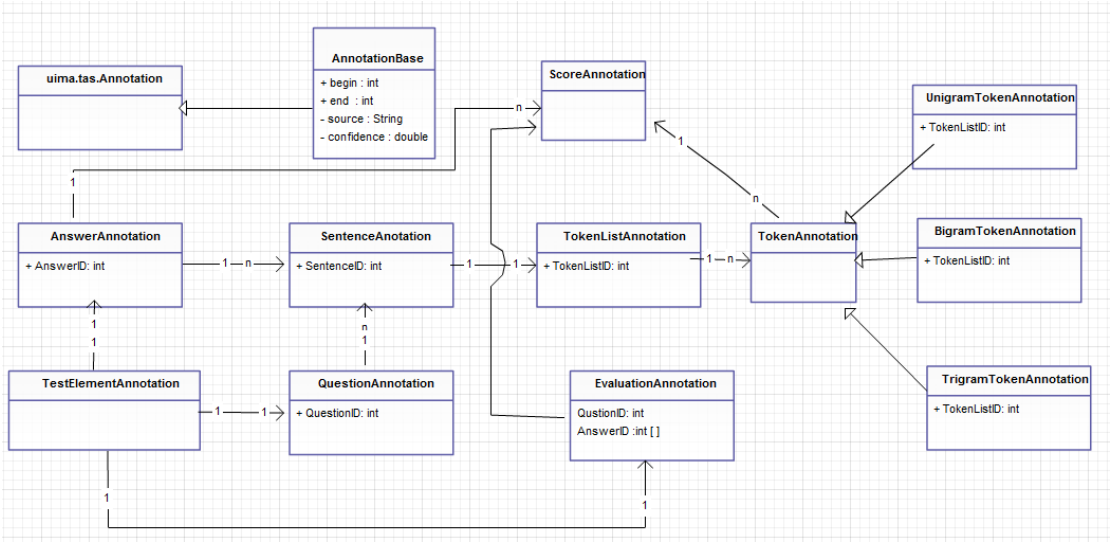
Hw1-jiacongh-Report

My designment of this project is shown in the class diagram as follow.



AnnotationBase:

In this picture, AnnotationBase is super class of all other classes. Because it is difficult to add arrow for all classes without messing the class diagram, I didn't add it in this picture.

AnnotationBase class extends the uima.tas.Annotation class, it has four attributes, the begin position, end position, source and the confidence. And this is extened by all other classes.

TestElementAnnotation:

When this pipeline is initiated, an TestElementAnnotation is first created to initiate the follow AnswerAnnotation, QuestionAnnotation and the final EvaluationAnnotation.

QuestionAnnotation:

Used for annotating questions. Assign each question a questionID to identify and better get the question. Questions may contains more than one sentence, so a QuestionAnnotation object may contains more than one SentenceAnnotation object.

AnswerAnnotation:

Used for annotating answers. Assign each answer an answerID to identify and better get the answer. Answer may contains more than one sentence, so a AnswerAnnotation object may contains more than one SentenceAnnotation object.

SentenceAnnotation:

Used for annotating sentences. One sentence will be extracted to only one tokenlist. We assign each sentence a sentenceID to identify and search them.

TokenListAnnotation:

Used for annotating tokenlist. TokenList is the list of tokens in one sentence. One tokenlist should contain more than one token. We assign each tokenList a tokenListID to identify and search them.

TokenAnnotation:

Token is the minimize unit in this pipeline. Because the system will annotate 1-, 2- and 3-grams of consecutive tokens. This token type is super type of unigram token, bigram token and trigram token.

ScoreAnnotation:

The system will incorporate a component that will assign an answer score annotation to each answer. The answer score annotation will record the score assigned to the answer.

EvaluationAnnotation:

The system will sort the answers according to their scores, and calculate precision at N (where N is the total number of correct answers).