15-400 Milestone Report II

Sonya Anopa

February 12, 2016

0.1 Milestone Progress

Since the last report, I've taken some steps that in some ways put me ahead and in some behind the milestone schedule. My stated goal was to have a prorotype of a tool that would extract information from WordNet into a useful intermediate representation that would then be converted into Scone elements, relations, and so forth. To simplify my goal, Dr. Fahlman suggested that I try to look at EPILOG and see whether I could use their representation as an intermediate representation since he had thought Dr. Allen's work generated such a representation. In fact, that was not the case, so exploring EPILOG did not provide me with useful information at first. However, after contacting Dr. Lenhart Schubert, the upkeeper of EPILOG, it turned out that a project associated with EPILOG called KNEXT (also written in Episodic Logic) has extracted some useful relationships from WordNet. Therefore, I have been writing a (mostly) Python tool that uses KNEXT's information. The tool at this stage roughly matches up with the next milestone - being able to extract is-a type links from definitions. Since those are encoded in KNEXT's statements, extracting them was pretty straightforward.

I have also pulled out equivalence links, not-links (concepts that are opposites), and branches of study (which are encoded separately). The not-links have also been pretty straightforward as they do not conflict with each other. However, the equivalence statements and the branches of study links are slightly more problematic. Branches of study are sometimes listed as both branches of a superclass and branches of each other - for example, bionomics and environmental science are both listed as branches of biology, but bionomics is also listed as a branch of environmental science. Figuring out how to resolve such cases is one of my future directions.

Secondly, the equivalence links are sometimes puzzling as well. They are originally represented as if-and-only-if statements, i.e. "something is a if and only if something is b." However, some cases are surprising - for example, process and physical process are listed in as being in such a relationship, even though process seems more general. This is also something I am going to be continuing to figure out.

0.2 Looking Ahead

Other than the plans I have already described, I hope to encode the other relationships from KNEXT as well. I know that the rest mostly have to do with proper names and

is-a links in a different form. KNEXT differentiates nouns by mass and count nouns, so I will attempt to preserve such classification in Scone. Some of the definitions also include plurals, which as far as I know are not representable in Scone, so I will need to figure out a way around that issue as well.

0.3   Revisions to Future Milestones

As I see it right now, my previous milestones could have "WordNet" replaced with "KNEXT" and otherwise stay the same; however, it remains to be seen exactly how successful the extraction will be.

0.4   Resources Needed

I have all the resources I need so far.