

Motivation Behind the Design

The functionality of the system can be concisely described as following:-

1. Disintegrate the data and reduce it to finer granularity.
2. Perform data enrichment via annotation.
3. So, the token types exhibit great deal of similarity but differing in terms of granularity.
4. On the other hand the annotations like BaseTokenAnnotation NgramToken diifer in terms of enrichment features.

TestElement

It provides an encapsulation for the text of a file. The text of the file is stored in the feature lines.

UIMACASAnnotatedDocument

This type is created by processing the TestElement. It consists of a question span and a list of answer spans.

BasicTokenList

It is a list of basic tokens and provides an encapsulation for it. After having sorted out question spans and answer spans the system will break them on white spaces and punctuation. From this step we will get BasicTokenList type.

NGramTokenAnnotation

It extends BasicToken and adds additional required features like the confidenceScore, partOfSpeech . This annotation will be the output of NgramToken Annotator. It will produced by taking in annotated words and using them as either 1,2 or 3 groups of words.

AnswerScoreAnnotation

This type will consist of answer and the score associated with it. After processing the NgramTokenAnnotation the system will assign the score to each answer span.

EvaluationReport

This type will be produced after the evaluation step. It will consist of all the answers sorted on the basis of their scores and the precision of the answer Span.