

Protege Annotation

Protege_Manual_Annotation_Guidelines.doc

Manual Annotation Guidelines

Installation

Annotation is performed using Protégé and the Knowtator plugin.
To get this software installed follow the steps below:

1. visit <http://protege.cim3.net/download/old-releases/3.3.1/basic/> and install **Protégé 3.3**
2. visit <http://sourceforge.net/projects/knowtator/files/Knowtator/Knowtator%201.7/> and download **knowtator-1.7.zip**
3. visit <http://knowtator.sourceforge.net/install.shtml> and follow **steps 1-5** of the first section to perform the installation.
4. restart protégé.
5. Create a new project
6. once protégé starts, select **knowtator>configure** from the top menu
 - a. in the section labeled 'default annotator' click the diamond icon with the plus sign and select yourself. If you do not exist in that list, click the diamond icon with the radiating lines and create yourself as a new 'human annotator'.
7. If you encounter memory issues, you will need to increase memory usage for protégé. To do this you will need to create a Protégé.lax file to be stored in the same directory as your protégé project:
 - a. create your own file:
 - i. open protégé
 - ii. click file > preferences
 - iii. click the 'property files' tab
 - iv. click the protégé.lax tab
 - v. locate the row labeled 'lax.nl.java.option.java.heap.size.initial' and change the number in the right column to '100000000'. (100mb)
 - vi. locate the row labeled 'lax.nl.java.option.java.heap.size.max' and change the number in the right column to '2000000000'. (2gb)
 - vii. click the floppy disk icon in the upper right corner and save this file as Protege.lax
 - viii. additional details in references section below.

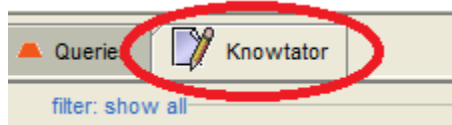
Annotation Types

You will need to create annotation classes in your protégé project. Annotations are broken up into 8 main categories:

1. patient
2. doctor
3. date
4. id
5. location
6. hospital
7. phone
8. age

Annotation

1. open protégé and click on the 'knowtator' tab

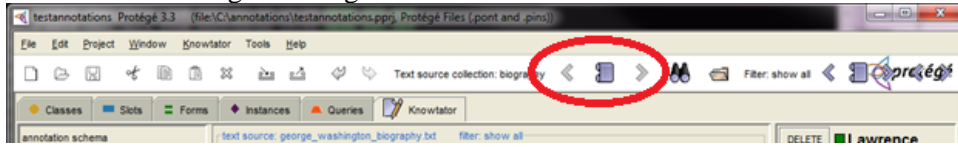


- a.
2. Increase font size by clicking on "Window > Increase Font Size" at the top of the screen.
3. open the directory containing the input files
 - a. in protégé, in the upper right corner, click the folder icon.



- b.
- c. navigate to the input directory

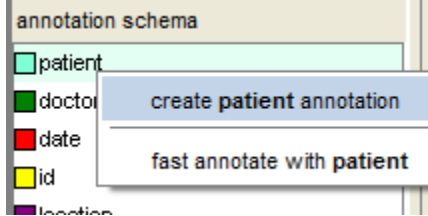
4. navigate between files using left & right arrows:



- a.

5. annotate

- a. highlight a span of text with your mouse
- b. left-click on the appropriate phi type on the left under 'annotation schema' and choose 'create annotation'

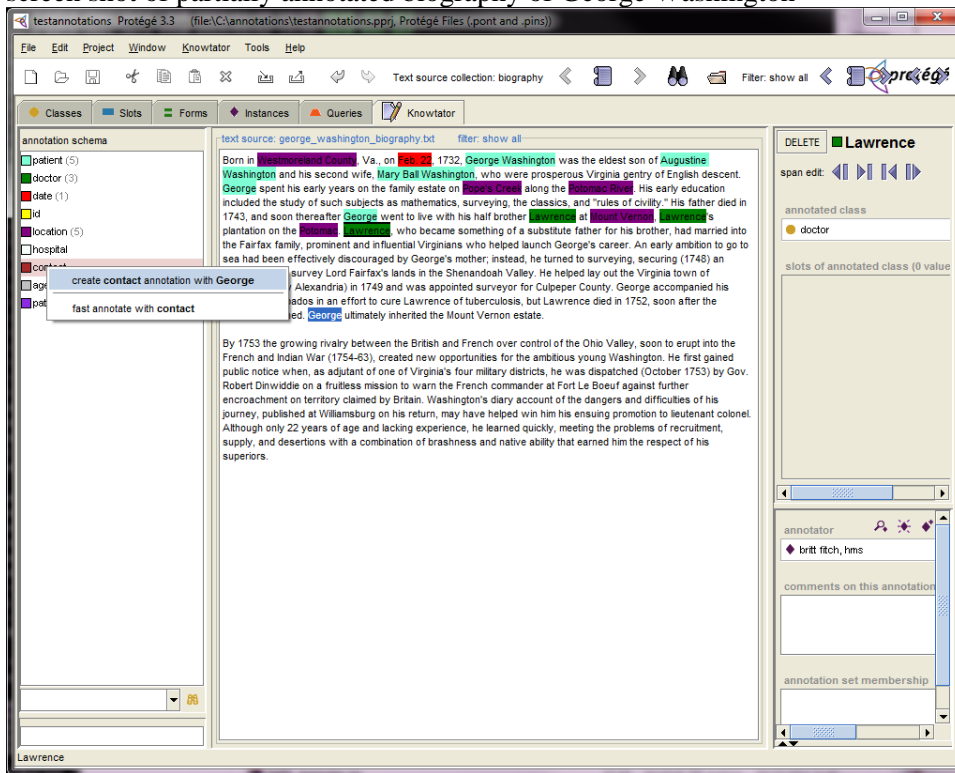


- c.

6. save

- a. be sure to save the annotations periodically.

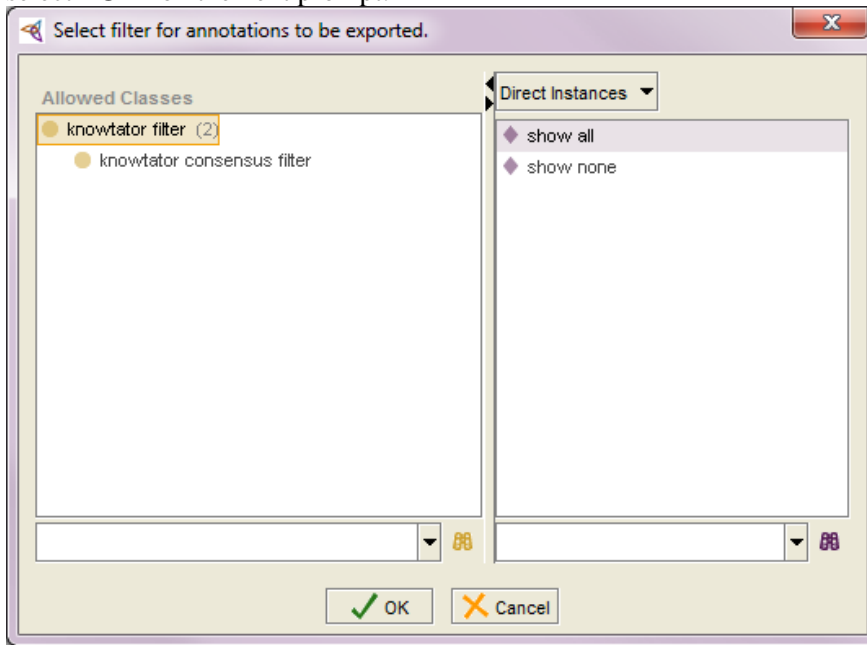
7. screen shot of partially annotated biography of George Washington



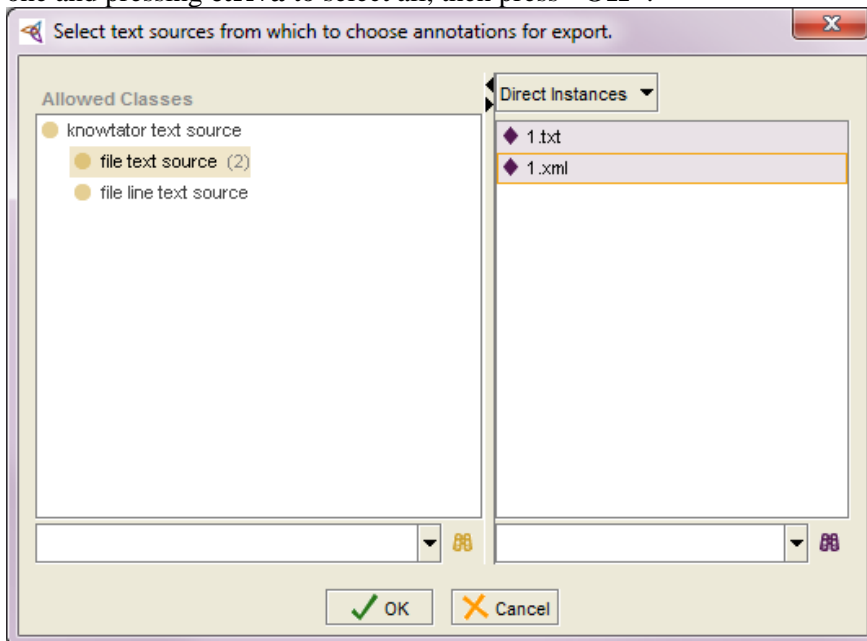
Extracting Annotations

1. From protégé, click **knowtator>export annotations to xml** and you will be prompted for an output directory.

2. select '**OK**' on the next prompt:



3. on the next prompt, select '**file text source**' on the left and **select all of the cases** on the right by selecting one and pressing **ctrl+a** to select all, then press '**OK**'.



Annotation Import

This section is useful if you need to modify previous annotation files.

1. At the top of the screen click 'knowtator>import annotations from xml'
2. When prompted for the location of the import files, navigate to the directory containing your xml annotation files and select all files in that window by clicking on one of them and pressing '**ctrl +a**'.
3. You should now be able to navigate between cases as described above. please do so and verify:
 - a. The annotations have been imported properly, i.e. you can see them and they do not all appear in a dark grey color. The left navigation should show counts of how many instances of each type of PHI are currently recorded for this case.
 - b. If there are any errors in importing, **DO NOT** save the project, just close the project and retry importing the xml files.

References

This section is for reference only, it is not required to get the minimal installation up and running.

1. quickstart guide for setting up a new project: <http://knowtator.sourceforge.net/quickstart.shtml>
2. general documentation: <http://knowtator.sourceforge.net/docs.shtml>
3. data management: <http://knowtator.sourceforge.net/data-management.shtml#merge>
4. filters: <http://knowtator.sourceforge.net/filters.shtml>
5. memory usage: http://protegewiki.stanford.edu/wiki/Setting_Heap_Size