15-400 Milestone Report V

Sonya Anopa

April 15, 2016

0.1 Milestone Progress

Since the last milestone I was working on adding relations and roles, realized by 'acl' and 'nmod' relations, which is exactly the progress planned for by this stage with my last progress report. I also worked a bit on improving the word sense disambiguation problem encountered earlier and on moving the script over to a server to be able to run it more freely.

First I added relations, which came from 'acl' dependencies in the original gloss. For example, the first part of the definition for 'cisterna' is 'a sac or cavity containing fluid'. We can learn from it that a cisterna is not only a sac or cavity, but also that it contains fluid. That information can be extracted as follows: the head of the dependency tree in here is 'sac' and 'containing' is related to 'sac' with the 'acl' relation. Also, 'fluid' depends on 'containing'. Extracting those words, we create a relation "contains" and add a new statement in Scone that means "cisterna contains fluid".

Next, 'nmod' relations are converted into roles in Scone. That is, if there is a group of words such as "population of India", we know that 'population' is a role that something can perform. So, from the original glosses I extract such relationships, which are defined by 'nmod' dependencies and 'case' dependencies (e.g., in "population of India", 'of' is the 'case' dependency). Then I create roles out of the heads of the dependency trees ("India" above) and a new statement of the form "x is a population of India" (in most cases, we are dealing with types rather than instances, so in general the "x-is-a-y-of-z" will be used, but "x-is-the-y-of-z" is possible as well) where "x" is the synset being defined.

Finally, because running both parsers takes a very long time on my computer, I started trying to move everything to a Scone server, but that has not been completely finished due to permissions and script dependencies that require those permissions.

0.2 Looking Ahead

Once I get everything on the server, I'll be generalizing the script to accept any definition and output the relations I have been adding until now. One definite problem to take care of is the duplication of effort. For example, sometimes the gloss defines a synset as "x or y", which means that all the extractions having to do with x should also work for y and any number of such relations. Currently, that would require to copy-paste a lot of code, which I for that reason have not done; instead, I'll refactor the current attempts to make such additions smooth. Also, I'll make sure to allow non-WordNet definitions to be processed.

0.3 Revisions to Future Milestones

There are no new revisions to future milestones - the ones described in the previous progress report apply.

0.4 Resources Needed

I have all the resources I need so far.