Text Mining Mini Project Reflection: Claire Kincaid, February 24th, 2016

My project is a keyword search engine for research papers. It is designed to take in the body of text from a research paper as a text file and create a dataset, then traverse the dataset and create a dictionary detailing the number of occurrences of keywords as specified by the user. It is designed so that researchers can easily identify relevant or effective research papers from the occurrences of specified keywords. From a process point of view, I find the code readable, and my doctests did their job very well. I picked doctests that would encompass as many situations as possible while remaining simple and not taking up a lot of space or becoming cumbersome and difficult to read or understand. My code runs smoothly and serves its intended purpose well. I could stand to reduce redundancy in my code or start to practice optimizing my code for shortest length. For its intended purpose, my project was very well scoped, 1 paper at a time. For a future project, I would love to create a higher order program that could determine relevancy of the paper based on user specified parameters of keyword occurrence, without the user having to decide on their own after viewing the final keyword histogram. Additionally, I'd like to increase the scope of the program so that it can take in a multitude of research papers and determine/rate relevancy of the research papers with respect to each other and user specified parameters.