11-791: Homework 2

Maya Tydykov

1   System Pipeline

My system works by first reading in a test file. There are two collection readers defined for reading in the test file, which can be swapped in and out in the CpeDescriptor.xml. One of them is BioNECollectionReader, which takes in just a test file. The other is BioNEGoldStandardCollectionReader, which contains two parameters - one for a test file and another for a gold standard for that test file.

When reading in these files, the system populates SentenceData objects for each sentence, containing the sentence text and id. If a gold standard is provided and the BioNEGoldStandardCollectionReader is used, GoldMention objects are also created for each gold standard mention in the gold standard file.

Next, the system has several choices of annotators, which can be swapped in and out in the CpeDescriptor.xml. The first is the BioNESimpleAnnotator, which uses PosTagNamedEntityRecognizer, the baseline that came with the assignment. A more sophisticated annotator is the BioNeLingPipeAnnotator, which uses the LingPipe toolkit with a trained model file[1] in order to identify gene mentions. The final version is BioNECleanLingPipeAnnotator, which uses LingPipe but also checks to make sure that a certain set of characters is not in the extracted gene mention. I added this annotator because I noticed that in general, many mistakes LingPipe made seemed to stem from characters like parentheses included in the final output, which seemed to me to be good indicators that it was not really a gene mention. Thus, I added this simple rule-based filtering step to this final annotator.

2   Experimental Results

The baseline (BioNESimpleAnnotator using PosTagNamedEntityRecognizer achieves recall of 54.7%, precision of 10.25% and F1-Score of 17.3%.

---

[1]http://alias-i.com/lingpipe/web/models.html

Figure 1: UML class diagram for HW 1.
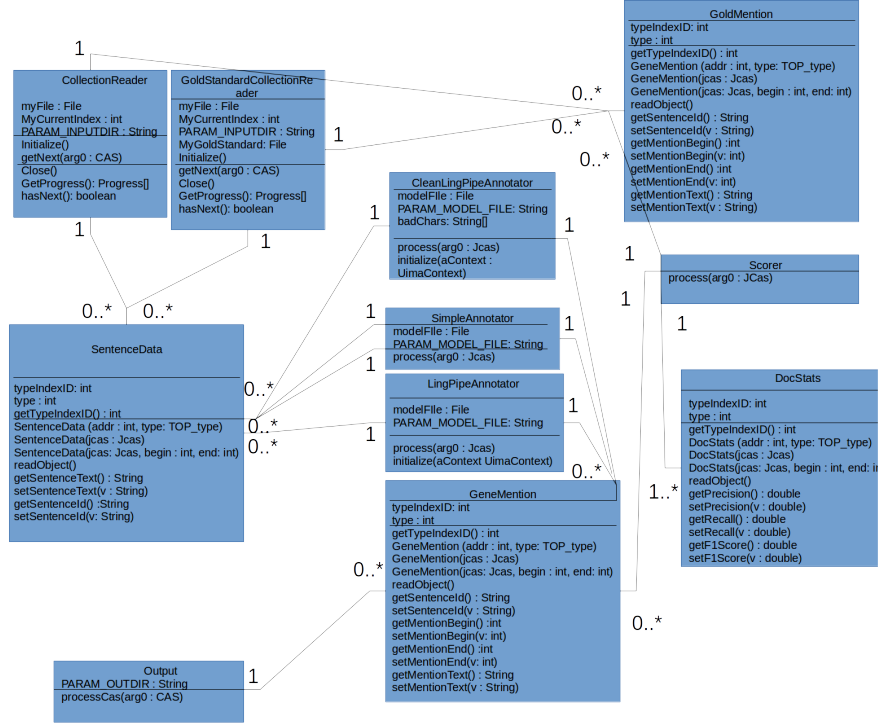
The BioNELingPipeAnnotator achieves recall of 84.9%, precision of 76.9% and F1-Score of 80.7%.

The BioNECleanLingPipeAnnotator achieves recall of 84.3%, precision of 78.3% and F1-Score of 81.2%.

Although the recall of the final version (LingPipe and rules together) achieves slightly smaller recall than using only LingPipe, the increase in precision is relatively higher than the decrease in recall, and the overall F1-Score is higher as well.