

## Overview

This project was created to count unique terms in text. The user can choose a text file to read or enter their own text by typing or copy/paste. The user can opt to find specific terms or exclude specific terms instead of counting all terms, but the option to count all terms is the default. The user can utilize the program on its own, or they can download list output to a spreadsheet.

### Currently Implemented Planning Specs.

Data input: the interface will allow users to enter data in multiple ways. Users will be able to:

- Enter their own text (i.e. typing) into the text field
- Copy/paste text into the upper text field
- Browse a local txt or pdf to read

Data manipulation: the interface will allow users to specify very limited output preferences. Users will be able to:

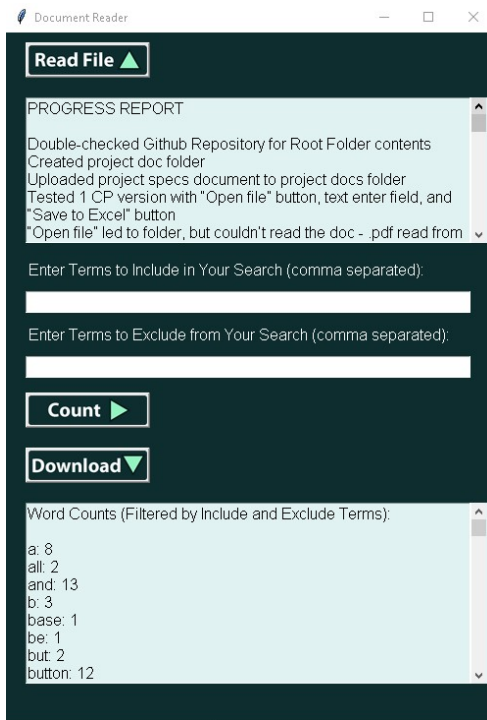
- Choose a specific term or terms to read, count and display in the output field and/or download
- Choose which terms should not be included in the counts or output field and/or download
- Choose to read and count all of the terms in a pdf and display all terms and counts in the output field and/or download

Data output: the interface will display the terms and counts of each term in the output field and also give the user the option to download as well as copy/paste, the data shown in the output display,

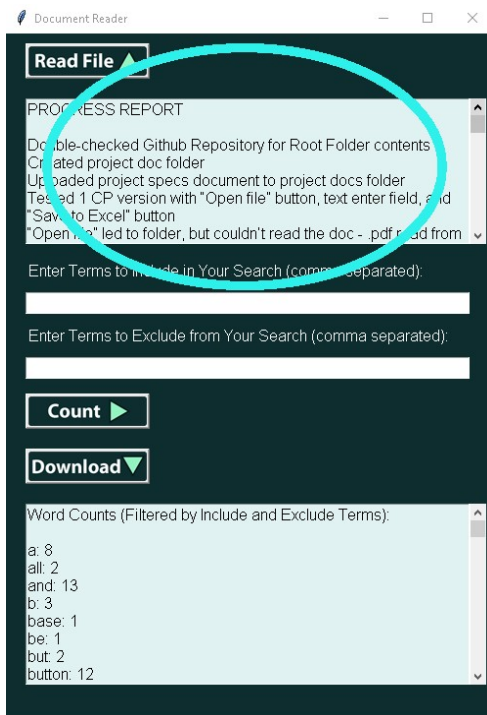
# DEVELOPER'S DOCUMENTATION

## Using the program

- Open the program

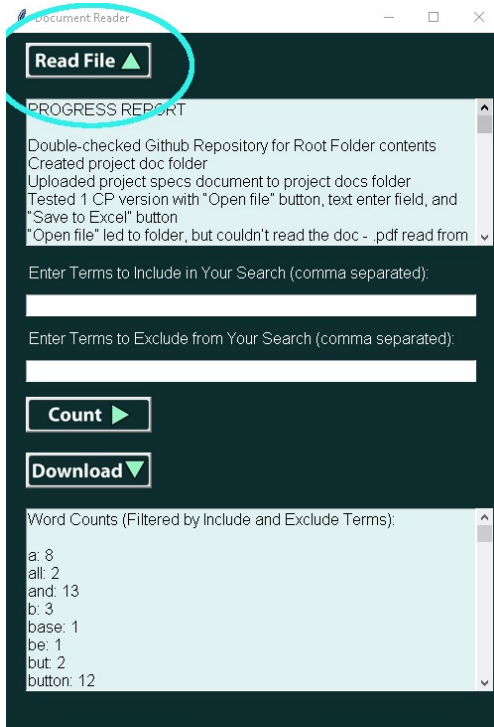


- If you want to use manual text, you can either type into the text input field or copy/paste your text into the text input field.

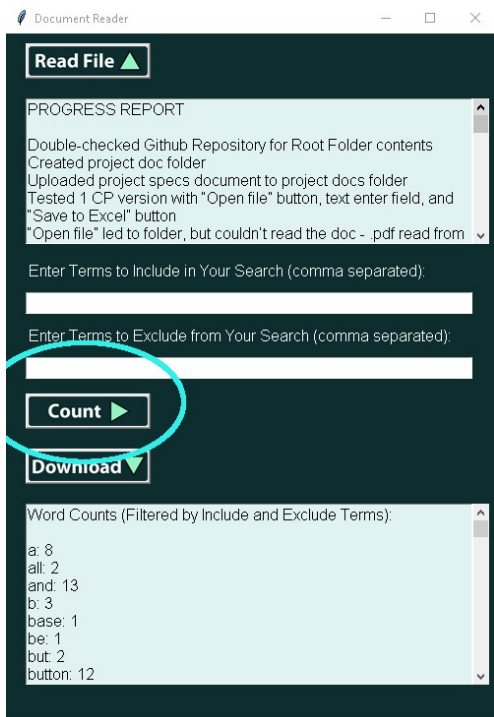


## DEVELOPER'S DOCUMENTATION

- Alternatively, you can click the "Read File" button to load a document with text from your computer.



- If you want a simple count of your manual text or opened document, click the "Count" button to see a list of unique terms from the text input field.



## DEVELOPER'S DOCUMENTATION

- If you want to search for specific terms, enter those terms in the search field for included terms and click the "Count" button to see a list of only the specified terms and their counts.

Document Reader

**Read File ▲**

PROGRESS REPORT

Double-checked Github Repository for Root Folder contents  
Created project doc folder  
Uploaded project specs document to project docs folder  
Tested 1 CP version with "Open file" button, text enter field, and "Save to Excel" button  
"Open file" led to folder, but couldn't read the doc - pdf read from

Enter Terms to Include in Your Search (comma separated):

Enter Terms to Exclude from Your Search (comma separated):

**Count ▶**

**Download ▼**

Word Counts (Filtered by Include and Exclude Terms):

a: 8  
all: 2  
and: 13  
b: 3  
base: 1  
be: 1  
but: 2  
button: 12

- If you want to exclude specific terms from your search (e.g., the, and, etc.), enter those terms in the search field for excluded terms and click the "Count" button to see a list of unique terms and their counts that don't include the terms you specified.

Document Reader

**Read File ▲**

PROGRESS REPORT

Double-checked Github Repository for Root Folder contents  
Created project doc folder  
Uploaded project specs document to project docs folder  
Tested 1 CP version with "Open file" button, text enter field, and "Save to Excel" button  
"Open file" led to folder, but couldn't read the doc - pdf read from

Enter Terms to Include in Your Search (comma separated):

Enter Terms to Exclude from Your Search (comma separated):

**Count ▶**

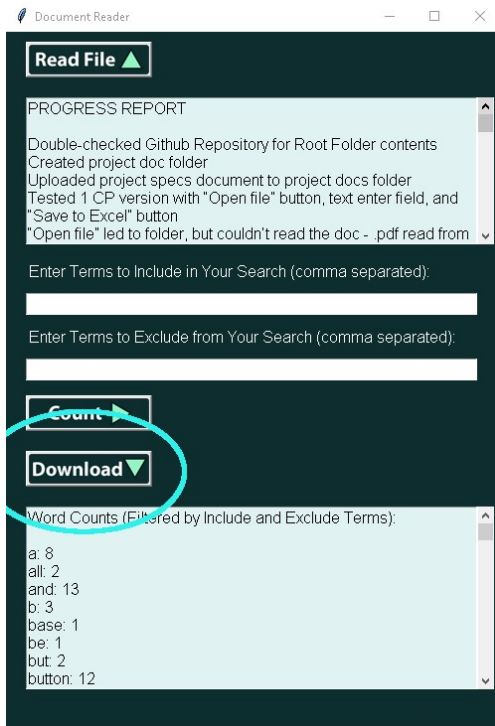
**Download ▼**

Word Counts (Filtered by Include and Exclude Terms):

a: 8  
all: 2  
and: 13  
b: 3  
base: 1  
be: 1  
but: 2  
button: 12

## DEVELOPER'S DOCUMENTATION

- You can copy/paste out of the text output field, or you can click the "Download" button to export the list of terms and counts to a spreadsheet and save it on your computer.



## CODE FLOW

### **Classes:**

DocumentReader : Combined functions for the Document Reader program

### **Functions:**

**`__init__`** : Used to assign values to object properties

**`create_widgets`** : Defines and assigns values to the Graphical User Interface (GUI)

**`open_file`**: Opens documents to be read by the Document Reader

**`count_words`**: Counts unique terms in text based on the user's options. May include or exclude terms. This function also converts words to lowercase and removes symbols to ensure singularly unique terms

**`display_word_count`**: Displays words and their counts in a small table in the output field. This function also sorts the word list in alphabetical order for easier output reading

**`export_to_excel`**: Downloads to a spreadsheet on the user's local machine

### **Possible future revisions or add-ons**

- Expand document types that can be read
- Ability to enter a URL for reading
- Possibly pytesseract and pillow (for image-based or scanned .pdfs)
- Possible APIs in the future – external mechanisms. Currently none exist for syllabi
- Web interface