Name: Wenqing Yuan
Andrew ID: wyuan

Logical Data Model and UIMA Type System

1. Type system

The type system consists of six types: baseAnnotation, question, answer, answerScoreAnnotator, TokenAnnotator and NGramAnnotator.

The **baseAnnotation** is a superclass of other five types because it has features that are common in all kinds of annotators, namely, the annotator's name who made the annotation and its confidence in the annotation.

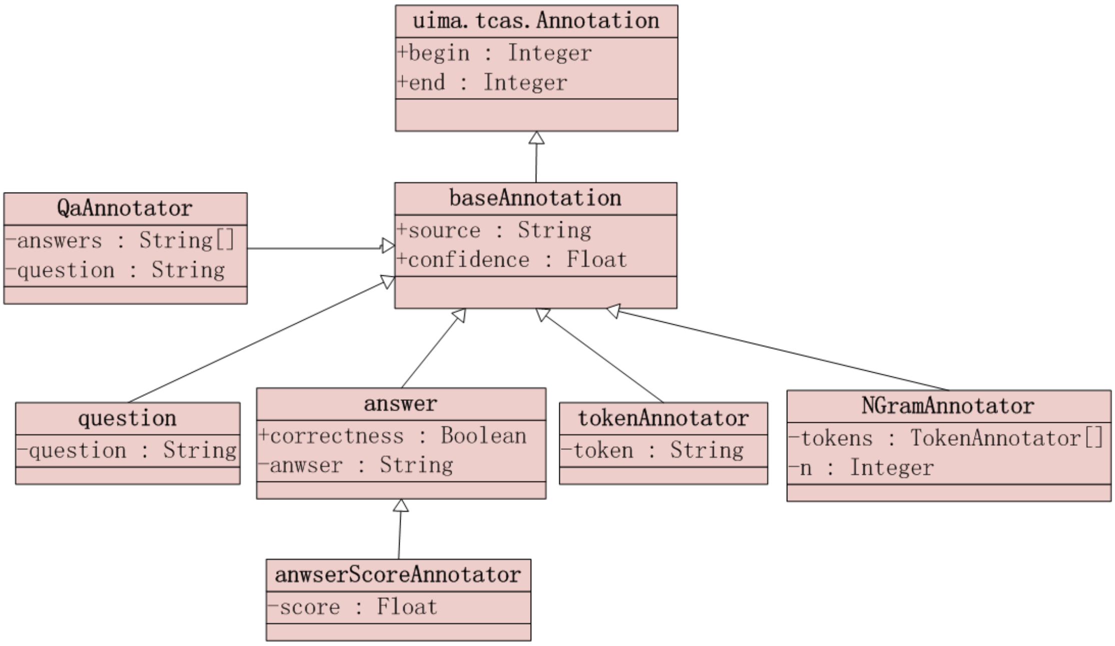
QaAnnotator extracts text lines from input file and create a question or answer object to save the text line. This is the first step of analyzing.

Type **question** and **answer** are separated because answers have one more feature (correctness) than questions, which can be seen in input files. Besides, answers will be tokenized and given a score.

tokenAnnotator corresponds to "token annotation" in the processing pipeline, and is responsible for annotating each token span in each question and answer.

NGramAnnotator annotates 1-, 2- and 3-grams of consecutive tokens. The Integer feature "n" can only be valued as 1,2 or 3 and annotator will annotate n-grams of tokens according to n. The feature "tokens" is an array of tokenAnnotator objects which are produced by tokenAnnotator and acts as input data.

AnswerScoreAnnotator is subclass of type answer because the two type objects are one-to-one correspondent. Besides, an AnswerScoreAnnotator object exists only when there is an instance of answer. The AnswerScoreAnnotator has only one feature to record the score annotation to an answer.



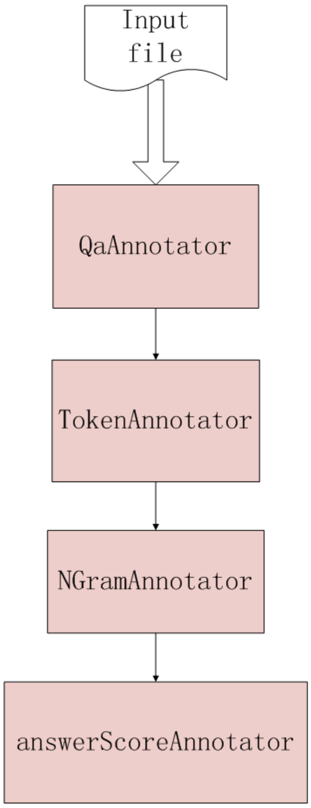
Picture 1

UML static class graph

2. Architecture design

The annotators work together according to the processing pipeline. The QaAnnotator firstly takes an input file which contains raw text question and answers. By reading each line of input file, it creates question object and answer objects which are then delivered to tokenAnnotator. TokenAnnotator tokenizes each question and answer objects and generates tokens as output. NGramAnnotator continues to annotate n-gram. Finally, all tokens and ngrams are sent to AnswerScoreAnnotator who assign a score annotation to each answer.

Type name	input	output
QaAnnotator	file	Question and answer
question		
Answer		
answerScoreAnnotator	Answer, token, ngrams	Scored answer
tokenAnnotator	Question, answer	token
NGramAnnotator	Question, answer, token	ngram



Picture 2

Data Flow Diagram