15-400 Milestone Report IV

Sonya Anopa

March 18, 2016

0.1 Milestone Progress

I have been exploring WordNet through both NLTK and parsing its definitions with the Stanford dependency parser. I began by trying to parse the definition of every noun and extract the root of the resultant dependency tree as the superclass or is-a link that I would generate between that concept and the superclass. That was in general successful; however, as in another class I am using NLTK (Natural Language Processing Toolkit, a Python kit), which contains some WordNet features, I decided to explore that first. It turned out that NLTK provided more exact information - I could search for links between concepts with an API call. Those links also included other useful relations such as meronymy/holonymy and additional possible words used when referring to a concept. Therefore, I used NLTK to DFS through the WordNet graph to ensure adding concepts to Scone only when their parents have been added, including the holonymy and alternate naming information. As I realized later, I could have parsed the same information from the raw WordNet data.

Next, I returned to parsing the definitions to retrieve useful information. I have looked further into some relations I identified earlier and some new ones, although I have not added them into Scone yet. The ones that seem promising are the 'amod', 'nmod' (with 'case'), 'conj', 'acl', and 'appos' relations. The 'amod' and 'conj' relations relate adjectives (and adjective groups, e.g. "small and loud") to the concept; however, I'm not entirely sure as to how to translate those into Scone, so I will add them upon consulting with Dr. Fahlman. The 'nmod' relation describes a possessive relation, for example that the intermaxillary suture is a suture of the skull; this again is something I am not sure of the translation for. I have to separate and explore the 'appos' relation a bit further - I believe it is responsible for equating two concepts, which would be represented with an eq link, but there are some cases where the dependency parser identifies that relation but it does not seem to be correct. Similarly for the 'acl' relation - I am reasonably sure it is going to be used as a "relation" relation - that is, the dependent word is going to represent a relation between the concept in question and some other concept, e.g. biceps as muscle with two origins (the relation would be "have"). However, while the 'acl' relation tells me the relation in Scone I would be creating, I still need to find the second "end" of that relation - "two origins" in the above example.

With that in mind, as my goal was to deliver some prototype extracting the types of relations I talked about in the previous progress report, I would say I have mostly met this goal.

0.2 Looking Ahead

I am going to continue adding the relations talked about into Scone, hopefully including all the relations I talked about above by the next milestone.

0.3 Revisions to Future Milestones

The next milestone prescribed the addition of roles into Scone; since I have been looking at roles and relations at the same time, the next milestone is going to be more like the original previous milestone (adding relations) and the coming one (adding roles) as one.

0.4 Resources Needed

I have all the resources I need so far.