

CS 8001 Big Data

HW #1: Word Count (10 points + 2 bonus points)

Spring 2016

(Due 1/27, Wednesday, midnight)

In this exercise, you will write Python code that outputs

1. (10 points) the 10 most common words in the first 3 sections of “The Sonnets” (about 330 words in total) from the [Complete Works of William Shakespeare](#) retrieved from [Project Gutenberg](#) and the number of occurrence of each of them.
2. (2 bonus points) the 20 most common words in the [Complete Works of William Shakespeare](#) retrieved from [Project Gutenberg](#) and the number of occurrence of each of them.

Submit electronically in Blackboard the following:

- 1) A brief description of your Python environment, code structure, and the execution process of the code.
- 2) Output of your code (10 or 20 most common words and their counts).
- 3) Execution time of your code in seconds.
- 4) Your code with appropriate comments.