CSSE 413: Natural Language Processing

Qishun Yu, Yiyu Ma

CM 2043, CM 1789

The purpose of this project is to evaluate NLP and use it in grading text summarization. In order to accomplish this goal, we used parser tree and triples to perform information extraction and developed a grading system. During this lab, we have conducted a variety of experiments improving the algorithm performance and we believe the grading algorithm is reasonable and comprehensive.

# Parsing and Information Extraction

The very first procedure we performed is parsing and information extraction. We used the Stanford NLP kit to represent the syntactic structure of the whole text file. Both the summary and the original paper were first cleaned and broken into sentences. After that, those sentences were analyzed and broken down into constituency-based parse tree.

The second step is to extract information from the parsing tree. At first, we tried to get useful information directly form the parsing tree. However, this attempt has been proven not feasible. The reason being matching two text files to each leaf nodes is extremely hard and detailed. Therefore. We use triples to group a sentence into Subject, Object, and Predicate. This process is shown in Figure 1.

A screenshot of a cell phone

Description automatically generated

Figure 1. A schematic of Parsing and Information Extraction.

# Grading Algorithm

After the summary and original text sentences are categorized in triples, four different grading rubrics are applied and weighted to produce a final score. We