Design and Engineering of Intelligent Information Systems (DEIIS)
Homework #2

Weston Feely
andrew ID: wfeely

September 23, 2013

System Description

I created annotators for each of the default data types, as well as my own "Eval" data type for the final evaluation. My annotators mark the begin and end character offsets for tokens, ngrams, questions, answers, and answer scores (same span as answers). I matched alphanumeric strings as tokens within the question and answer sentence texts, excluding the beginning 'Q' and "A 0/1" for each. I later had to include apostrophes in my regular expression that I used to match tokens, because this caused issues with my tokens in the second document. I maintained the sentence number for each token, in a sentenceId field, because this helped me break ngrams across sentence boundaries. My ngrams were unigrams, bigrams, and trigrams, built up of FSarrays of tokens. I added an order field to keep track of whether each ngram was a 1, 2, or 3-gram. I also maintained the document text covered by tokens, ngrams, questions, and answers, for later use. Aside from the begin and end character offset annotations, I added FSarrays of unigrams, bigrams, and trigrams belonging to each question and answer as a field to each question and answer object. I calculated scores for my AnswerScores by summing the number of unigrams, bigrams, and trigrams that matched from an answer to its question, and then divided by the total number of unigrams, bigrams, and trigrams belonging to the question. Finally, the Eval annotator I created ranks the AnswerScore objects by their score, calculates precision at N for each Q/A set, and prints this information to standard out.

Here is my final output of my system, on the input data provided:

```
Question: John loves Mary?
+ 1.0 John loves Mary with all his heart.
+ 1.0 John loves Mary.
+ 0.33 Mary is dearly loved by John.
- 0.33 Mary doesn't love John.
- 0.33 John doesn't love Mary.
Precision at 3: 1.0

Question: Booth shot Lincoln?
+ 1.0 Booth shot Lincoln.
- 0.5 Lincoln shot Booth.
+ 0.5 Lincoln was shot by Booth.
- 0.5 Booth was shot by Lincoln.
+ 0.33 Booth assassinated Lincoln.
- 0.33 Lincoln assassinated Booth.
+ 0.33 Lincoln was assassinated by Booth.
- 0.33 Booth was assassinated by Lincoln.
Precision at 4: 0.5
```