

## **HW2-Report**

### **TypeSystemExtensions**

#### 1)edu.cmu.deiis.subTypes.AnnotatedToken

AnnotatedTokenType extends the edu.cmu.deiis.types.Annotation by adding the:-

tokenText: stores the text which was annotated.

pos: (part of speech) which part of speech the words belong to.

#### 2)edu.cmu.deiis.subTypes.AnnotatedNGram

The type extends the edu.cmu.deiis.types.NGram and adds :-

nGramToken:it is the nGramText

#### 3)edu.cmu.deiis.subTypes.AnnotatedAnswer

It extends the answer score and adds the following :-

text: it is the sentence text

answerId:A unique ID associated with the answer

sentiment:Whether the sentence is negative or positive.

#### 4)edu.cmu.deiis.subTypes.AnnotatedQuestion

It extends the question type and adds the following feature:-

text: it is the sentence text

sentiment:Whether the sentence is negative or positive.

#### 5)edu.cmu.deiis.subTypes.Document

It contains:-

question:An AnnotatedQuestion to store details of the question in the document.

Answers:An Array to hold the answers

threshold:A confidence level associated with a document. Below this answers will be marked as false else true.

#### 6)edu.cmu.deiis.subTypes.TokenizedSentence

This type extends the UIMA Annotation and adds the following:

annotatedtokens:An array to hold tokens

#### 7)edu.cmu.deiis.subTypes.TokenizedDocument

This type extends the UIMA Annotation and adds the following features:

tokenizedQuestion: This is of type AnnotatedQuestion to hold the question.

tokenizedAnswers:This is an array to hold answers

#### 8)edu.cmu.deiis.subTypes.NGramMatrix

This type extends the UIMA Annotation and adds the following

matrix: An  $X \times N$  array. Where X represents the the number of answers and N represents size of Ngram.

It stores all the Ngrams associated with an answer,

## **Annotators**

### 1)TestElementAnnotator

It takes in a document as CAS type.

It separates the question and answers

To each answer it assigns whether it is true or false.

It assigns sentiment to the question and each answer i.e. -1 for -ve & +1 for +ve.

It creates a AnnotatedQuestion and an AnnotatedAnswerArray which are associated with a edu.cmu.deiis.subTypes.Document.

### 2)TokenAnnotator

It will receive the edu.cmu.deiis.subTypes.Document and from it extract the AnnotatedQuestion and AnnotatedAnswerArray

It will split them into tokens.

These tokens will be assigned additional metadata.

The generated tokens will be stored in edu.cmu.deiis.subTypes.TokenizedDocument.

### 3)NGramAnnotator

This annotator will receive the edu.cmu.deiis.subTypes.TokenizedDocument.

It will take the tokens and combine it into NGrams tokens.

i.e. For a tokens of a given sentence it will generate the unigram, bigram,... ngram tokens associated with the sentence.

### 4)AnswerScorer

It will check the sentiment of question and answer.

If they are incompatible i.e one has a sentiment score of 1 and other -1 it will arbitrarily assign it a confidence and score of 0.

If they are of compatible sentiments, it will take all the nGram for a given and answer and try to match them with question.It will average the number of NGrams matched.

In case the answer text is longer than the question it will multiply the answer scores by a factor of (Alength/Qlength). This is done so that in a case where answer has more tokens in a question, the mismatches which occur due to data being absent in question should not reduce the confidence.

The above steps give the confidence of each answer.

### 5)Evaluator

It will sort the answers based upon their confidence scores using a comparator.

It will display the answers in decreasing orders of their confidence.

Along with the answer text it will display the answer confidence, the score(1 or 0) assigned to the answer assigned by the system and the gold standard score.

```

classDiagram
    class edu_cmu_deiis_types_Annotation {
        casProcessorId : Integer
        confidence : Double
    }
    class edu_cmu_deiis_types_Answer {
        isCorrect : Boolean
    }
    class edu_cmu_deiis_types_AnswerScore {
        score : Float
        answer : edu_cmu_deiis_types_Answer
    }
    class edu_cmu_deiis_subTypes_AnnotatedAnswer {
        text : String
        answerId : String
        sentiment : String
    }
    class edu_cmu_deiis_subTypes_AnnotatedToken {
        tokenText : String
        pos : String
        newAttr : Integer
    }
    class edu_cmu_deiis_subTypes_AnnotatedQuestion {
        text : String
        sentiment : String
    }
    class edu_cmu_deiis_subTypes_AnnotatedNGram {
        text : String
    }
    class edu_cmu_deiis_subTypes_NGram {
        elementType : String
        elements : Array
        newAttr : Integer
        newOperation()
    }
    class edu_cmu_deiis_subTypes_NGramMatrix {
        matrix : edu_cmu_deiis_subTypes_AnnotatedNGram
    }
    class edu_cmu_deiis_subTypes_Document {
        question : edu_cmu_deiis_subTypes_AnnotatedQuestion
        answer : edu_cmu_deiis_subTypes_AnnotatedAnswer
        threshold : Double
    }
    class edu_cmu_deiis_subTypes_TokenizedDocument {
        question : edu_cmu_deiis_subTypes_TokenizedSentence
        answers : edu_cmu_deiis_subTypes_TokenizedSentence
    }
    class edu_cmu_deiis_subTypes_TokenizedSentence {
        annotatedTokens : edu_cmu_deiis_subTypes_AnnotatedToken
    }

    edu_cmu_deiis_types_Annotation <|-- edu_cmu_deiis_types_Answer
    edu_cmu_deiis_types_Annotation <|-- edu_cmu_deiis_subTypes_AnnotatedToken
    edu_cmu_deiis_types_Annotation <|-- edu_cmu_deiis_subTypes_AnnotatedQuestion
    edu_cmu_deiis_types_Annotation <|-- edu_cmu_deiis_subTypes_AnnotatedNGram
    edu_cmu_deiis_types_Annotation <|-- edu_cmu_deiis_subTypes_NGram
    edu_cmu_deiis_types_Annotation <|-- edu_cmu_deiis_subTypes_Document
    edu_cmu_deiis_types_Annotation <|-- edu_cmu_deiis_subTypes_TokenizedDocument
    edu_cmu_deiis_types_Annotation <|-- edu_cmu_deiis_subTypes_TokenizedSentence

    edu_cmu_deiis_types_Annotation "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedToken
    edu_cmu_deiis_types_Annotation "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedQuestion
    edu_cmu_deiis_types_Annotation "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedNGram
    edu_cmu_deiis_types_Annotation "1" -- "1" edu_cmu_deiis_subTypes_NGram
    edu_cmu_deiis_types_Annotation "1" -- "1" edu_cmu_deiis_subTypes_Document
    edu_cmu_deiis_types_Annotation "1" -- "1" edu_cmu_deiis_subTypes_TokenizedDocument
    edu_cmu_deiis_types_Annotation "1" -- "1" edu_cmu_deiis_subTypes_TokenizedSentence

    edu_cmu_deiis_subTypes_AnnotatedAnswer "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedToken
    edu_cmu_deiis_subTypes_AnnotatedAnswer "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedQuestion
    edu_cmu_deiis_subTypes_AnnotatedAnswer "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedNGram
    edu_cmu_deiis_subTypes_AnnotatedAnswer "1" -- "1" edu_cmu_deiis_subTypes_NGram
    edu_cmu_deiis_subTypes_AnnotatedAnswer "1" -- "1" edu_cmu_deiis_subTypes_Document
    edu_cmu_deiis_subTypes_AnnotatedAnswer "1" -- "1" edu_cmu_deiis_subTypes_TokenizedDocument
    edu_cmu_deiis_subTypes_AnnotatedAnswer "1" -- "1" edu_cmu_deiis_subTypes_TokenizedSentence

    edu_cmu_deiis_subTypes_Document "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedQuestion
    edu_cmu_deiis_subTypes_Document "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedAnswer
    edu_cmu_deiis_subTypes_Document "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedNGram
    edu_cmu_deiis_subTypes_Document "1" -- "1" edu_cmu_deiis_subTypes_NGram
    edu_cmu_deiis_subTypes_Document "1" -- "1" edu_cmu_deiis_subTypes_Document
    edu_cmu_deiis_subTypes_Document "1" -- "1" edu_cmu_deiis_subTypes_TokenizedDocument
    edu_cmu_deiis_subTypes_Document "1" -- "1" edu_cmu_deiis_subTypes_TokenizedSentence

    edu_cmu_deiis_subTypes_TokenizedDocument "1" -- "1" edu_cmu_deiis_subTypes_TokenizedSentence
    edu_cmu_deiis_subTypes_TokenizedDocument "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedToken
    edu_cmu_deiis_subTypes_TokenizedDocument "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedQuestion
    edu_cmu_deiis_subTypes_TokenizedDocument "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedNGram
    edu_cmu_deiis_subTypes_TokenizedDocument "1" -- "1" edu_cmu_deiis_subTypes_NGram
    edu_cmu_deiis_subTypes_TokenizedDocument "1" -- "1" edu_cmu_deiis_subTypes_Document
    edu_cmu_deiis_subTypes_TokenizedDocument "1" -- "1" edu_cmu_deiis_subTypes_TokenizedSentence

    edu_cmu_deiis_subTypes_TokenizedSentence "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedToken
    edu_cmu_deiis_subTypes_TokenizedSentence "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedQuestion
    edu_cmu_deiis_subTypes_TokenizedSentence "1" -- "1" edu_cmu_deiis_subTypes_AnnotatedNGram
    edu_cmu_deiis_subTypes_TokenizedSentence "1" -- "1" edu_cmu_deiis_subTypes_NGram
    edu_cmu_deiis_subTypes_TokenizedSentence "1" -- "1" edu_cmu_deiis_subTypes_Document
    edu_cmu_deiis_subTypes_TokenizedSentence "1" -- "1" edu_cmu_deiis_subTypes_TokenizedDocument
    edu_cmu_deiis_subTypes_TokenizedSentence "1" -- "1" edu_cmu_deiis_subTypes_TokenizedSentence

```