CPSC 230: Computer Science I
Spring 2023
Programming Assignment 5
Due: May 8th, 2023 @ 11:59pm

Overview

A major application of artificial intelligence is natural language processing (NLP), where we train a computer to be able to "understand" human language. The approaches can be simple or not so simple, but at its heart, NLP is basically about figuring out the structure that underlies language. With the Internet acting as a repository of unlimited information, a basic task of companies such as Google is to be able wade through huge volumes of text to find content of interest. To do this efficiently, they must first throw away words that offer very little insight into what text is actually about. These words are called "stop words," and removing stop words is often the first step of NLP. Stop words are basically words that appear so frequently they don't mean much. For example, in English, some common stop words are "and", "a", "but", "the", and so on.

Your job for this assignment is to write a python program that processes The Bee Movie script (provided) to determine likely stop words. Your program should go through the file and count the number of times each word appears, taking into account case and punctuation (bee Bee and Bee. should all count as the same word). Your program should then create a file, counts.txt, that consists of each word followed by the number of times it appears in the text, from most frequent to least frequent.

Your module should implement the following functions, read_file, write_file, and build_dictionary.

- read_file: will take in a file to be read and return the contents of the file as a string
- write_file: will take in a dictionary and an output file name and write each word with it's corresponding count to the output file (in **descending** order)– this is where you will sort your dictionary (this function will not have a return statement)
- build_dictionary: will take in a string and build a word count dictionary – again, be sure to account for case and punctuation

To sort a dictionary, you can use a call from the operator module, so be sure to *import operator*. This will sort a dictionary*, a_dict*, in **ascending** value order (try it out and convince yourself it works):

```
sorted_list = sorted(a_dict.items(), key=operator.itemgetter(1))
```

Keep in mind that you should be calling functions within other functions when necessary. Your main program may not need to be more than 5 lines.

Due Date

This assignment is due at 11:59 pm on 5-8-2023. Submit via Canvas; create a zip file with all your files in it. It should be labeled firstinitiallastname_Assignment5.

Please make sure to include all the required files (README, source files).

Grading

Your program will be evaluated for correctness and elegance. In particular, you should make sure your code is properly commented and obeys standard naming conventions.