Name: William Ibekwe
Andrew ID: wibekwe
Date: September 11, 2013

Homework 1 Report

Description of the TypeSystem

Looking at the input, output, and the processing for the pipeline we are creating, there are some key types needed in the type system in order to successfully go through the pipeline. The input is broken up into two sections: Question and AnswerList types. These data types are encapsulated within another data type of type TOP called Input.

The Question type in the input has only one question asked, therefore inside Question is are type that contain the entire question string, counts the number of tokens in the question, contains list of all the tokens from the question string, and contains the cardinality of n-grams for every quantity set. TokenList and NGramList are different in the sense that TokenList just has all the different tokens, where as NGramList houses ever combination of tokens with the specific string. Breaking down the string into tokens and n-gram sets are important because we can use these types for Token and NGram Annotation in the pipelines. Keeping the question string is also important for the Test Element Annotation. It is used to create the token set of type TokenList and the n-gram set of type NGramList. The token list is broken down into tokens and contains information about whether or not a specific token is in the beginning or the end of a string. It also contains information about the type of speech the token is in a language model i.e. noun, adjective, verb etc. The n-gram list is also broken down into cardinality sets of 1-grams, 2-grams, and n-grams, where n is the number of tokens in a string.

The AnswerList type in this specific type system contains types that are isomorphic to those in the Question type. The difference is AnswerList has an extra level of abstraction since there are multiple answers in this model while question only contains one. AnswerList contains the number of answers in the list as well as each individual answer. AnswerList can contains anywhere from one to and infinite number of answers of type Answer. Answer is similar to Question, containing the answer string of a particular answer, a list of tokens for that specific answer string of type tokenList, and the weight of that answer of type double. The weight will be used for Answer Scoring and Evaluation purposes. Each answer will have a specific score or weight weight is used to evaluate each of the answer in comparison to the question asked. The token list is the same type used in the question data type.