Motivation Behind the Design

The functionality of the system can be concisely described as take input data and perform enrichment(in form of annotations) on it of various degrees at various steps. Hence, what we have is after each step the data is being disintegrated, evolving and being specialized. The system is best viewed as a a layered architecture. So, the design methodology which I have chosen is a reflection of this approach.

Walk through of the system types

(Items listed in the order of processing)

TestElement

A test element will consist of all the lines in the file. The text of the file is stored in the feature lines.

UIMACASAnnotatedFile

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

GenericTokenList

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 GenericTokenList type. It is a list of generic tokens and provides an encapsulation for it.

NGramTokenAnnotation

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.It extends GenericToken and adds additional required features like the confidenceScore, partOfSpeech & producer (the moule producing the token).

AnswerScoreAnnotation

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

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.