# 11791 Design and Engineering of Intelligent Information Systems
## Fall 2013 Assignment 1

Kuo Liu

kuol@andrew.cmu.edu

September 10, 2013

## Implementing Logical Data Model with UIMA SDK

Figure 1 illustrates the overall data flow that occurs among the different types of components which make up the Analysis Engine.



Figure 1: Annotator components

# Type System

The Type System defines six types: *QuestionAnswer*, *Question*, *Answer*, *AnswerScore*, *Token* and *NGram*.

**QuestionAnswer**:

QuestionAnswer is the father of Question and Answer, it has three features, which are qaString (recording the content), annotator (recording the annotator) and score (recording the confidence score).

**Question**:

Question inherits the features of QuestionAnswer.

**Answer**:

Answer inherits the features of Question and it has another feature called yesOrNo, which records whether this answer is true or false (if yesOrNo="Yes", then the answer is true, otherwise, false). Answer is also the father of AnswerScore.

**AnswerScore**:

AnswerScore inherits the features of Answer.

**Token**:

Token has three features which are token (recording the content), annotator (recording the annotator) and score (recording the confidence score).

**NGram**:

NGram has four features which are ngram (recording the string of ngram), tokens (recording each token of the ngram string), annotator (recording the annotator) and score (recording the confidence score).
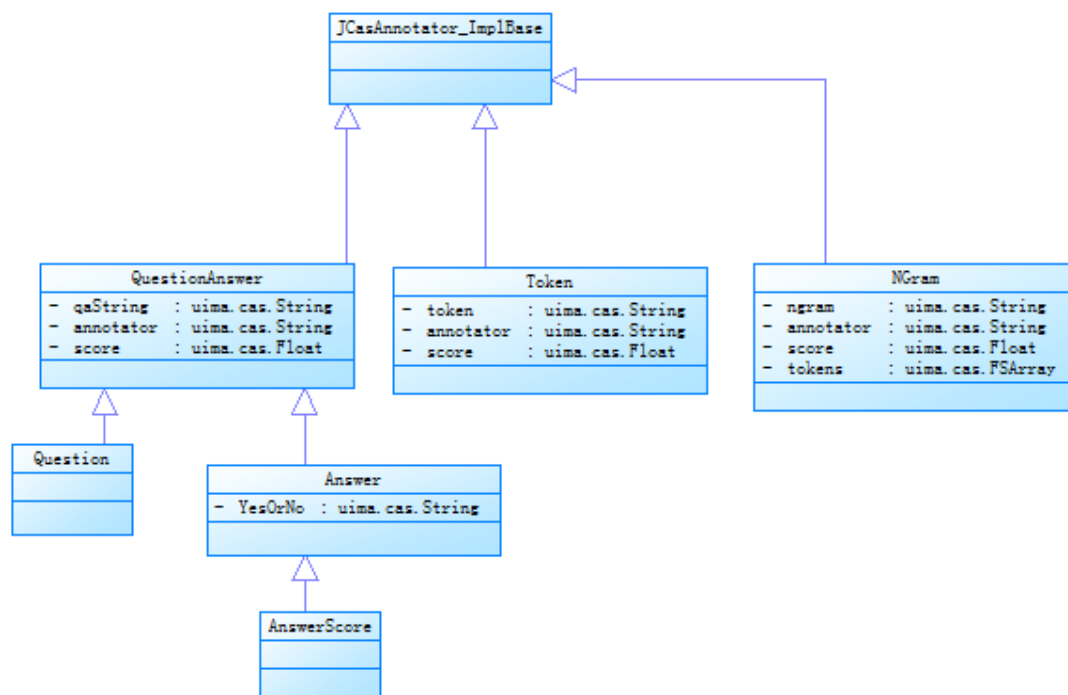
The Class Diagram is as follows.



Figure 2. Class Diagram for the type system.

# Annotator

**InformationProcessingAnnotator**:

InformationProcessingAnnotator is an aggregate Analysis Engine made up from *QuestionAnswerAnnotator*, *TokenAnnotator*, *NGramAnnotator* and *AnswerScoringAnnotator*, and the order for the four annotators is just as listed.

**QuestionAnswerAnnotator**:

QuestionAnswerAnnotator is used for annotating all the Questions and Answers.

**TokenAnnotator**:

TokenAnnotator is used for annotating all the Tokens. This annotator makes use of the index of Question and Answer.

**NGramAnnotator**:

NGramAnnotator is used for annotating NGrams (1-, 2-, 3-). This annotator makes use of the index of Token, Question and Answer.

**AnswerScoringAnnotator**:

AnswerScoringAnnotator, which annotates the AnswerScores, gives the score for each answer, sorts the answers according to their scores and calculate the precision at N (where N is the total number of the correct answers). Moreover, it prints the result to standard output.