn` extension, the files listed below have the project name *BiomedGT*. If your project name is different, delete the files that match your project name but have the same extensions.

Using the file manager, delete the following files **for BiomedGT**:

- `BiomedGT-accepts.pprj`
- `BiomedGT-accepts.pont`
- `BiomedGT-accepts.pins`
- `Thesaurus.pjrn`

Delete the following files **for the NCI Thesaurus**:

- `Thesaurus-accepts.pprj`
- `Thesaurus-accepts.pont`
- `Thesaurus-accepts.pins`

Post-Comparison Procedures

To restart the servers and log out of the Protégé host computer, follow these steps:

1. Start the Explanation server:

- a. For **BiomedGT**, type the following command:

```
cd /usr/local/protége/Explanation.Server-x.y.z
```

or

For the **NCI Thesaurus**, type the following command:

```
cd /usr/local/protégeqa/NCIThesaurusGroup/Explanation.Server-
x.y.z
```

- b. Press ENTER.

- c. For **BiomedGT**, type the following command:

```
./start_explanation_server.sh -p 8090 --url jdbc:mysql//
cbiows501.nci.nih.gov:3610/protége -t BiomedGT_Baseline -U
protége -P pradmin#652
```

or

For the **NCI Thesaurus**, type the following command:

```
./start_explanation_server.sh -p 8092 --url jdbc:mysql//
cbiows501.nci.nih.gov:3610/NCIt_Demo -t Thesaurus_Baseline
-U protége -P pradmin#652
```

- d. Press ENTER.

If you need to *restart* the server for any reason, follow these steps. Otherwise, skip to step 2:

- a. For **BiomedGT**, type the following command:

```
cd /usr/local/protége/Explanation.Server-x.y.z
```

or

For the **NCI Thesaurus**, type the following command:

```
cd /usr/local/protégeqa/NCIThesaurusGroup/Explanation
.Server-x.y.z
```

- b. Press ENTER.

- c. Type the following command:

```
./restart_explanation_server.sh
```

- d. Press ENTER.

2. To confirm that the explanation server started (or restarted) successfully, type the following at the appropriate `Explanation.Server` command prompt:

```
tail -f nohup.out
```

When the server has started (or restarted), the following message appears:
Server started listening on port 8090 (BGT) or 8092 (NCIt)

3. Start the Protégé server:

- a. For **BiomedGT**, type the following command:

```
cd /usr/local/protege/Protege.Server-x.y.z
```

or

For the **NCI Thesaurus**, type the following command:

```
cd /usr/local/protegeqa/NCIThesaurusGroup/Protege.Server-x.y.z
```

- b. Press ENTER.

- c. Type the following command:

```
./run_protege_server_nci.sh
```

- d. Press ENTER.

4. Verify that the Protégé server is running:

- e. For **BiomedGT**, type the following command:

```
cd /usr/local/protege/Protege.Server-x.y.z
```

or

For the **NCI Thesaurus**, type the following command:

```
cd /usr/local/protegeqa/NCIThesaurusGroup/Protege.Server-x.y.z
```

- f. Press ENTER.

- g. Type the following command:

```
tail -f console.txt
```

- h. Press ENTER.

5. Verify that the Explanations tab is communicating with the Explanation server:

- a. Re-open the Protégé client.

- b. Click the **Explanations** tab.

- c. Using the Class Browser on the left, select a class.

The Named Superclass and Named Subclasses panels (in the center of the tab) show the selected class and its children.

- d. Select a named subclass.

An explanation appears in the far right panel. If no explanation appears, or if an error message appears, then the communication failed. If this happens

Note: A communication failure between the Explanations tab and the Explanation server is considered an serious error. The problem should be immediately reported to the EVS Operations team by phone instead of through GForge.

6. Log off of the host computer using Reflection X.

CHAPTER 3 CLASSIFICATION

This chapter explains how to classify the NCI Protégé database, accept or reject proposed changes, and prepare and save a classification report for distribution.

Topics in this Chapter

- [Overview](#) on this page
- [Pre-Classification Procedures](#) on page 39
- [Classification Procedures](#) on page 41
- [Post-Classification Procedures](#) on page 46

Overview

To perform a periodic inferencing analysis of the NCI Protégé database, you run an automated classification process using a tool called a *reasoner*. The reasoner examines the relationships in the class hierarchy, computes the subclass relations between every named class, and determines whether possible modeling errors exist.

About the Classification Process

You can run the reasoner in either normal client server mode or remote server mode with the Protégé server shut down. Table 3.1 summarizes the process stages for classification in remote server mode.

Stage	Action	Responsibility
Pre-Classification Procedures on page 39		
1.	Send e-mail or chat to alert modelers that the NCI Protégé database will be shut down and that they need to log out.	Manager
2.	Log out of the database.	Modeler

Table 3.1 Classification process

Stage	Action	Responsibility
3.	Log into the Protégé host computer.	Manager
4.	Shut down the Protégé server.	Manager
5.	Start the remote Protégé client and connect to the BiomedGT project.	Manager
<i>Classification Procedures</i> on page 41		
6.	Configure the remote Protégé client to display the OWL Classes tab.	Manager
7.	Start the reasoner.	Manager
8.	Review the results, accepting or rejecting retreeing suggestions according to established guidelines.	Manager
9.	Using the previous classification report reviewed and discussed by the modelers, accept or reject items as determined.	Manager
10.	Save the current classification results as an external text-delimited file, then open the file in Microsoft Excel and save it as an Excel workbook for circulation among the team.	Manager
11.	Exit the Protégé project without saving it.	Manager
<i>Post-Classification Procedures</i> on page 46		
12.	(Optional) Restart the Explanation server to make sure that it is synchronized with the database.	Manager
13.	Restart the Protégé server and verify that it is running.	Manager
14.	Log off of the Protégé host computer.	Manager

Table 3.1 Classification process (Continued)

About the Pellet Reasoner

NCI Protégé uses the Pellet reasoner, an open-source tool developed and commercially supported by Clark & Parsia LLC (<http://clarkparsia.com/>). As stated on the Pellet Web site, “Pellet supports the full expressivity of OWL DL, including reasoning about nominals (enumerated classes).”

To learn more about the Pellet reasoner, visit <http://pellet.owldl.com/>.

Running the Pellet Reasoner in Normal Client Mode

To run the reasoner in normal client mode (from your desktop), follow these steps:

1. Open Protégé.
2. Using both the Collaborative Protégé tab and e-mail, notify editors that the reasoner will be run and that they should not be logged in while it is running.
3. Follow these steps in order, referring to the pertinent sections:
 - a. Configure Protégé to display the OWL classes tab.

See *Enabling the OWL Classes Tab* on page 41.

- b. Start the reasoner.
See [Starting the Reasoner](#) on page 42.
- c. Review the results, accepting or rejecting retreeing suggestions according to established guidelines.
See [Asserting \(Accepting or Rejecting\) Current Classification Results](#) on page 42.
- d. Using the previous classification report reviewed and discussed by the modelers, accept or reject items as determined.
See [Asserting \(Accepting or Rejecting\) Results from the Previous Classification](#) on page 44.
- e. Save the current classification results as an external text-delimited file, then open the file in Microsoft Excel and save it as an Excel workbook for circulation among the team.
See [Saving the Current Classification Results as a Report](#) on page 45.
- f. Exit the Protégé project without saving it.

Pre-Classification Procedures

Before you run the reasoner, complete the following procedures:

- [Notifying Protégé Users of Database Shutdown](#) on this page
- [Starting the Remote Protégé Client](#) on page 39

Notifying Protégé Users of Database Shutdown

Before running the reasoner, let all users know that the database will be shut down. This gives them a chance to save their work and log off.

To notify users of a database shutdown, follow these steps:

1. Using e-mail or chat, announce that the Protégé database will be shut down.
2. Log out of the Protégé client on your own machine.

Starting the Remote Protégé Client

Note: In the following instructions, the current NCI Protégé directory is represented as `Protege.Server-x.y.z`. When typing commands, substitute the directory name with the current Protégé version number (**Example:** `Protégé.Server-1.2.3`).

To start the Protégé client on the remote server, follow these steps:

1. Log in to the Protégé host computer:
 - a. Open Reflection X.
 - b. Select the **direct** connection method.
 - c. Enter the host machine name.

Tip: If you have previously entered the host machine name, you can select it from the drop-down list.

- d. Enter your user name and password.

2. Open a Terminal window.
3. Shut down the Protégé server:
 - a. For **BiomedGT**, type the following command:

```
cd /usr/local/protége/Protege.Server-x.y.z
```

or
For the **NCI Thesaurus**, type the following command:

```
cd /usr/local/protégeqa/NCIThesaurusGroup/Protege.Server-x.y.z
```
 - b. Press ENTER.
 - c. For **BiomedGT**, type the following command:

```
./shutdown_protege_server.sh localhost:2633
```

or
For the **NCI Thesaurus**, type the following command:

```
./shutdown_protege_server.sh localhost:2822
```
 - d. Press ENTER.
4. Start the remote Protégé client:
 - a. Open a second Terminal window.
 - b. For **BiomedGT**, type the following command:

```
cd /usr/local/protége/Protege.Client-x.y.z
```

or
For the **NCI Thesaurus**, type the following command:

```
cd /usr/local/protégeqa/NCIThesaurusGroup/Protege.Server-x.y.z
```
 - c. Press ENTER.
 - d. For **BiomedGT**, type the following command:

```
./run_protege_BGT.sh
```

or
For the **NCI Thesaurus**, type the following command:

```
./run_protege_NCIT.sh
```
 - e. Press ENTER.
 - f. When prompted to select a project, select **BiomedGT** or **Thesaurus**.

Classification Procedures

The following topics explain how to perform the procedures that correspond to each stage of the classification process:

- [Enabling the OWL Classes Tab](#) on this page
- [Starting the Reasoner](#) on page 42
- [Asserting \(Accepting or Rejecting\) Current Classification Results](#) on page 42
- [Saving the Current Classification Results as a Report](#) on page 45

Enabling the OWL Classes Tab

After running the reasoner, you can view the results using the OWL Classes tab. This tab is not enabled by default.

To enable the OWL Classes tab, follow these steps:

1. Select the following menu command: **Project > Configure...**

The Configure File window opens.

2. In the **Visible** column, check the box to the left of the **OwlClassesTab** widget.
3. Click **OK** to close the Configure File window.

The OWL Classes tab now appears in the main Protégé window. The tab is positioned on the far right ([Figure 3.1](#)).

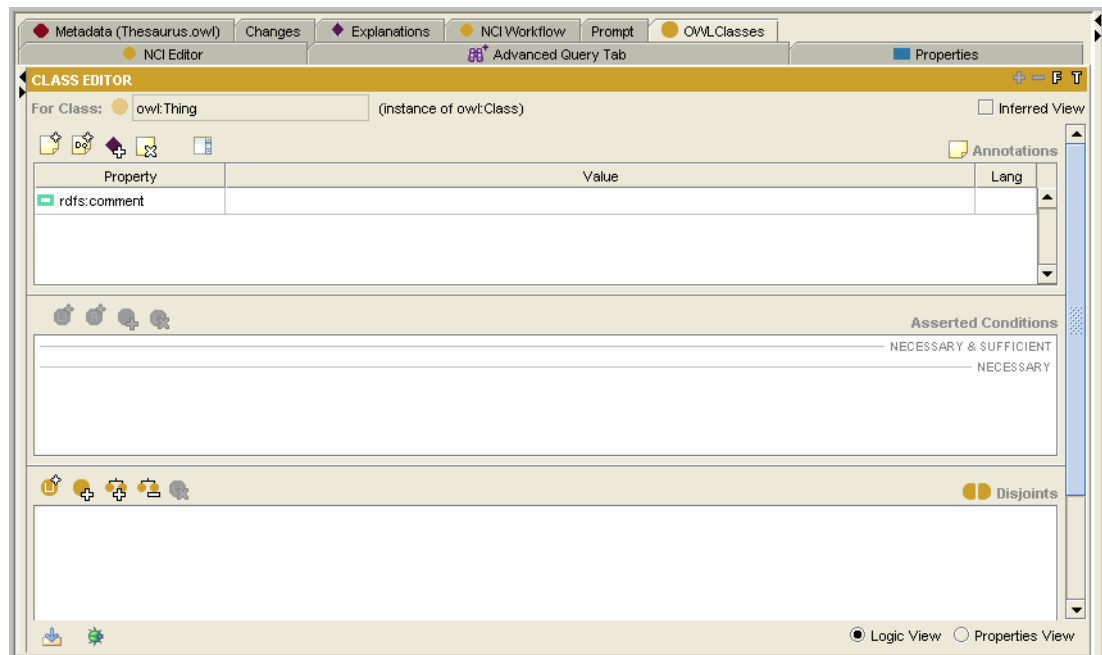


Figure 3.1 Owl Classes tab

Starting the Reasoner

Prerequisite procedure: [Enabling the OWL Classes Tab](#) on page 41

To start the reasoner from the OWL Classes tab, select the following menu command:

OWL > Classify taxonomy

A status window opens. This window monitors the status of the classification process. When the Task Complete message shows ([Figure 3.2](#)), click **OK** to close the window.

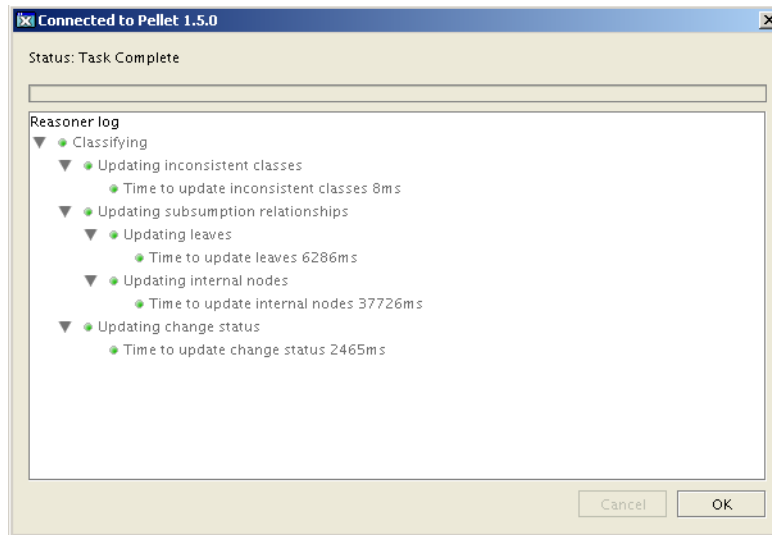


Figure 3.2 Window showing status during classification

Asserting (Accepting or Rejecting) Current Classification Results


After the reasoner runs, click the OWL Classes tab and examine the layout. Note the following aspects of the tab panels (illustrated in [Figure 3.3](#) on page 43):

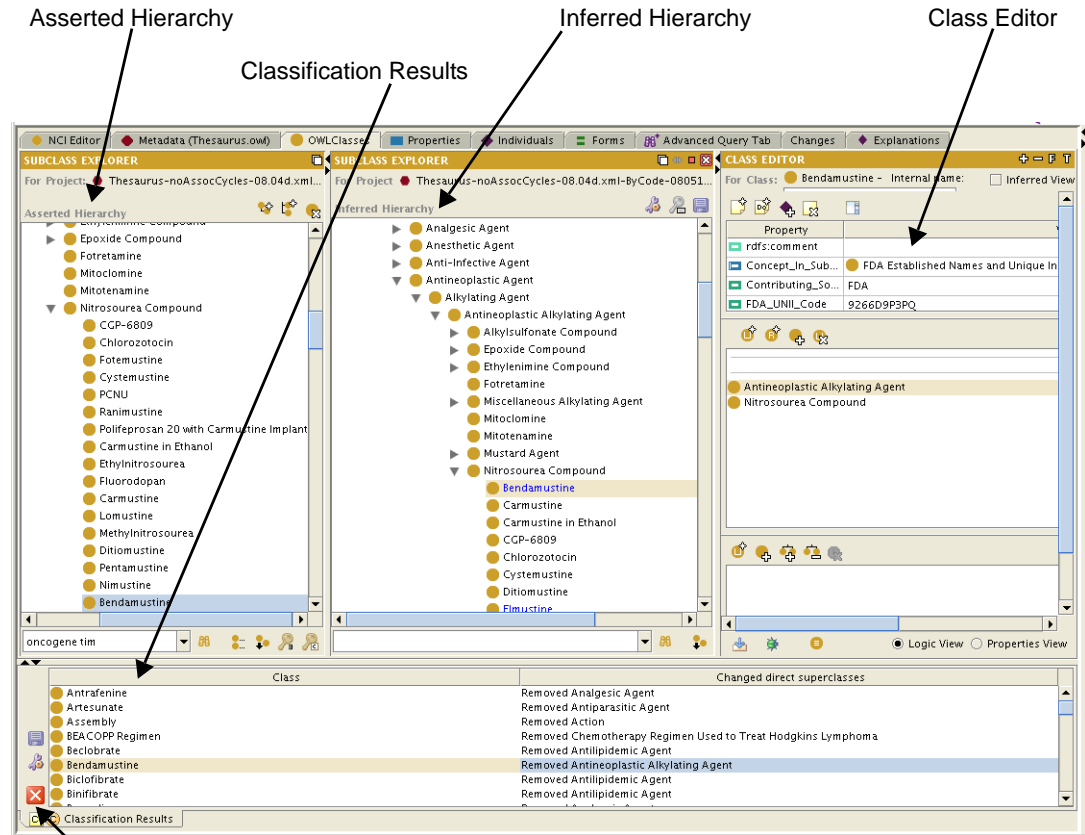
- The two Subclass Explorer panels each represent a different view of the class hierarchy:
 - The left panel displays the *Asserted Hierarchy*. This is the hierarchy resulting from current modeling.
 - The middle panel displays the *Inferred Hierarchy*. This is the hierarchy as established by the reasoner during classification.
- The Class Editor panel is a fully functional editor, enabling you to view and edit properties for a selected class or subclass.
- The Classification Results panel at the bottom of the tab shows class names on the left and changed direct subclasses on the right.

To read the following scenarios for accepting and rejecting classification results, click the right-pointing arrow to the right of each scenario name.

Scenario 1: Accepting a Suggestion

In *Figure 3.3*, the *Bendamustine* class is selected in the class results. The reasoner has suggested that the redundant parent class *Antineoplastic Alkylating Agent* (shown in the Class Editor panel) be removed.

To accept the selected change, click the **Assert Selected Change** button  on the left of the Classification Results panel (*Figure 3.3*). The button icon resembles a wrench.



Buttons, in order: Save Results, Assert Selected Change, and Close Results Panel

Figure 3.3 OWL Classes tab with classification results

Scenario 2: Rejecting a Suggestion

In [Figure 3.4](#), the reasoner has suggested that Complement Components 3 and 5 be re-treed as children of Complement Component 4. This suggestion is probably not valid and should alert the modeler that either (1) the DL model is incorrect and needs to be reconsidered, or (2) the DL model has been incorrectly applied and the modeler needs to modify one or more of the concepts in question.

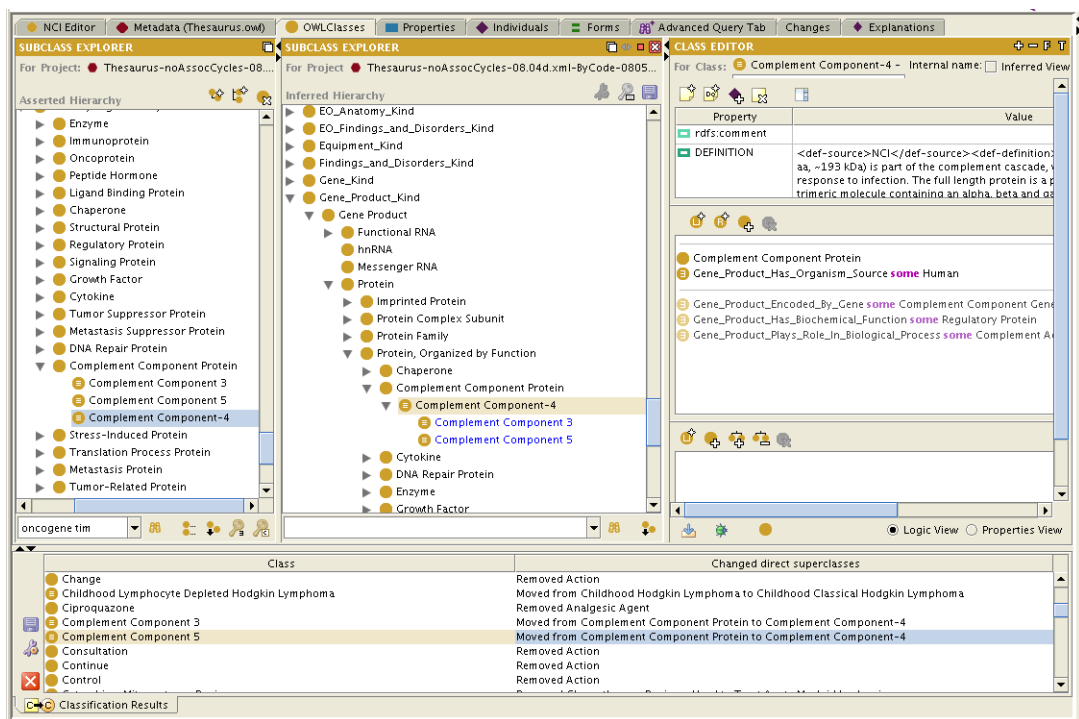


Figure 3.4 Classification Results panel with suggestions for Components 3 and 5

Note: The Classification Results panel does not include a button or any other mechanism for rejecting a suggestion. You reject suggestions by not explicitly accepting them. For the current scenario, ignore the suggestion and make a note to follow up on the three sibling classes.

Asserting (Accepting or Rejecting) Results from the Previous Classification

After asserting results from the *current* classification, review the report from the previous classification. Assert any outstanding items that have been resolved through discussion with the modelers.

For more information about creating a report after each classification session, see [Saving the Current Classification Results as a Report](#) on page 45.

Saving the Current Classification Results as a Report

Once you have identified any modeling issues that you need to discuss with the team, you can create a report for circulation.

To create a report showing the current classification results, follow these steps:

1. In the Classification Results (bottom) panel, click the **Save Classification Results** button on the left (*Figure 3.5*).

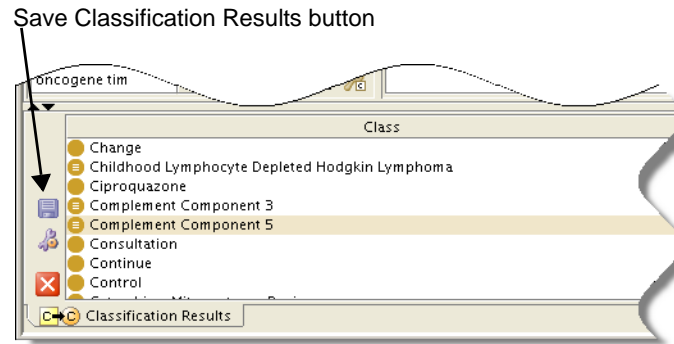


Figure 3.5 Save Classification Results button

2. Save the file to a specified directory, noting where you saved it.
This step creates a text file that is delimited by equal signs (=).
3. Open Microsoft Excel.
4. Select **File > Open**, then open the file that you saved in step 2.
The Text Import Wizard launches.
5. To set up the workbook, use the following Wizard settings:
 - a. On the Step 1 of 3 screen, select **Delimited** as the file type, then click **Next**.
 - b. On the Step 2 of 3 screen,
 - Clear the **Tab** check box under the **Delimiters** section.
 - Select the **Other** check box.
 - Type an equal sign (=) in the box following the word *Other*.
 - Click **Next**.
 - c. On the Step 3 of 3 screen, click **Finish**.
The delimited data now appear in workbook columns.
6. Using **File > Save As**, save the file in Microsoft Excel (.xls) format.
7. Close Excel.

Exiting Protégé Without Saving the Project

When you have completed all of the classification procedures (enabling Prompt, running the comparison, accepting and rejecting changes, and saving the report), you are ready to exit Protégé.

Note: When you exit, Protégé prompts you to save the current project. *Be sure to exit without saving.* Remember that, as part of the classification process, you configured the open Protégé project to display the OWL Classes tab. This tab is not a part of the standard configuration for all users, so exiting without saving will prevent the tab from showing when you restart the database.

Post-Classification Procedures

To complete the classification process, follow these steps:

1. Start the Protégé server:

a. For **BiomedGT**, type the following command:

```
cd /usr/local/protege/Protege.Server-x.y.z
```

or

For the **NCI Thesaurus**, type the following command:

```
cd /usr/local/protegeqa/NCIThesaurusGroup/Protege.Server-  
x.y.z
```

b. Press ENTER.

c. Type the following command:

```
./run_protege_server.sh
```

d. Press ENTER.

2. Verify that the Protégé server is running:

a. Type the following command:

```
tail -f console.txt
```

b. Press ENTER.

Note: You can also verify that the Protégé server is running by re-opening the Protégé client and then closing it.

3. (Optional) Restart the Explanation server to make sure that it is synchronized with the database:

a. For **BiomedGT**, type the following command:

```
cd /usr/local/protege/Explanation.Server-x.y.z
```

or

For the **NCI Thesaurus**, type the following command:

```
cd /usr/local/protegeqa/NCIThesaurusGroup/Explanation.Server-  
x.y.z
```

b. Press ENTER.

c. Type the following command:

```
./restart_explanation_server.sh
```

d. Press ENTER.

4. Log off of the host computer using Reflection X.

INDEX

A

accepting

- classification results 42 – 44
- comparison results 26

accepts and journal files 33

annotation ontology, about 30

asserting classification results 42 – 44

assigning workflow item 12

B

baseline version

- changing number 29
- naming convention 28

BiomedGT overview 5 – 7

C

change ontology

- See* annotation ontology

changing baseline version number 29

classification process

- asserting results 42 – 44
- overview 37 – 38

comparing baseline files

- See* database comparison process

concept history, saving 27

contacting support 4

D

database comparison process

- about 13
- accepting and rejecting results 26
- deleting extraneous files 33
- interpreting results 19
- starting 18

E

Explanation server

- shutting down 15
- starting and verifying 34

exporting project file 31

I

importing packages from wiki 12

J

journal and accepts files 33

N

naming convention for baseline version 28

NCI Workflow tab 10

notifying users of shutdown 15, 39

O

Owl Classes tab, enabling 41

P

Pellet reasoner

- about 38
- starting 42

project file, exporting 31

Prompt plug-in, about 15

Prompt tab

- disabling 27
- enabling 17

Protégé server

- shutting down 15
- starting and verifying 34

R

reasoner (Pellet)

- about 38
- starting 42

reassigning workflow item 12

rejecting

- classification results 42 – 44
- comparison results 26

remote client, starting 15, 39

results (classification), accepting and
rejecting 42 – 44
results (comparison)
accepting and rejecting 26
interpreting 19

S

Semantic MediaWiki
See BiomedGT overview
starting remote client 15, 39
support, contacting 4

T

Table View subtab 20
Tree View subtab 22

U

users, notifying of shutdown 15, 39
Users subtab 24

W

workflow overview
BiomedGT to Protégé 7
internal NCI process 8 – 11
workflow procedures
assigning item 12
importing packages from wiki 12