Detecting Plagiarism, Phase II

Fritz, Caleb

The purpose of this experiment was to enhance a plagiarism detection technique known as bag of words from phase I and apply a graphical user interface (GUI), allowing users to employ the program to detect plagiarism. The procedure involved enhancing the bag of words algorithm and developing a GUI that would allow a user to interact with it to compare Microsoft Word and Notepad documents. Testing involved comparing two documents discussing the same topic by applying varying degrees of text from one document, plagiarizing it at certain levels, and applying it to the second document at both the paragraph and sentence levels. Testing was also done to determine if verbatim plagiarism could be detected, similar documents could be distinguished, and found the average processing time for comparisons. Data from the experiments suggested that the program was capable of detecting verbatim text, copy and paste, and limited amounts of rogeting, or shake and paste plagiarism, while it failed to detect paraphrased text. The program reported cached and non-cached processing times below the 3 - 5 second limit for 50 documents, but exceeded the threshold for the non-cached Microsoft Word documents, where the processing time was nearly 9 seconds. The program did not provide users with the ability to enter URLs for website comparison due to dependency and reliability concerns.

Awards Won: