

# Lab Assignment 6 Design Document

## **Introduction (1-5 sentences, starting with lab 2)**

The program, "Word Frequency," analyzes any text set (The Phantom Tollbooth in this case) and outputs the fifty most common words. It will exclude all stop words such as conjunctions, pronouns, prepositions, and articles. The program will separate each word in the text, count the number of occurrences of each unique word, filter out the excluded words, and store them in a data set. The output will have each word and their frequency starting from most frequent to least.

## **Functional Requirements (starting with lab 3)**

- The program will take input text from the 'get\_text()' function in the 'phantom\_tollbooth.py' file.
- It will split and store each unique word in the text.
- Stop words will be excluded from the data set.
- The program will find the fifty most common words based on the highest frequencies.
- Will generate and print out each word and its frequency.

## **Design Requirements (starting with lab 4)**

Data:

- Input text from the 'get\_text()' function
- Counts of words/their frequencies (dictionary)

Loops:

- Iterate over each word in the text
- Counting words and update in the dictionary
- Filters to exclude/remove stop words

Conditions:

- Checking if a word is meant to be excluded
- Checking for words similar to stop words (there, they're, their, there's, etc.)
- Checking for errors

## **Testing Predictions Results (starting with Lab 6 and on)**

Tested inputs:

- Input text without excluded words.
- Input text with excluded words.

Predictions:

- Should count and report the fifty most common words in the text.
- Stop words should not appear in the output.

Actual results:

- To be determined.