Guillermina Morales
CS256
Lab Assignment 6
2/25/2024

Lab Assignment 6 Design Document

Introduction:

The program, tentatively named "Phantom Tollbooth Word Counter", aims to analyze the text of "The Phantom Tollbooth" to determine the fifty most common words, excluding articles, prepositions, conjunctions, and pronouns. The input to the program is the text of the book obtained from the get_text() function in the phantom_tollbooth module. The program will process the text by counting word frequencies while excluding the specified words. The output of the program will be a report of the fifty most common words along with their occurrences.

Functional Requirements: Input: The text of "The Phantom Tollbooth" obtained from the get_text() function. Counting word frequencies. Excluding articles, prepositions, conjunctions, and pronouns. Output: Report the fifty most common words and their occurrences.

Data Structures:

text: String - to store the text of the book.

word_counts: Dictionary - to store word frequencies.

excluded_words: List - to store articles, prepositions, conjunctions, and pronouns.

Loops: Behavior: Counting word frequencies. Loop Type: for loop - to iterate through each word in the text. Exit Condition: When all words have been processed.

Conditions: Conditions Checked: Whether the word is in the list of excluded words. Whether the word is already in the word_counts dictionary. Resulting Behavior: If the word is excluded, it will not be counted. If the word is not in word_counts, it will be added with a count of 1. If it's already there, its count will be incremented.

Flow of Control Diagram:

Output Format: Print the fifty most common words along with their occurrences.

Functions:

remove_stop_words(text): Remove stop words from the text.

count_words(text): Count word frequencies in the text.

Comments:

Include comments to explain major sections of the code, data structures, and conditions.

Adhere to Python's style guide for variable and function names, line lengths, and empty lines.

Testing Predictions Results: Variations of Input:

Test with the text of "The Phantom Tollbooth" to ensure correct functionality.

Test with different books to check if the program handles various texts correctly. Predictions:

Guillermina Morales
CS256
Lab Assignment 6
2/25/2024

Expect the program to correctly count word frequencies while excluding stop words. Actual Results:

Verify that the program accurately identifies the fifty most common words and their occurrences.

Reflection and Questions

My initial decomposition strategy was to first understand the input, then identify the processing steps required, and finally determine the output format.

The first thing I figured out was the need to remove stop words from the text before counting word frequencies.

During the design process, I uncovered questions regarding the most efficient way to handle stop words and the best approach to sorting and reporting the most common words.

I was able to answer most of my questions through research and experimentation, but I still have some questions regarding the optimization of certain processes.

I made a second attempt to decompose the problem by focusing more on the specific requirements and breaking down the tasks into smaller, more manageable steps.

I believe I achieved around 90% completion in terms of design, with minor adjustments required during implementation.

I feel most confident about understanding the input and output requirements, but I still need to refine my understanding of the most efficient algorithms for word counting and stop word removal.