Final Submission Write-Up

1. Project Design

**State of the Union Speeches**

***Project Design***

**Main Functions:**

* **process\_file() 🡪 Open the file for readability**
* **process\_words() 🡪 Function to get a cleaned list of words from the text**
* **process\_sentences() 🡪 Function to get a cleaned list of sentences from the text**
* **avg\_word () 🡪 Function to find the average word length**
* **avg\_sentence () 🡪 Function to find the average sentence length**
* **wordCount() 🡪 Function that display the number of words and its average words**
* **sentenceCount() 🡪 Function that display the number of sentences and its average sentences from the text**
* **characterCount() 🡪 Function that display the number of characters from the text**
* **freq\_words() 🡪 Function to find the 15 most frequently used words from the text**
* **longest\_words() 🡪 Function to find the top 10 longest words from the text**
* **plot\_cloud () 🡪 Function to plot a word cloud**
* **wordCloud() 🡪 Function that creates a word cloud object and plot it by calling the plot\_cloud() function**
* **programmInfo() 🡪 Function that display programmer information**
* Print Welcome
* While user input is not ‘done’
  + Ask user to enter name of file
  + Process file function and print readability score
  + Append the user file to a list of files
  + If user inputs done, end while loop
* End While-loop
* For eachFile in list of files, call the functions for each file
  + **longest\_words( )**
  + **freq\_words( )**
  + **wordCount( )**
  + **sentenceCount( )**
  + **characterCount( )**
  + **wordCloud()**
* **ProgramInfo()**
* **End program**

2. Test Cases

Test Case 1 – WORST CASE

Enter name of files on at a time. When ready to compile, enter 'done': Biden.txt

Readibility Score: 28.994388378193918

Enter name of files on at a time. When ready to compile, enter 'done': KimKardashian.txt

File not found, try again.

Enter name of files on at a time. When ready to compile, enter 'done': done

The 10 longest words in Biden.txt

immunocompromised

hospitalizations

responsibilities

apprenticeships

totalitarianism

administration

infrastructure

revitalization

semiconductors

pharmaceutical

The 15 most frequently used words in Biden.txt

you : 75

with : 57

have : 50

more : 48

can : 44

america : 41

will : 37

they : 34

from : 33

at : 32

americans : 31

people : 30

let : 30

all : 29

my : 29

There are a total of 7680 words with about 4.565104166666667 letters per word in Biden.txt.

There are a total of 567 sentences with about 13.514991181657848 words per sentence in Biden.txt.

There are a total of 35060 characters in Biden.txt.

The word cloud of Biden.txt:

Thanks for using this Program

Author: Jose Valdivia Rojas

Course: CMSC 206

Assignment: Project 4B (State of the Union Speeches)

© 2022 JOSE VALDIVIA ALL RIGHTS RESERVED

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input Type** | **User Input** | **Expected Output** | **Actual Output** | **Pass?** |
| User input | Biden.txt | Readibility Score: 28.994388378193918 | Readibility Score: 28.994388378193918 | Yes |
| User input | KimKardashian.txt | File not found, try again. | File not found, try again. | Yes |
| User input | done | The 10 longest words in Biden.txt  immunocompromised  hospitalizations  responsibilities  apprenticeships  totalitarianism  administration  infrastructure  revitalization  semiconductors  pharmaceutical … | The 10 longest words in Biden.txt  immunocompromised  hospitalizations  responsibilities  apprenticeships  totalitarianism  administration  infrastructure  revitalization  semiconductors  pharmaceutical … | Yes |

Test Case 2: GOOD CASE

Enter name of files on at a time. When ready to compile, enter 'done': Obama.txt

Readibility Score: 4.730274884166846

Enter name of files on at a time. When ready to compile, enter 'done': Trump.txt

Readibility Score: 24.056162857963173

Enter name of files on at a time. When ready to compile, enter 'done': done

The 10 longest words in Obama.txt

transportation

transformation

representative

personalizing

extraordinary

breakthroughs

manufacturing

international

congressional

understanding

The 15 most frequently used words in Obama.txt

have : 37

america : 32

you : 30

people : 26

they : 26

with : 26

now : 25

will : 25

can : 25

just : 24

when : 24

work : 22

all : 22

world : 22

american : 20

There are a total of 5652 words with about 4.437367303609342 letters per word in Obama.txt.

There are a total of 374 sentences with about 16.106951871657753 words per sentence in Obama.txt.

There are a total of 25080 characters in Obama.txt.

The word cloud of Obama.txt:

The 10 longest words in Trump.txt

administrations

congratulations

administration

groundbreaking

pharmaceutical

infrastructure

constitutional

servicemembers

extraordinary

neighborhoods

The 15 most frequently used words in Trump.txt

you : 73

have : 51

with : 40

america : 38

american : 36

very : 34

will : 34

thank : 33

all : 29

has : 28

my : 27

was : 27

new : 24

at : 23

than : 23

There are a total of 6217 words with about 4.808267653208943 letters per word in Trump.txt.

There are a total of 440 sentences with about 14.263636363636364 words per sentence in Trump.txt.

There are a total of 29893 characters in Trump.txt.

The word cloud of Trump.txt:

Thanks for using this Program

Author: Jose Valdivia Rojas

Course: CMSC 206

Assignment: Project 4B (State of the Union Speeches)

© 2022 JOSE VALDIVIA ALL RIGHTS RESERVED

Test Case 3: GOOD CASE

Text

Description automatically generated

3. Learning Experience

The project 4 was a challenging assignment but helpful to master Regular expression. The assignment was about to process different text files, which contains speeches from different authors such as Trump, Biden, John F Kennedy, Obama. Then, process each file to obtain relevant information from these speeches like total word count, average word length, average sentence length, most frequent used words, top 10 longest words, and a word cloud for each speech text. I was able to code this program by using regular expression to filter the texts. First, I coded the menu which was the easier part from this assignment since I only had to code a while statement to keep processing texts according to the user inputs. On the other hand, the tedious part was coding the functions to make the program work well. So, I started building the skeleton of this assignment, which is to define the functions that process each text and covert them in lists of sentences and lists of words. While I was doing these functions, I learned the process and coded the algorithm for formatting words and sentences, and mastering lists.

Moreover, I also learned how to create a word cloud image on Python. By importing wordcloud library, I was able to create a wordcloud object to plot the image with words in it. To enhance this part of the program, I decided that for each text speech, it will have their own image according to the person who does the speech. For example, for the Biden speech, the word cloud will have Biden as background in the word cloud.

At the end, the program worked successfully and had not complications with user-program interaction. Each possible incorrect input was acknowledged so that the user re-enters another answer. Furthermore, Overall, it was a fun program to code, I am happy with the outcome of this assignment, and I am ready for the next challenge coding program and projects.

4. Assumptions

* The user will choose a file for displaying it. If the user enters a file that is not found. The program will ask the user to re-enter another file.
* Quotes sentences may end a sentence in ( “ ) instead of a period ( . )
* Words may be compound or joined by a dash ( - ) or double dash ( -- )
* Words may be compound by (‘ll) or (‘s) or (‘t)
* Words may have special characters (()\"#$/@;:<>`+=~)
* All of these assumptions were addressed in the program