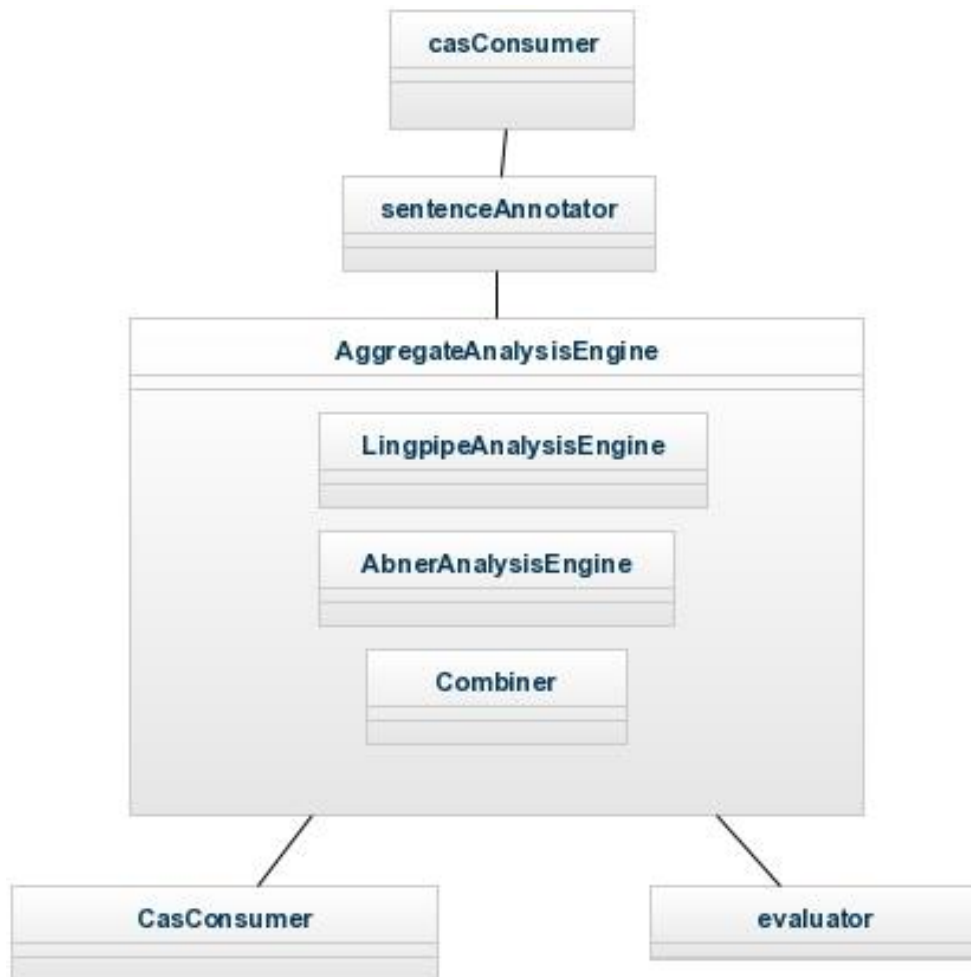# HW2-jianhel report

1.The UML graph

It is obvious in the graph that there are four annotators in the graph, and the three annotators LingpipeAnalysisEngine annotator, AbnerAnalysisEngine annotator and the Combiner annotator can combine to a AggregateAnalysisEngine annotator.

2.CollectionReader
The CollectionReader.java file is responsible for reading the sample.in file and put the content into the JCas.

3.SentenceAnnotator
The SentenceAnnotator.java file is responsible for make the sample.in file from the a whole String type into every sentence, and then put sentence into JCas to store.

4.SentenceTypeSystemDescriptor

The SentenceTypeSystemDescriptor.xml file define Annotation type of SentenceTypeSentence, which contains two type, the String SentenceID and the String SentenceContent.

5.typeSystemDescriptor

The typeSystemDescriptor.xml file define Annotation type of jianhelTypeSentence, which contains four types, the String fnTypeSystem1, which is used to store the sentences' ID, the Integer fnTypeSystem2, which is used to store the genes' begin positions, the Integer fnTypeSystem3, which is used to store the genes' end positions, and the String fnTypeSystem4, which is used to store the genes' names. Through the typeSystemDescriptor, we can do some action to annotate the genes from the file sample.in.

6.LingpipeAnalysisEngine

The LingpipeAnalysisEngine.java file is defined to get the JCas from the CollectionReader.java and use the external model Lingpipe to annotate the genes from the sample.in file, also memorize the sentences' ID and the start and end position of the genes.

7.AbnerAnalysisEngine

The AbnerAnalysisEngine.java file is defined to get the JCas from the CollectionReader.java and use the external abner.jar file and its' class to annotate the genes from the sample.in file, also memorize the sentences' ID and the start and end position of the genes. In the Abner model, we can choose to annotate the "protein", "dna", "rna" and so on.

8.CombineAnnotator

The CombineAnnotator.java file is define to combine the results from the Lingpipe analysis engine and Abner analysis engine. And then we can figure out the final results. We can improve the result's precision and recall by combining these two analysis engine to form a aggregate analysis engine.

9.CasConsumer

The CasConsumer.java file is defined to catch the JCas and the use the FileWriter and BufferWriter to print out the result. Also when we print out the result, we invoke the function from Evaluater class.

10.Evaluater

The Evaluater is defined to calculate the precision, recall, F-score and so on of the output results. Combining with the CasConsumer.java file the Evaluater.java together can print out the detail report of the output result.

11.SimpleRunCPE

SimpleRunCPE.java is defined to run the whole program, I thought that it is the main function of the whole program, it can automatic invoke the other .java file from the project to run the whole project.

12.cpeDescriptor

cpeDescriptor.xml file is the configuration file the SimpleRunCPE.java. I can change the source file in the cpeDescriptor.xml file to change the procedure of the whole process. So the cpeDescriptor.xml file just like the configuration file of the whole program.