HW2 Report

Student: Xiangzhu Long

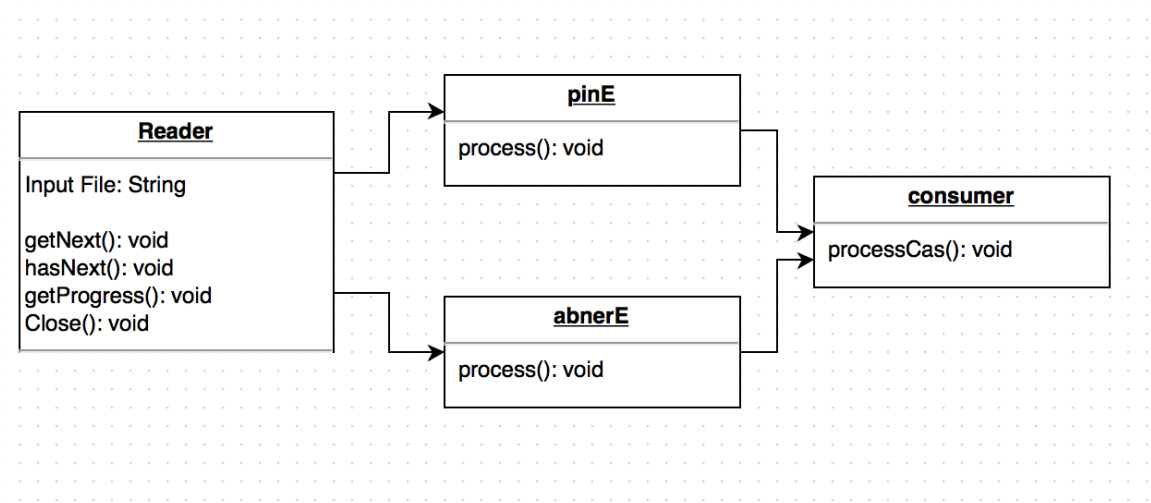
Andrew ID: xiangzhl

Date: 10/10/2014

1. Requirement

Design and implement an aggregate analysis engine in UIMA framework to detect Gene name entity in given dataset.

2. Design



3. Implement

3.1 Type System

This part is quiet similar with the Type system in hw1 except a middle Type that stores the information of engines number for a candidate token because I implement two AEs and one merge engine.

3.2 Collection Reader

The collection reader gets the gene information from the input sentence by sentence, sends the sentence ID directly to the consumer and put context to the aggregate AE.

3.3 Aggregate Analysis Engine

The AAE is designed to evaluate a token using two different tools. This project implements two AEs, abner and pipeline, and output a score for each token.

3.4 CAS Consumer

The CAS Consumer retrieves the gene names from the AAE and the sentence ID from the reader, and print them in an output file.

4. Reference:

LingPipe: http://alias-i.com/lingpipe/

Abner: http://pages.cs.wisc.edu/~bsettles/abner/