Andrew Coalwell

CS256

Lab Assignment 6

11/10/2024

Lab Assignment 6 Design Document

*This program is being made to count all of the words in Phantom Tollbooth and print the 50 most common words excluding*

*1.Process*

*The program will use get\_text() function from the module to get the full text from the Phantom tollbooth.*

*2.The program will convert text to lowercase to make sure that capitalization won’t mess up the word frequencies*

*3.It will remove all punctuations but leave word spaces.*

*Words it will not count will be pronouns and filler words (the, and, but, like)*

*4.Counting*

*The program will iterate through each word in the book and count each unique word*

*5. It will store the word in a dictionary*

*6. Then it will number each time the word appears.*

*7. What will happen with each time a word is found*

*8. Before counting the word the program will check if it's on the don't count list.*

*9. If it is a word that is not counted it will be skipped then it will go to the next word*

*10. If the word is on the list it will be add to the dictionary*

*Outputs*

*11.The program will sort the dictionary by word frequencies in descending order.*

*12. Then the top 50 will appear then with their frequencies*

*Text filtering*

1. *Purpose loop through words then filter through the words then increase the counter*
2. *Exit after looking through all the words.*
3. *Sort the 50 most common words.*
4. *Testing*
5. *I will test this by testing functions by themselves*
6. *I will compare my output to the output on tagcloud*
7. *Then making sure the output order is most to least.*
8. *Test with small groups of text*
9. *Test and make sure pronouns and other words dont break it.*
10. *Test the case sensitivity of the program.*
11. *Test on a larger scale*
12. *My guess for the most common word would be said or some other word of similar meaning.*

**Reflection and Questions (Lab 6 and on)**

*Reflect about your experience designing this program. Be sure to answer all the following questions:*

1. *I started with breaking down the program to input process output. I identified each part then made sure to exclude pronouns and other words like that.*
2. *The first thing I found out was that I needed to make sure capitals don’t have there own section so I had to filter those out and make sure they all go into one section.*
3. *I wasn’t sure how specific I needed to be with what words to count and which to overlook and how specific I needed to be with capitals.*
4. *I don’t have any further questions.*
5. *On my second attempt I checked each section to make sure it worked as I was putting it in so it would be easier than looking through it all at the end and not where the problem could be.*
6. *I believe I’ve done everything that was asked of me so I believe I got 100%*
7. *I could have been more thorough with my coding making it better and more efficient but it works as is. I feel pretty confident with my word filtering process.*