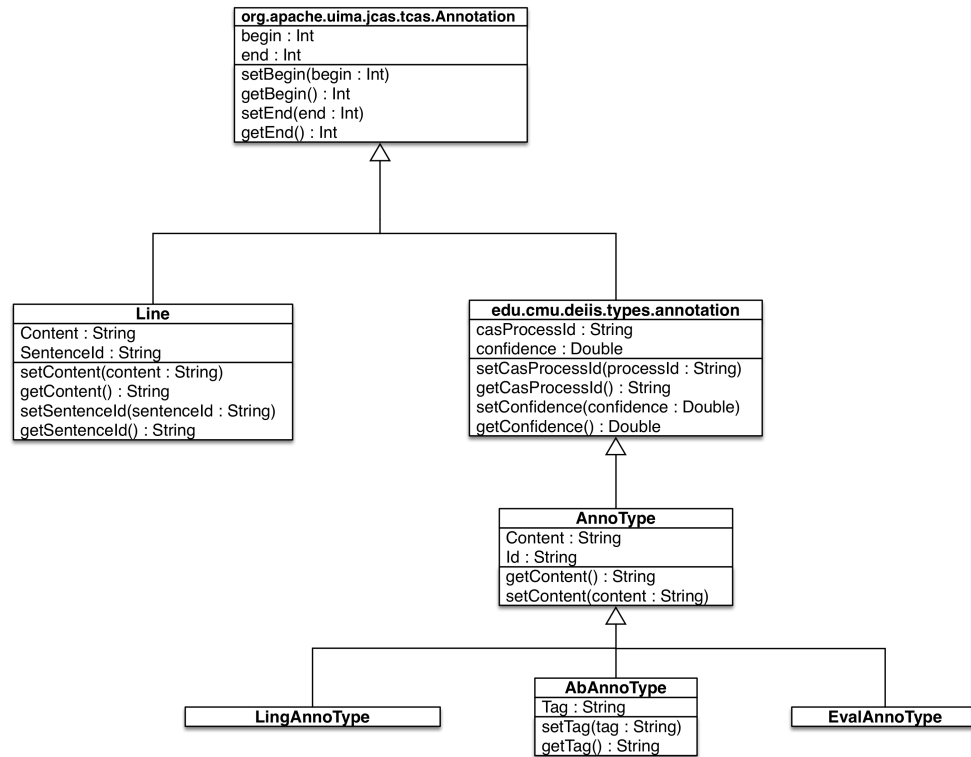


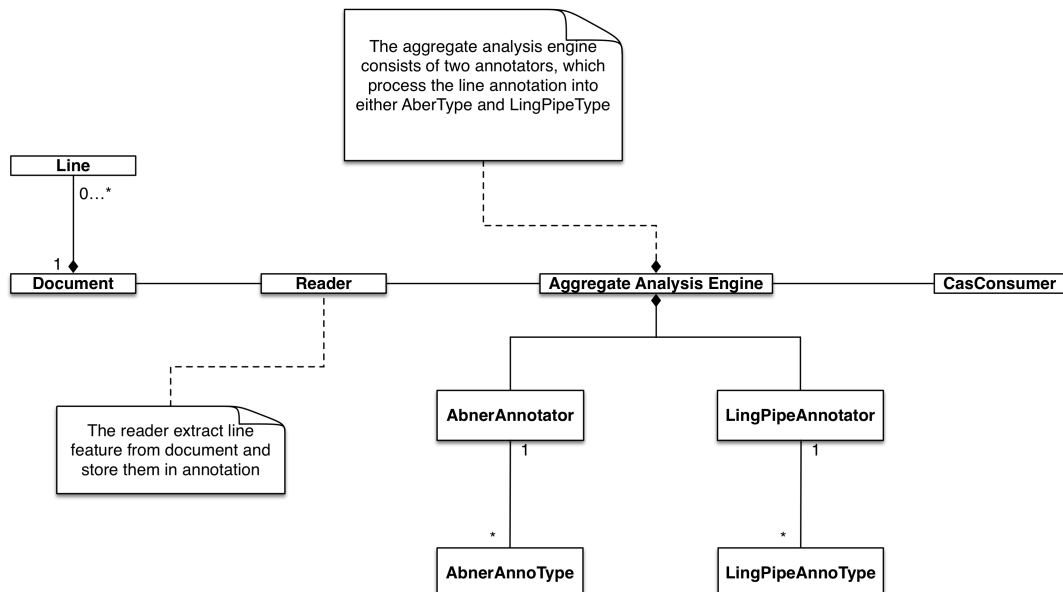
HW2 Final Report

UML Diagram

Type System Class Diagram



Collection Processing Engine Design Diagram



Collection reader

The reader will read one line of document into system at one time.

Aggregate Annotators

The aggregate analysis engine consists of three annotators.

The first one is based on LingPipe, which recognize the name entity based on the metric of confidence by Confidence Named Entity Chunking.

The second annotator is based on Abner. This annotator will give each entity a special tag based on its category of biological terms.

The last annotator will merge the results produced by the previous two analysis engines. In terms of the threshold of confidence, all of the entities which have a confidence below that threshold will be excluded from the results given by LingPipe-based annotator. However, the corresponding results from Abner-based annotator are taken into consideration by the last annotator, which may increase the F-measure result.

CAS Consumer

CASConsumer will write all of the gene that the system analyzed into one single file according to the given format. In addition, the precision, recall and F-measure can also be calculated according to the given sampleout file.

Running result

```
Precision: 0.614186064144  
Recall: 0.886996988776  
F1 Score: 0.725802477432
```

Reference:

Alias-i, L. (2008). 4.1.0. URL <http://alias-i.com/lingpipe>.

ABNER: A Biomedical Named Entity Recognizer. (n.d.). Retrieved October 9, 2014.

URL <http://pages.cs.wisc.edu/~bsettles/abner/>