Homework 2

CS 2000: Python Programming Language

Wassnaa AL-mawee

Western Michigan University

January 26, 2018

(use chapter 12?)

In this assignment we will be counting the frequency of words in a text file. It is worth 5 points and is due **(02/05/2018)**. We will use the Wikipedia article on Character Masks: <https://en.wikipedia.org/wiki/Character_mask>

For the purposes of this assignment, please use the copy of the text file (charactermask.txt) that I have provided on E-learning.

**What is a word?**

**-** For the purposes of this assignment we will define a word as a sequence of **more than one letter** (a-zA-Z) characters separated by one or more of any other character (whitespace, non- pintables, numbers, punctuation, etc). For example, the string:

**‘simple is1902better a than !a14complex?’**

should be split into the following words:

**[‘simple’, ‘is’, ‘better’, ‘than’, ‘complex’]**

**-** Words should be case-insensitive, i.e. ‘the’, ‘The’, and ‘THE’ should all be considered the same word. The above criteria yields the following list of words for the first line in the text file provided:

(**2 points**)

[‘character’, ‘mask’, ‘german’, ‘charaktermaske’, ‘in’, ‘the’, ‘marxian’, ‘sense’, ‘is’, ‘character’, ‘disguised’, ‘with’, ‘different’, ‘character’, ‘the’, ‘term’, ‘was’, ‘used’, ‘by’, ‘karl’, ‘marx’, ‘in’, ‘various’, ‘published’, ‘writings’, ‘from’, ‘the’, ‘to’, ‘the’, ‘and’, ‘also’, ‘by’, ‘friedrich’, ‘engels’, ‘it’, ‘is’, ‘related’, ‘to’, ‘the’, ‘classical’,

‘greek’, ‘concepts’, ‘of’, ‘mimesis’, ‘imitative’, ‘representation’,

‘using’, ‘analogies’, ‘and’, ‘prosopopoeia’, ‘impersonation’, ‘or’,

‘personification’, ‘as’, ‘well’, ‘as’, ‘the’, ‘roman’, ‘concept’, ‘of’,

‘persona’, ‘but’, ‘also’, ‘differs’, ‘from’, ‘them’, ‘see’, ‘below’,

‘the’, ‘notion’, ‘of’, ‘character’, ‘masks’, ‘has’, ‘been’, ‘used’, ‘by’,

‘neo’, ‘marxist’, ‘and’, ‘non’, ‘marxist’, ‘sociologists’, ‘philosophers’,

‘and’, ‘anthropologists’, ‘to’, ‘interpret’, ‘how’, ‘people’, ‘relate’,

‘in’, ‘societies’, ‘with’, ‘complex’, ‘division’, ‘of’, ‘labour’, ‘where’,

‘people’, ‘depend’, ‘on’, ‘trade’, ‘to’, ‘meet’, ‘many’, ‘of’, ‘their’,

‘needs’]

**What to turn in:**

(Remember READABILITY COUNTS!)

**-** A **Hard Copy**: you will be turning in a print out of the following information :

* The 10 most frequent words in the document and their associated frequency (**1 point**),
* The 10 most infrequent words and their associated frequency

(**1 point**),

* The number of unique words in the document (**1 point**),
* Your code.

**-** Please format your output in the following manner

**# Most Frequent Words:**

most frequent word: frequency

2nd most frequent word: frequency

.

.

.

10th most frequent word: frequency

**# Least Frequent Words:**

most infrequent word: frequency

2nd most infrequent word: frequency

.

.

.

10th most infrequent word: frequency

**# Number of Unique Words:**

number

**-** Attach to your output a printed copy of the code you’ve written

**-** On e-Learning, submit your Python file whose name <hw#2\_LastName.py>. Please format your Python code in the following manner:

# Name: <your name here>

# Date: <#/#/#>

# Homework: <#>

# Your code