Dennis Guevarra
CS256
Lab Assignment <Number here>
11/12/2023

Lab Assignment 6 Design Document

Introduction

The objective of this program is to analyze the text of the book "The Phantom Tollbooth" by performing the following tasks:

1. Convert the text to lowercase.
2. Remove punctuation and special characters.
3. Split the text into words.
4. Remove common articles, prepositions, conjunctions, and pronouns (stop words).
5. Count the occurrences of each remaining word.
6. Display the top 50 words along with their frequencies.

Functional Requirements *(starting with lab 3)*

The program makes use of libraries to simplify the task needed to be done by allowing users to implement code without the hassle of hardcoding everything.

```
import phantom_tollbooth
import re
from collections import Counter
```

```
book = phantom_tollbooth.get_text()
```

#pulls text from the phantom tollbooth book

```
book = book.lower()
```

#converts the book's text to lowercase

```
book = re.sub(r'\W+', ' ', book)
```

#removes punctuations and special characters such as apostrophes

```
words = book.split()
```

#Split the text into singular words

```
stop_words = {…}
words = [word for word in words if word not in stop_words]
```

Remove articles, prepositions, conjunctions, and pronouns from the words list

```
word_counts = Counter(words)
```

#counts the words and each time it the words occurs which then creates a dictionary to those words into

Design Requirements *(With Libraries)*

- Make use of Libraries
- Use the counter function to count the words
- Use the re.sub to remove special characters
- Use stopwords to remove unnecessary words
- Use the sort() function to sort from highest to lowest

Design Requirements (*No Libraries)*

- Make use of for loops
- Use if-else branches
    - If the word is in the excluded word set, it does not execute
    - If it isn't, executes and looks to see if the word was counted and adds a counter

Testing Predictions Results *(With Libraries)*

The re.sub() function created words that were only letters which added on extra code to remove them from the program. The output of several test showed multiple singular and double letters such as s, t, and ll, which were all separated and turned into words due to the re.sub() function.

Testing Predictions Results (*NO Libraries)*

The coding was similar to the program with libraries except this program makes use of nested if-else statements which checks for the words that appear in the string. The excluded_words set is similar to the stopwords set in the previous program in which I had to repeatedly test for various words until all unnecessary words stopped appearing.

Reflection and Questions

There was a lot of overthinking in this lab which made me look to the internet rather than carefully understand the assignment. My assumptions made me think that there was no other way to do this assignment without libraries and the help of the internet, but it was actually simple. After coding with the use of libraries, I realized that with loops and if-else statements, the outcome would be the same.