The labeling tool can help us label protein and cell names, and also specify some properties of the labels(I also call them annotations). For example, your confidence for labeling this as a protein name (1.0 is the highest confidence and is set to default) and why you label it as a protein name (based on your knowledge or any resources you refer to) would be helpful annotations.

Here I'll show you how to use the labeling tool to:
Add labels
Add annotations
Remove unwanted labels
Label the identical names within one caption
Save results Please remember to save before you close the main window
View the original papers
Use shortcut keys
by going through some examples.

Startup

First download the minorthird.jar, your data directory and urlmap.txt from goblin, start the labeling tool by running

java –cp PATH_to_minorthird.jar edu.cmu.minorthird.text.gui.TextBaseEditor DATA_DIRECTORY LabelsFlName urlmap.txt

where
PATH_to_minorthird.jar is the relative path to the minorthird.jar
DATA_DIRECTORY is where your documents are stored,
LabelsFlName is where you would like to save your labels (minorthird will create a LabelsFlName if it does not exist),
and **urlmap.txt** is a file containing information about the original papers.

For example, I'll put minorthird.jar, my data directory 'zhenzhen' and urlmap.txt at /home/zkou/data_label/, 'cd /home/zkou/data_label/' and run

java –cp minorthird.jar edu.cmu.minorthird.text.gui.TextBaseEditor zhenzhen/ zkou.labels urlmap.txt

to start labeling.

A window that looks like Figure1 will appear:

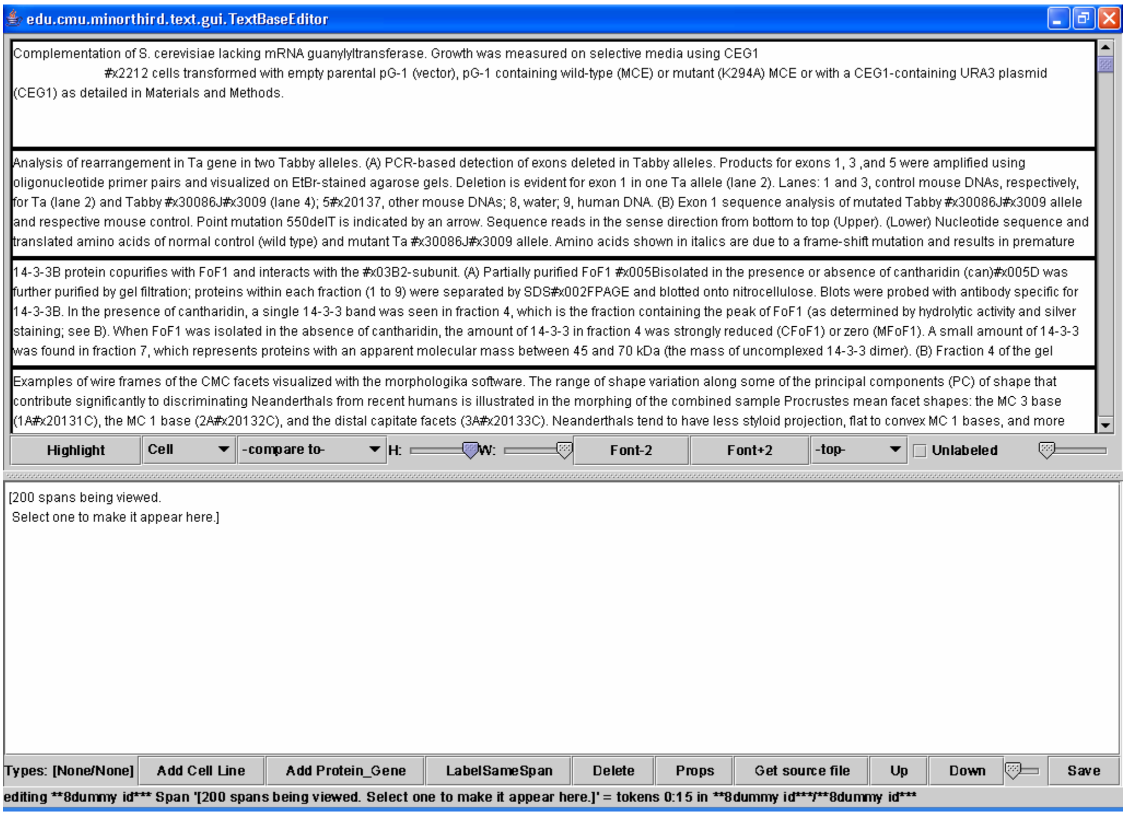

Figure 1

Please select the label ‘Cell’ from the pull-down menu 1 in Figure 2 and ‘Prot’ from the pull-down menu 2.

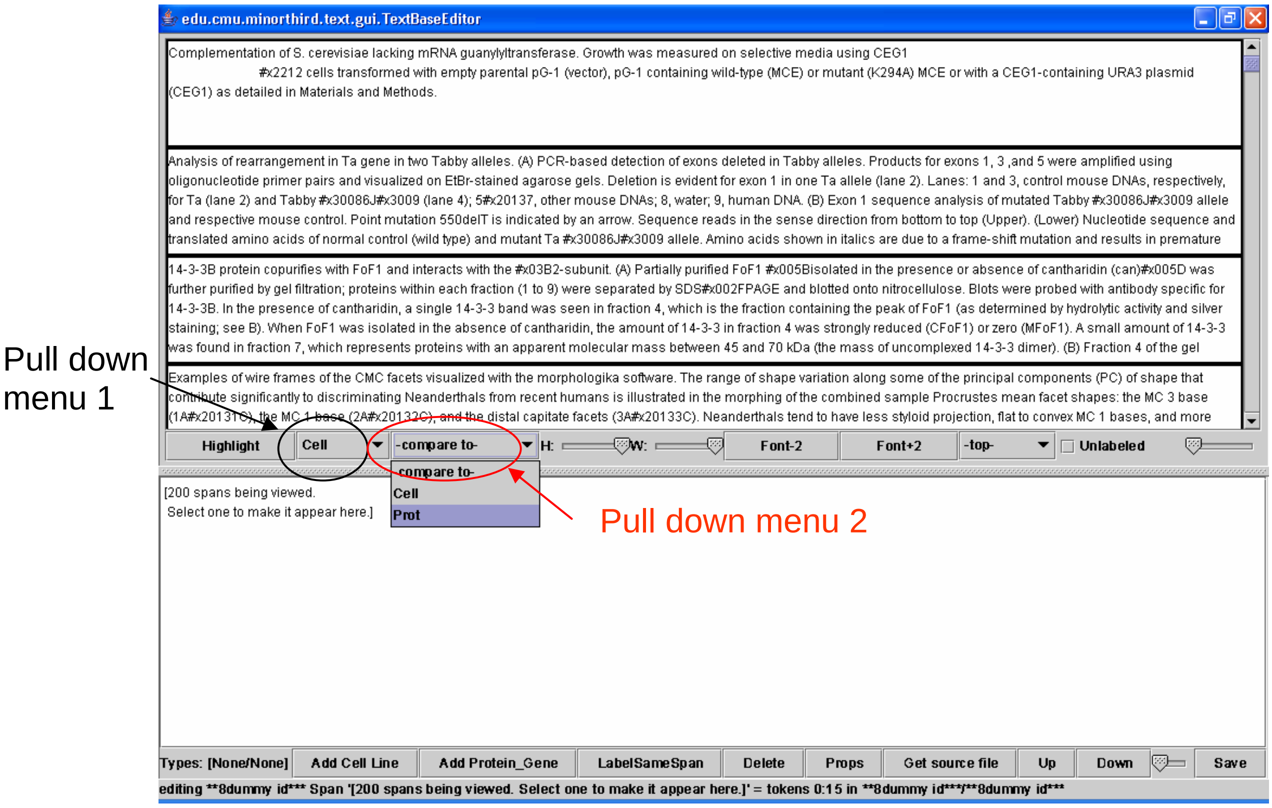

Figure 2

Then you'll get a working area like Figure 3. You can adjust the sliding bars to adjust the height and width of the tope panel where the documents are listed.

Figure 3

To select a document to label, click on it in the top panel and the text from that document will appear in the bottom panel, as shown in Figure 4.

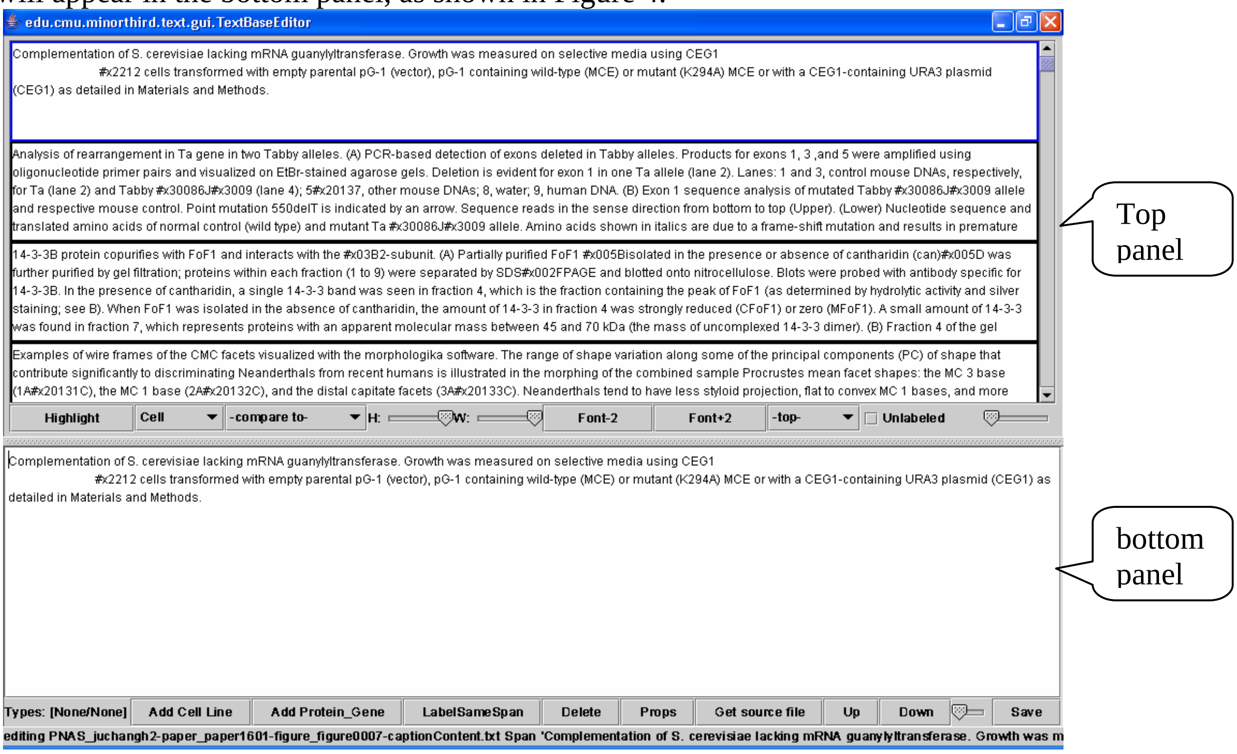

Figure 4

Add a label

To label the currently selected document, you **have to** work in the bottom panel. Select the words you want to label with the mouse, and then click the ‘Add Cell Line’ or ‘Add Protein_Gene’ button, as shown in Figure 5.

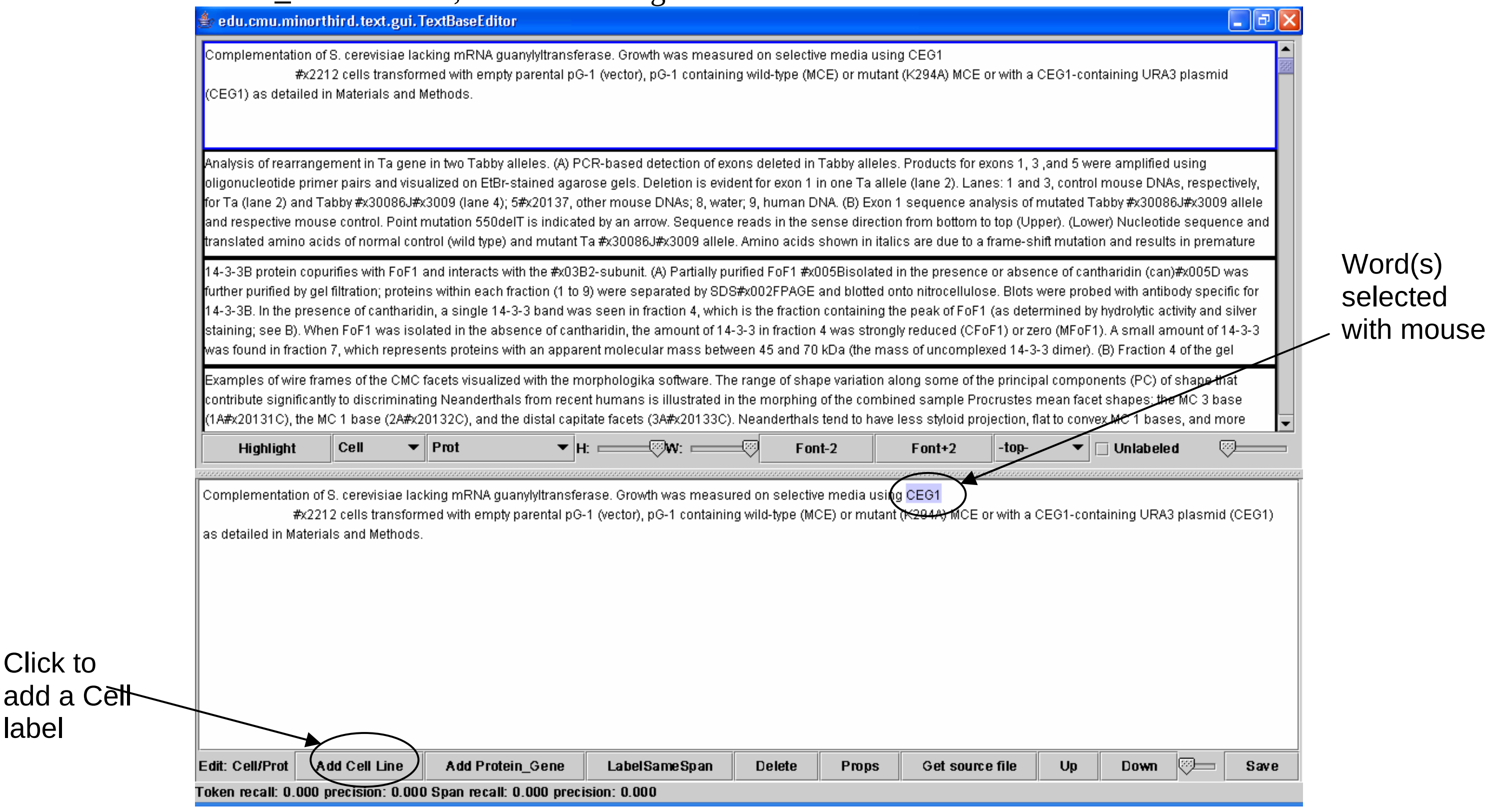

Figure 5

Once successfully labeled, the names will be highlighted as yellow (for category specified in pull down menu 1, here it’s ‘Cell’) or blue (for category specified in pull down menu 2, here it’s ‘Prot’). Figure 6 shows a successfully labeled ‘CED1’ as ‘Cell’ .

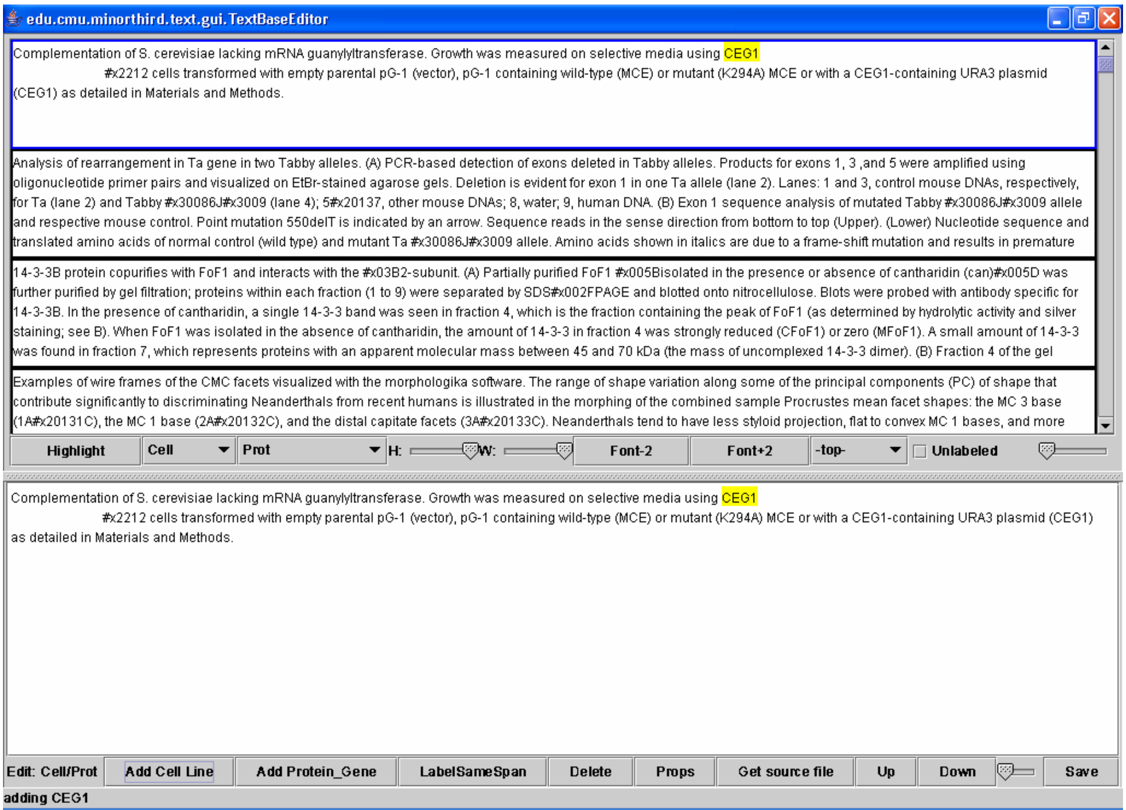

Figure 6

Figure 7 shows another example, with a cell name and a protein name successfully labeled.

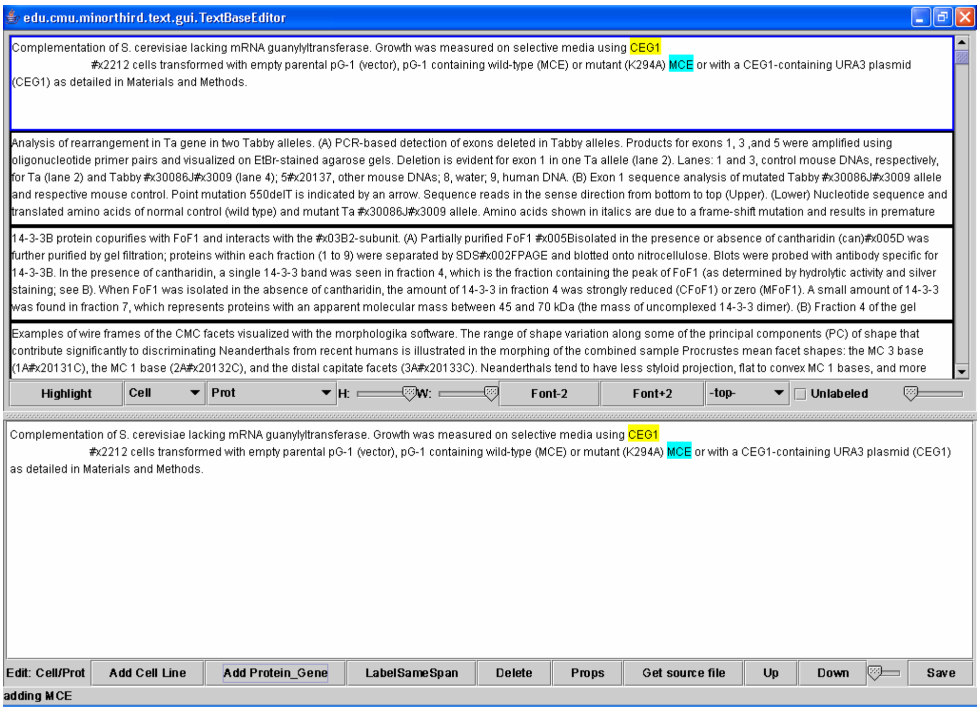

Figure 7

Put some annotations

Our labeling tool is able to let you input some annotations when you label the terms. The default annotations generated are:

Method of labeling: manually
Confidence of labeling this term: 1.0
Reference: I know it is a protein (or cell) name
Conjuctive: false
Family: false
Complex: false
gene/protein: protein
comments: additional things you want to add

If you refer to any sources, or feel not that confident to label one term, it would be helpful to record them as annotations. To view and edit the annotations, select the term you've labeled, and click the 'Props' button (as shown in figure 8), a window like Figure 9 will pop up.

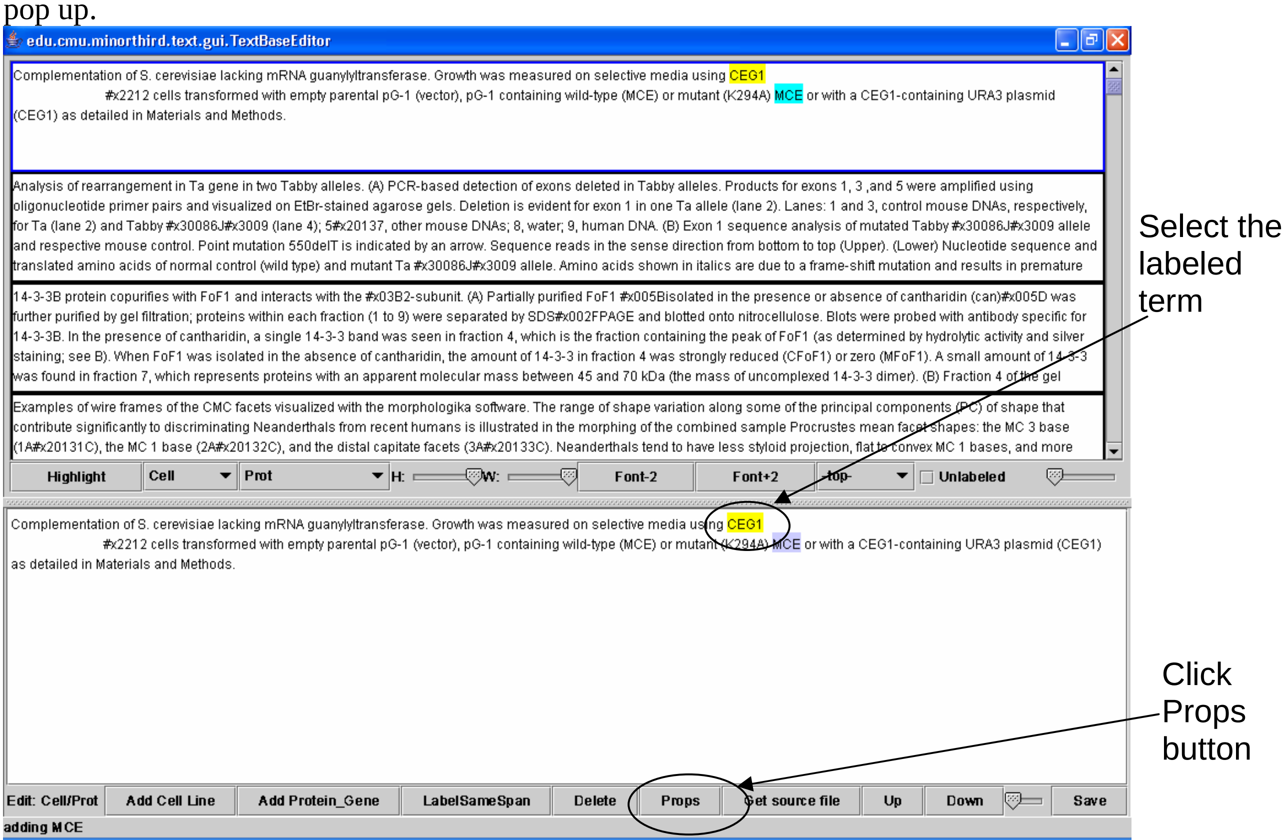

Figure 8

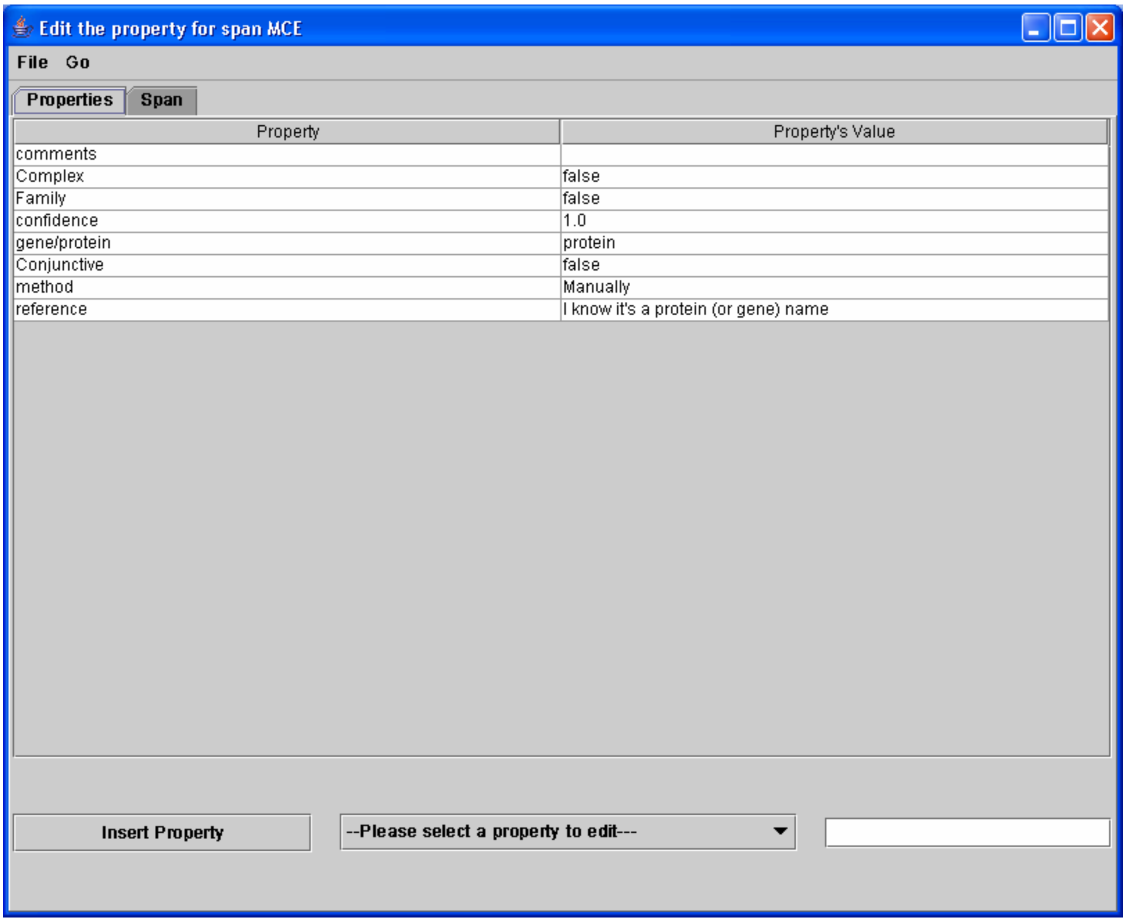

Figure 9

The window will display the default annotations. You can edit them by select the item you want to modify from the pull down menus and input your annotation into the text box, then click ‘Insert Property’(shown through Figure 10~11).

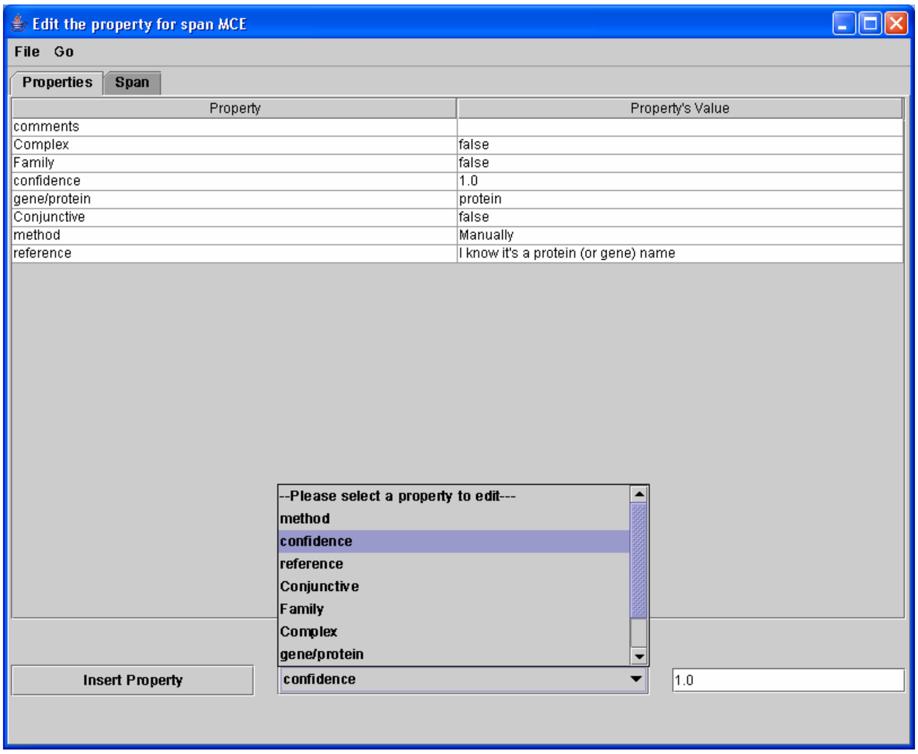

Figure 9

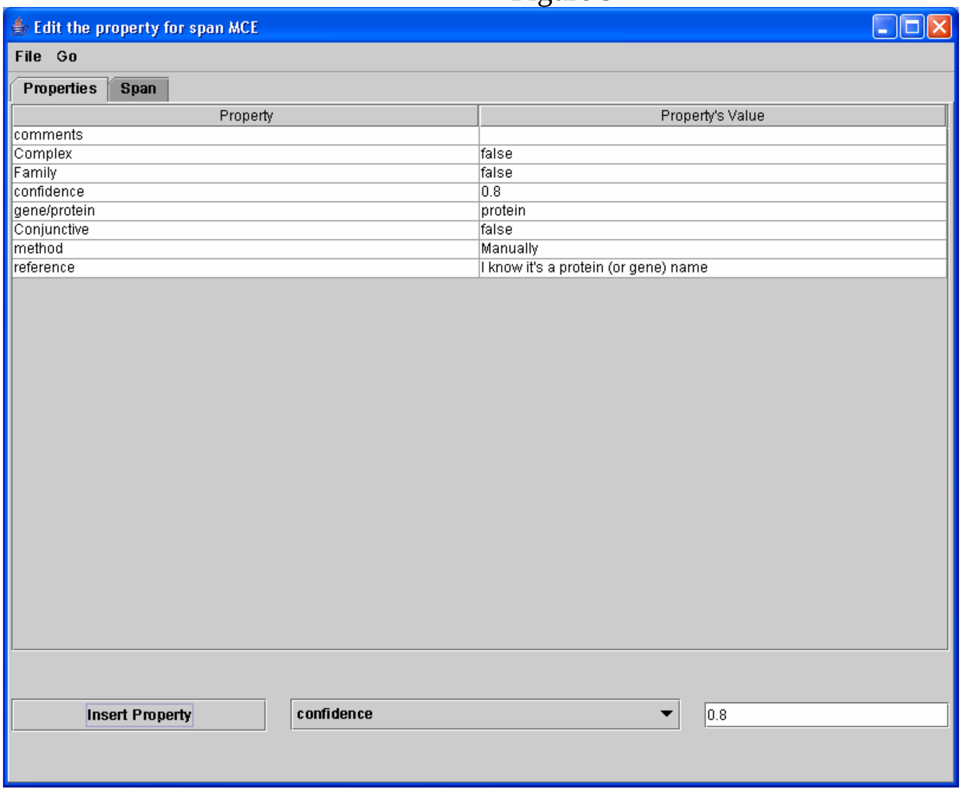

Figure 10

Delete a label

If you mis-labeled something, select the mis-labeled word(s) with mouse, and click the 'Delete' button.

Label the identical names within one caption

One name might be mentioned several times in a caption. The labeler provides a convenient way of labeling the remaining identical names after you've labeled one of them.

For example, if I want to label the remaining 'CEG1' as 'Cell', select the labeled one and click on 'LabelSameSpan' button, as shown in Figure 11.

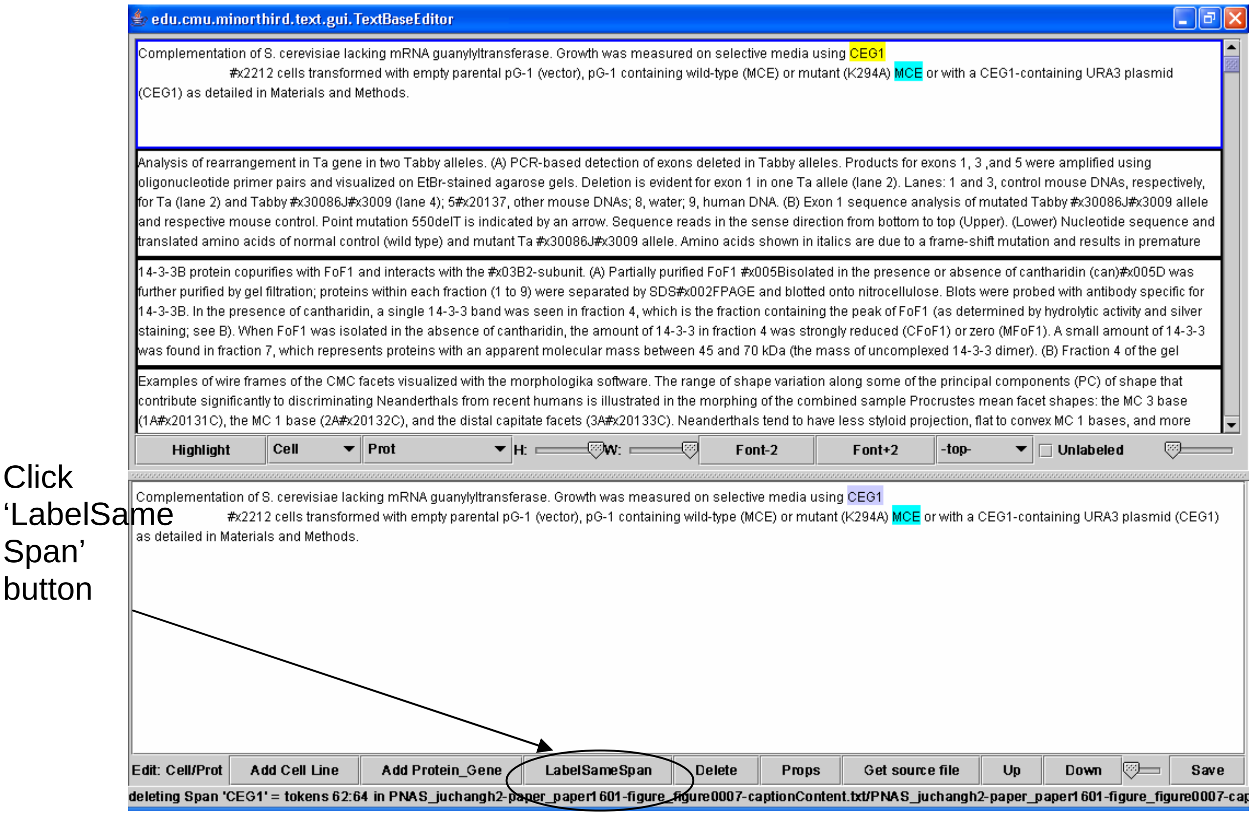

Figure 11

We'll get the remaining 'CEG1' labeled, as shown in Figure 12.

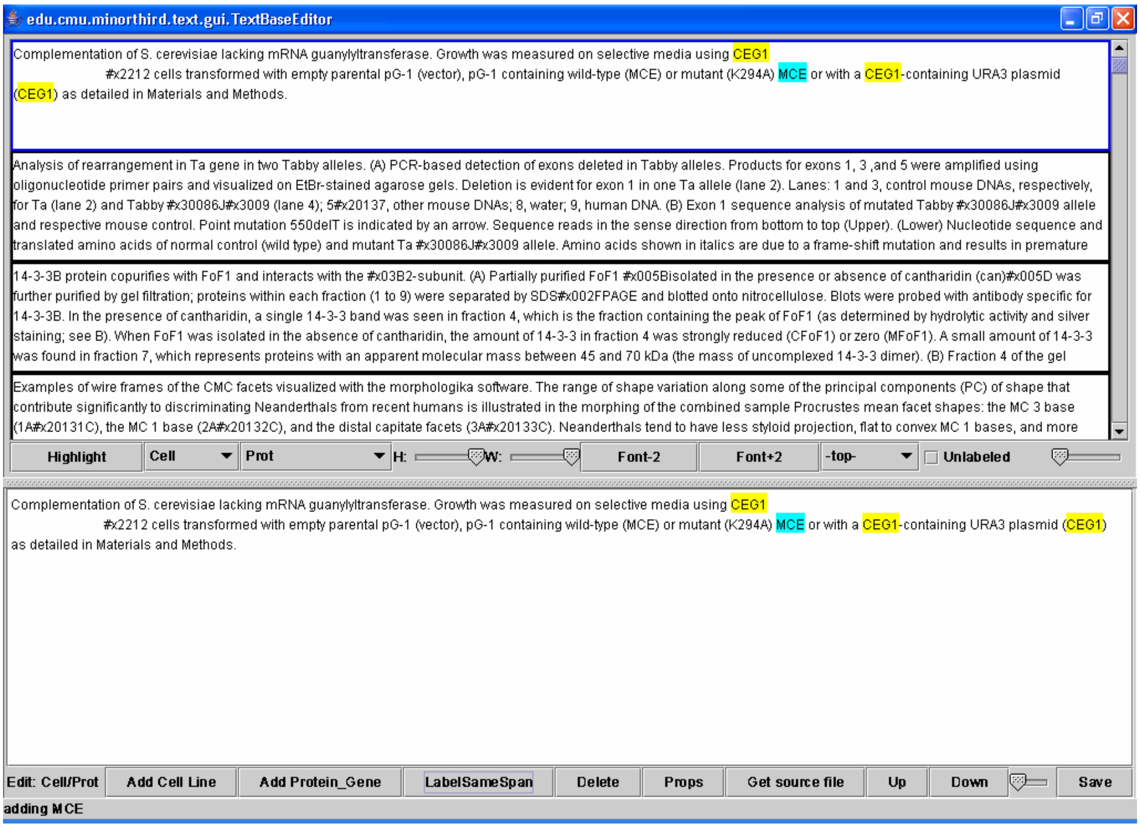

Figure 12

If you've modified the annotations of 'CEG1', they will be copied to the other 'CEG1' when you use 'LabelSameSpan' to label them.

Save your result

You can save your result at any time and come back again. Click the 'Save' button to save your labeling effort.

Please remember to save before you close the main window. There won't be warning message pop-out to remind you to save your results when you close the main window (sorry I haven't figured out how to register a listener for this).

View the original paper

If you want to look at the original paper the caption was extracted from, click the ‘Get source file’ button and you’ll see a dialog containing the url to the paper, as shown in Figure 13. You can copy the url from the text field and open it in your browser.

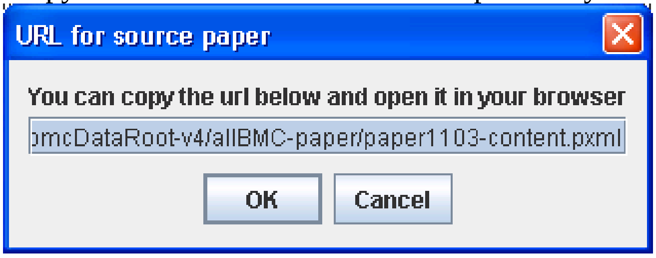

Figure 13

Shortcut keys

To speed up the labeling, some shortcut keys are defined:

To add a Cell, you can use ‘c’ or ‘C’
To add a Protein_gene, you can use ‘p’ or ‘P’
To modify the properties, you can use ‘ctrl+p’
To delete a label, you can use ‘Del’
To save, you can use ‘ctrl+s’
You can also use ‘left arrow’ and ‘right arrow’ to move around the entities you’ve labeled.

Additional information

1. **The ‘UP’ and ‘Down’ button in the main window:** this is designed to help you move from one document to the adjacent document in the top panel. Yet you can always select a document to label by clicking on it in the top panel.
2. The term labeled **both as ‘label specified in pull down menu 1’ and ‘label specified in pull down menu 2’** will be highlighted as green. This is unlikely to happen in our case, when labeling protein names and cell names. This will be helpful to cross-validate things labeled via difference sources.