Design and Engineering of Intelligent Information Systems (DEIIS)

Homework #3

Weston Feely
andrew ID: wfeely

October 9, 2013

System Design

I created a CPE to replace the document analyzer from assignment 2, which uses the default FileSystemCollectionReader and XmiWriterCasConsumer from uimaj-examples. I copied the collection reader and xmi-writer cas consumer descriptors and java files from uimaj examples, because I wanted to make sure my CPE had everything it needed to run, all within my project folder. I then refactored the evaluation annotator that I made in assignment 2 into an evaluation cas consumer. This new cas consumer performs the same function as my previous annotator, since my annotator simply didn't create annotations, it just printed the precision and answer rankings to the console.

After that, I created a Stanford Core-NLP client to run Stanford Core-NLP over the documents, and I added this as an additional analysis engine to my CPE. I noticed the annotations produced by Stanford Core-NLP aren't quite right, because Stanford doesn't know that the format of our question and answers includes a 'Q' before question text, and a 'A [01]' before each answer text. The Stanford Core-NLP annotations considered these to be tokens, and part of the question and answer sentences. This caused erroneous named-entity mention annotations to be created like "Q Booth" and "A 1 Booth". I kept this in mind for the next part of the assignment. I then modified my evaluation cas consumer to include named entity matches between each question and its answers as part of the scoring metric for each answer. I did this by modifying the answerScore objects right before I ranked them, so that the fraction of matching named entities between the question and

the current answer was weighted as one-fourth of the answer score, and the original score was weighted by three-fourths. I then summed the two together to get the new scores. Since the bad "Q Booth" and "A 1 Booth" named entity mention annotations seemed incorrect, I disregarded these bad annotations during my matching by excluding any named entities that had a mention type of "null", which was the case for all these bad annotations, but not for the good annotations like "Booth" and "Mary", which were of type "PERSON". I found that the named entity matching didn't change my rankings, and thus didn't change my precision results, for either document. This was the case for when I disregarded the bad annotations, or if I didn't disregard the bad annotations.

I then made a UIMA AS deploy and client for my AAE, as well as a new CPE for my client. I tested out my new CPE, and it successfully produced xmi with my annotations and the Stanford annotations, as well as still printing the answer rankings and precision scores to console.