CS 421 Project Report – Spring 2015

Arthur Nishimoto (anishi2) and Jenny Sum (sum1)

Syntax / Grammar (1)

Our spell checking system as implemented for part (1a) works well with the following statistics for all provided essays (both training sets and the competition set). While our program is capable of running both the original and

Spelling Error (min|avg|max)

High Score: 1|7.4666667|18

Medium Score: 2|9.533334|24

Low Score: 0|11.866667|50

For each grade of low, medium, and high, there are clear divisions in how each set of essays were graded. One of the challenges of developing this part of the project was attempting to use existing NLP tools for spellchecking. We attempted to use the Apache Lucene and Jortho spellcheckers with minimal results. While the libraries were successfully integrated, both failed to meet standards for spell checking the essays after multiple attempts to use different dictionaries and configurations. The glaring issue was Lucene’s apparent inability to correctly detect common words of three letters or less (A, him, her, the, or, etc.). Our final solution was a more straightforward reading of a large text file of words from (<http://wordlist.aspell.net/>). This list contained over 310,000 words plus an additional set to include common proper English nouns. Our system also has several functions for checking the root of words in the dictionary using simple string manipulations – this is not the most elegant solution, but does a reasonable job. Obviously being able to leverage the spellchecking capabilities of one of the established NLP tools could improve this section.

For Subject-Verb Agreement (1b) and Verb Tense (1c) errors, the statistics are less clear.

Subject-Verb Agreement Error (min|avg|max)

High Score: 0|7.7|22

Medium Score: 0|6.8|18

Low Score: 0|4.233333|18

Verb Tense Error (min|avg|max)

High Score: 0|1.2666667|4

Medium Score: 0|1.3666667|13

Low Score: 0|0.33333334|2

While the error counts are somewhat inconsistent with what would be expected, the grading weights still calculate the overall grade of the essays within reasonable standards. These standards are defined as known high graded essays evaluated in our system as high or medium and known low graded essays evaluated in our system as low or medium. As long as no high essays are scored low or low essays scored high, we consider this within acceptable ranges. Clearly this section could be improved by more careful addition and refinement of the rules governing the Part-of-Speech (POS) tag scoring for this section.

Sentence Formation Error (min|avg|max)

High Score: 0|3.3333333|7

Medium Score: 0|3.6666667|12

Low Score: 0|3.1333334|11

We consider our Sentence Formation checker (1d) to be successful in providing reasonable scoring results based on the above statistics. Similar to part (1b) and (1c) the sentence formation checker uses the POS tagger provided by OpenNLP. In addition to the POS tags, part (1d) also utilized the chunker model to further associate words and POS tags with noun and verb groups. This additional information was useful for getting the clearer division between the low, medium, and high ranges for this section.

Semantics / Pragmatics (2)

Length of Essay (3a)