Andrew id: Wyuan
My name: Wenqing Yuan

11-791

Logical Architecture and UIMA Analysis Engines Design

& Implementation

1. Design

AnswerScoreAnnotator
If TokenAnnotator is used, then AnswerScoreAnnotator gives every answer a score based on the overlap rate of tokens in answers and the question.

Evaluator
No matter which annotator and scoreAnnotator are used, Evaluator calculates the accuracy of the prediction.

NGramAnswerScoreAnnotator
If NgramAnnotator is used, then NGramAnswerScoreAnnotator gives every answer a score based on the overlap rate of Ngrams in answers and the question. It really does the same thing as AnswerScoreAnnotator. The reason why I create the two answerScoreAnnotator is that they have different objects, namely, token and Ngram.

QuestionAnswerAnnotator
This annotator annotates Question and Answers. Question and Answers can be distinguished from their initial word, namely, A or Q. If it is an answer, annotator sets it truth value.

TokenAnnotator
This annotator uses regular expression to tokenlize question and answers which are generated from QuestionAnswerAnnotator.
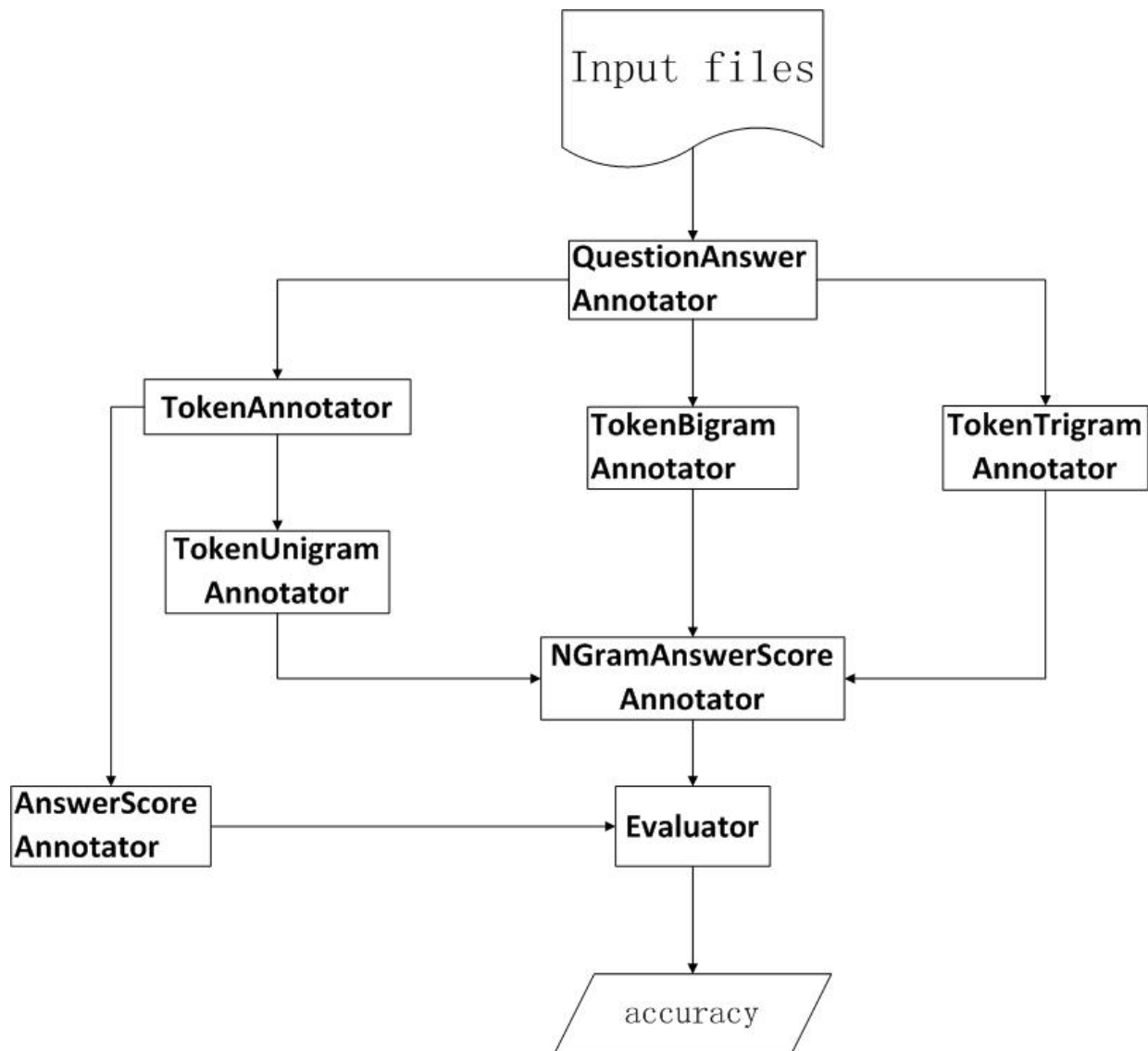
TokenBigramAnnotator
It combines two consecutive words within one sentence into a bigram.

TokenTrigramAnnotator
It combines three consecutive words within one sentence into a trigram.

TokenUnigramAnnotator
It uses the tokens generated by TokenAnnotator and simply create unigrams.

This figure points out the pipeline of the analysis engine in which annotators collaborate to output accuracy of prediction.

2. Accuracy

I experimented four different annotators (TokenAnnotator, TokenUnigramAnnotator, TokenBigramAnnotator and TokenTrigramAnnotator) separately with two input files(q001.txt and q002.txt). The diagram below is the output accuracy of them.

If the score of an answer is larger than a threshold, this answer will be judged as correct. I experimented with two thresholds and they get different accuracies.

Threshold = 0.5

	q001.txt	q002.txt
TokenAnnotator	0.5	0.2
TokenUnigramAnnotator	0.5	0.2
TokenBigramAnnotator	0.625	0.6
TokenTrigramAnnotator	0.625	0.6

Threshold = 0.333

	q001.txt	q002.txt
TokenAnnotator	0.5	0.6
TokenUnigramAnnotator	0.5	0.6
TokenBigramAnnotator	0.625	0.8
TokenTrigramAnnotator	0.625	0.6