Name: Chen Wang

Andrew ID: chenw1

In order to complete the goal in this homework, I designed and implemented six annotators, one for each type in the given type system. The annotators are named as AnnotationAnnotator, AnswerAnnotator, AnswerScoreAnnotator, NgramAnnotator, QuestionAnnotator and TokenAnnotator.

It is not clear what each type in the type system refers to and what the feature of each type could be used for. In my implementation, edu.cmu.deiis.types.Annotation type points to the whole input document. edu.cmu.deiis.types.Answer points to each line that starts with letter "A" excluding the letter "A" and the number 1 or 0 at the beginning of the line. The feature isCorrect in Answer type is true if the answer is in a line that starts with "A 1 " and false if the line starts with "A 0". edu.cmu.deiis.types.AnswerScore is a wrapper of Answer type with a numerical score assigned to that answer. edu.cmu.deiis.types.Question points to the line that starts with letter "Q" excluding the "Q" in the beginning of the line. edu.cmu.deiis.types.Token points to those words in the document separated by non-word characters. 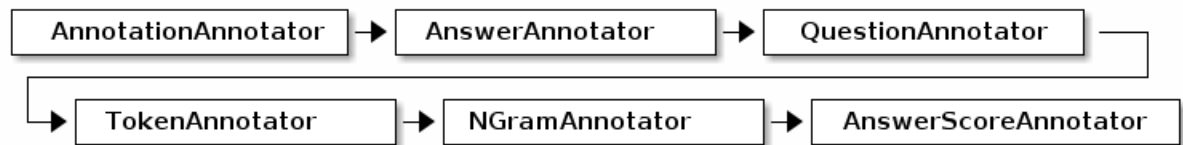And edu.cmu.deiis.types.NGram is a generalized edu.cmu.deiis.types.Token. Two or three tokens are considered as a 2-Gram or 3-Gram respectively, if and only if those tokens are in the same line and right next to each other. It is unclear what purpose the elements, elementType, edu.cmu.deiis.types.Question, and edu.cmu.deiis.types.Token feature inside edu.cmu.deiis.types.NGram serve so I do not use them.

The following provides a graphical view of the structure of the pipeline:

```
AnnotationAnnotator → AnswerAnnotator → QuestionAnnotator ┐
┌──────────────────────────────────────────────────────────┘
└→ TokenAnnotator → NGramAnnotator → AnswerScoreAnnotator
```

The last step in the pipeline, AnswerScoreAnnotator, serves as the final "merging" component that takes all the previous annotations and assign scores to each answer based on all existing annotations. It adopts N-Gram scoring method in which the score of an answer is calculated based on the ratio between question N-Grams that are found in answer N-Grams and the total number of answer N-Grams (for 1-, 2- and 3-grams). This annotator also prints out the final result to console, including the gold answer, the calculated result, original text input, and precision at N.