Kaitlen Birkner
CS256
Lab Assignment 6
2/25/24

Lab 6 Assignment Design Document

Introduction

The program, named "PhantomTollboothWordCounter", takes the text of the book "The Phantom Tollbooth" as input, processes it to count the frequency of each word, and outputs the top 50 most frequently occurring words. The text is cleaned to remove punctuation, numbers, and converted to lowercase. The words are then counted and sorted based on their frequency.

Functional Requirements

- **The program will obtain the text of the book using the phantom_tollbooth.get_text() function.**
- **The program will clean the text by removing punctuation, numbers, and converting all text to lowercase.**
- **The program will split the cleaned text into words.**
- **The program will count the frequency of each word using a dictionary.**
- **The program will sort the dictionary items by frequency in descending order.**
- **The program will print the top 50 most frequent words to the console.**

Design Requirements

- **The program will use a string variable book to hold the text of the book.**
- **The replace() function will be used to clean the text.**
- **The split() function will be used to split the cleaned text into words.**
- **A dictionary named words will be used to count the frequency of each word.**
- **The sorted() function will be used to sort the dictionary items by frequency.**

Kaitlen Birkner
CS256
Lab Assignment 6
2/25/24

- **The program will use a for loop to iterate over the words and count their frequency.**
- **The program will use an if-else statement to check if a word is already in the dictionary.**
- **The program will use the print() function to print the top 50 most frequent words.**

Version 1: Basic Word Counter

This version of the program will simply count the frequency of each word in the text without any text cleaning or sorting.

What to look for:

- **The program correctly splits the text into words.**
- **The program correctly counts the frequency of each word.**
- **The program handles edge cases correctly, such as multiple spaces between words.**

Version 2: Advanced Word Counter

This version of the program will include text cleaning by checking if isalpha to remove punctuation and numbers, convert to lowercase, and sorting to print the top 50 most frequent words.

What to look for:

- **The program correctly cleans the text by removing punctuation and numbers, and converting it to lowercase.**
- **The program correctly counts the frequency of each cleaned word.**

Kaitlen Birkner
CS256
Lab Assignment 6
2/25/24

- **The program correctly sorts the words by frequency and prints the top 50 words.**

Testing Predictions Results

- **I predict that the program will correctly count the frequency of each word and print the top 50 most frequent words.**
- **The actual results will be compared with a manually verified list of the top 50 most frequent words in "The Phantom Tollbooth".**