

Bronx Community College
Department of Mathematics and Computer Science
CSI31 Fall 2018

Programming Assignment 7: worth 40 points
Assigned on December 5, 2018
Due on December 18, 2018

Assignment 7: Improve the program wordfreq.py by adding two features that will give important information about the document that was processed.

1. Improve the report by omitting from the report the common words that are on the list of stop words posted on the course website.

The program should still find the frequency of every word, including the stop words. The part of the program that you should modify is the part that prints the words and their counts from the sorted list. For each word in the list, you should check to see if it is a stop word, and print the word and its frequency only if it is not a stop word. The number of words printed will still be the number received as input, so you will need to count the number of words printed. You will need to read the stop words from the file into a list so you can check whether a word is in the list.

This improvement to the word frequency list will make it easier for you to decide the subject area of the document.

2. Add to the program the collection of a list of all lengths of all the sentences in the document. Use the list to find the average sentence length for the document.

For our purposes, a sentence is the sequence of words between one period, exclamation point or question mark, and the next one. The program as written first breaks the text into a list of words without any punctuation or special characters. To compute the length of sentences with our definition, you should leave in periods, exclamation points, and question marks at first in order to do the counting, and then remove them later.

The average sentence length gives you an idea of the reading level of the document. Technical documents usually have average sentence length between 15 and 20 words.

We will discuss both of these programming problems in class.

There will be three documents posted on the website for you to use in developing and testing your program. You should test your program on each of those files, and include the testing results when you submit your program. Submit your program file and the report file by email to me at prof.persinger@gmail.com by midnight on December 18, 2018.