Will Childs
CS256
Lab Assignment 6
11 10 2024

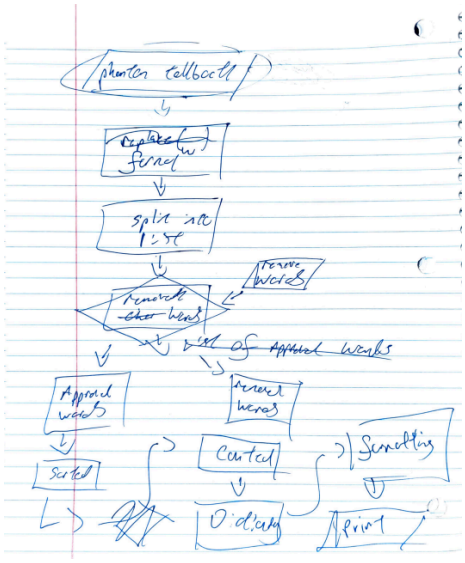
Lab Assignment 6 Design Document

Introduction:
Lab 6 provides a string of the book, *The Phantom Tollbooth*, and we must transmit that string into a format that shows the book's 50 most popular words. This program uses the split command to make the book into a list of words. Then that list is crossed with a few lists to remove unwanted words. That list is formed into a dictionary with the values being the frequency of the word. The dictionary is then formatted to display the 50 most popular words.

Functional Requirements:

1. The only input is the text for *The Phantom Tollbooth*, which is provided via import.
2. The program manipulates the text as described in the introduction.
3. The program prints a formatted version of the dictionary which shows the 50 most popular words and their frequencies.

Design Requirements:

1. Variables:
    a. book -> string of *The Phantom Tollbooth* that becomes edited
    b. book_list -> list made from splitting book with spaces
    c. articles, prepositions, conjunctions, pronouns, extra -> list of words to remove from book_list
    d. filtered_words -> list of words that were not in articles, prepositions, conjunctions, pronouns, extra but were in book_list
    e. sorted_words -> list of filtered words using sorted command
    f. word_count -> dictionary of sorted_words as keys and frequency used as value
    g. sorted_word_dict -> word_dict sorted via frequency to select 50 highest frequency words
    h. sorted_50_words -> dictionary of sorted_word_dict with word (keys) sorted a-z
    i. keys, values -> words and frequencies from sorted_50_words for printing formatting
2. Operators, functions:
    a. get_text -> takes string from phantom_tollbooth import
    b. replace -> replaces grammatical characters that are unwanted with a space
    c. lower -> converts all characters to lowercase
    d. split -> converts string to list with spaces removed
    e. append -> adds approved words to filtered_words
    f. sorted -> reorganizes list, dict a-z
    g. items -> reorganizes dictionary to output keys and values
3. Data structure:
    a. phantom_tollbooth.py is our imported text
4. Coding concepts:
    a. Strings, list, dictionary, branching
5. Conditions my program needs to handle
    a. Needs to input a string, should work as long as the input is a string ?
6. 

7. Not relevant
8. Output:

    a. Prints keys and values of dictionary
9. Not relevant

Testing Predictions Results:

1. One data set for input
2. Tests were continued to remove words that TagCloud disapproved of
3. __

Reflection and Questions (Lab 6 and on):

1. I had a general idea of the process and worked from the top down, fixing problems as they arrived
2. __
3. How to: use split, sorted, reverse, and functions with dictionaries
4. __
5. __
6. My design looks choppy and should be reformatted and made more concise, possible use of functions, but this is a 4 credit class and I need to be weary of the time allotted. It functions properly, it is logical, but it could be cleaner. 70%.
7. I understand everything well aside from the dictionary sorting

Collaboration ***(starting with lab 2– optional)***

I chatted with Darian about a few roadblocks, and I used copilot to handle possible errors and to write lists of words that are not wanted in my word cloud. I did not copy and paste any code; all code is mine.