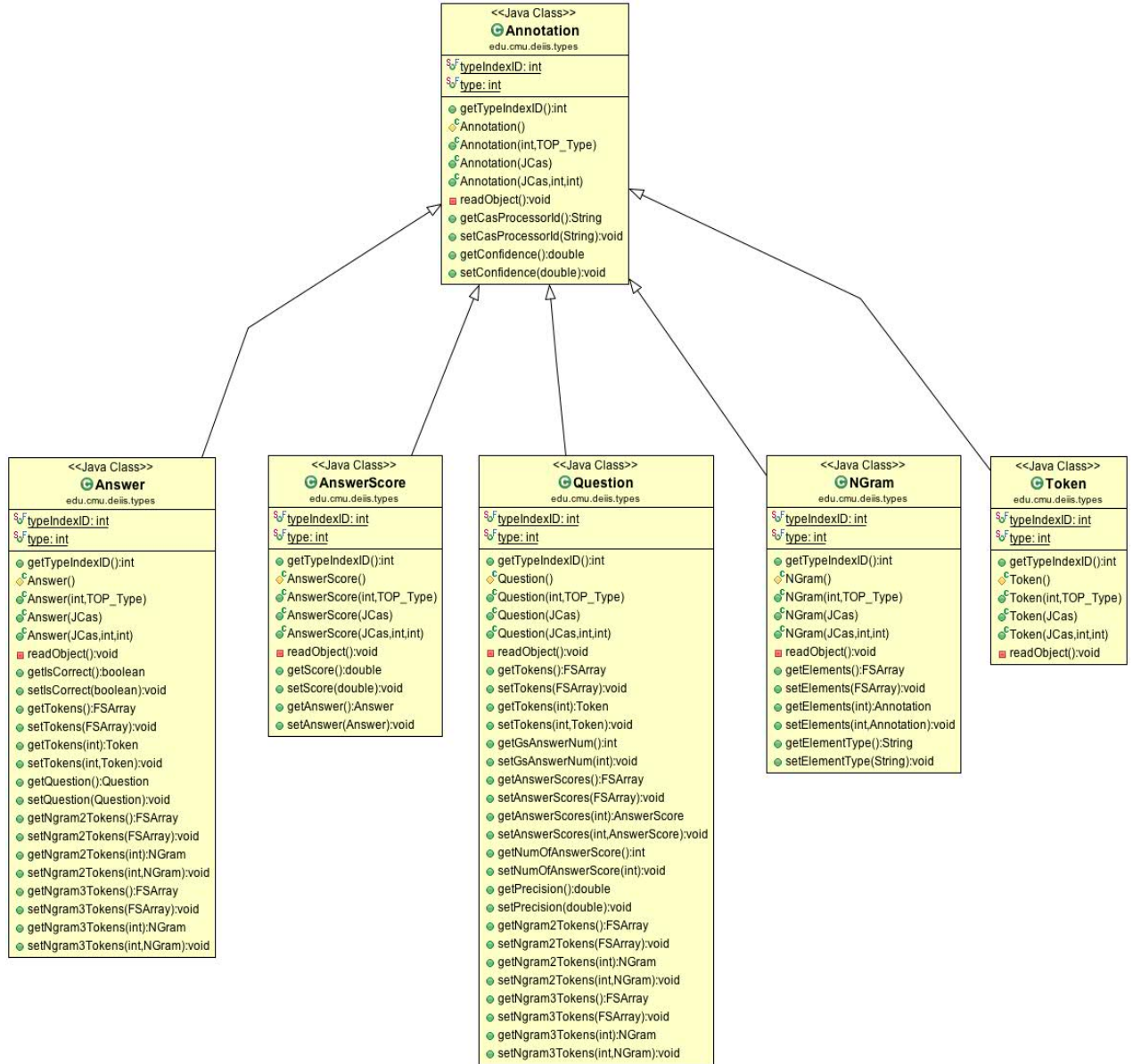


Homework II Report

Ran Zhao 2013/9/24

Part I. Type System Overview



Based on the type system provided, I put several additions on each type as followed.

(1) Type Name: Answer

Addition field:

Question: It indicates the question corresponding to this answer.

Tokens: The tokens of this answer.

Ngram2Tokens: The NGram2 tokens of the answer.

Ngram3Tokens: The NGram2 tokens of the answer.

(2) Type Name: Question

Addition field:

Tokens: The tokens of this question.

gsAnswerNum: The number of the golden standard correct answer.

numOfAnswerScore: The number of current available answerscores to this question.

Precision: The precision of this question.

Ngram2Tokens: The NGram2 tokens of the answer.

Ngram3Tokens: The NGram2 tokens of the answer.

Part II. Analysis Engine Implementation

Input: Two Q&A files.

Process:

- **TestElementAnnotator**: In this part, the annotator will parse the input file. It will use regular expression pattern to separate the question and its corresponding answers. The output of this AE is Question Annotation and Answer Annotation.
- **TokenAnnotator(including NGram Annotator)**: In this part, the annotator uses Lingpipe package to do the tokenizer and NGram annotation task. The output of this AE is Token annotation, NGram annotation which includes N=1, N=22, N=3. Each question and answer has its own token list, NGram2 list and NGram3 list. It will be convenience for future overlap scoring.
- **AnswerScoreAnnotator**: In this part, I use token overlap method and NGram overlap as they were described in the class. Firstly, I will assign each answers based on the proportion of the token and NGram overlaps between question and answer. Then I give each of the proportion a weight for calculation the socre.
- **Evaluation**: In this part, firstly, I will order the answers of each question based on their assigned scores. Secondly, as I have the number of golden standard correct answer in the TestElementAnnotator step, I can calculate the precision

of the answers at the number of golden standard correct answer.

Output format:

Question:Booth shot Lincoln?

+Booth shot Lincoln. 1.0

-Lincoln shot Booth. 0.16666666666666666

+Booth assassinated Lincoln. 0.11111111111111111

-Lincoln assassinated Booth. 0.11111111111111111

+Lincoln was shot by Booth. 0.09999999999999999

-Booth was shot by Lincoln. 0.09999999999999999

+Lincoln was assassinated by Booth. 0.06666666666666667

-Booth was assassinated by Lincoln. 0.06666666666666667

Precision 4: 0.5

Question:John loves Mary?

+John loves Mary. 1.0

+John loves Mary with all his heart. 0.28253968253968254

-John doesn't love Mary. 0.08333333333333333

-Mary doesn't love John. 0.08333333333333333

+Mary is dearly loved by John. 0.05555555555555555

Precision 3: 0.6666666666666666