Report for Design and Engineering of Intelligent Information System and Software Methods for Biotechnology Homework 2

Logical Architecture and UMIA Analysis Engines Design and Implementation

Name: Trisha Black

ID: trishab

The logical architecture for capturing and calculating the tokens of provided input data require Annotators in for the following methods.

1) Question
2) Answer
3) Test Element (AllToken)
4) Token Annotation
5) NGram
6) Answer Evaluator
7) Evaluation

The aim for each type of annotator requires specific input and output to construct a working aggregate analysis engine.

The Question Descriptor capabilities required the Type Systems output.

The Answer Descriptor capabilities required the Type Systems output.

The Token Descriptor capabilities required the Answer and Question Type System input and the Token Type System Output.

NGrams Descriptor capabilities requires the Token Type System Input.

AnswerScore Descriptor capabilities requires the Token Input and NGrams Output.

The Aggregate Analysis Engine:

The methods described for each annotator determine the text that is assigned to Annotator. For the purpose of this example Ngrams determine an AnswerScore. The Ngrams are defined by the number of Tokens. A UniGram is one Token, A BiGram is two Tokens and TriGram is three Tokens. The AnswerScorer is based upon the number of Tokens in a NGram and whether or the statement is True or False. Some NGrams are the same in a false statement as they are in true statement. The calculation for determining the AnswerScore depends on the weight (1 or 0) of the Answer if it is either True or False, respectively.

To improve the methods for determine the AnswerScore it might be best to eliminate any words that are fillers and are not meaningful features to determine what Answer best fits the Question.