

Bible.com Scraping Project

App type:

The project I chose is the build your own project. I will be creating a program that includes a GUI to web scrape bible.com

Description:

The user will be able to select from available bible versions via the GUI and get statistics about:

- The book they specified
- The version they specified
- The chapter/verse they specified

The GUI would have buttons that would specify the version the user wanted to use; the buttons would then be connected to the python requests module to scrape the selected webpage on bible.com.

Some statistics they could request would be word count, chapter count, and word frequency. They could toggle these options on and off through either buttons or a drop-down menu on the GUI.

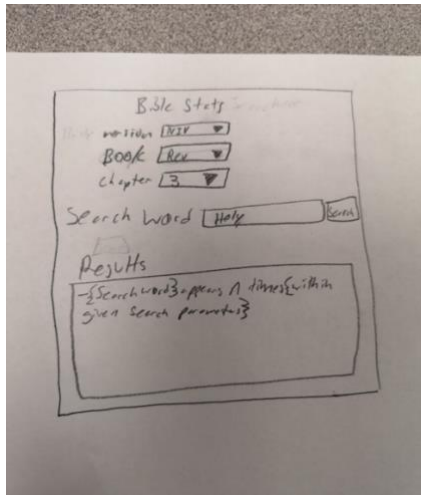
Tools:

I will be using python 3.9 for the project. I will be using the requests module in python to get the raw html from bible.com and the BeautifulSoup 4 module to parse the html from the requests module. I will of course also be using bible.com for the text information.

Feature List:

- Change between bible versions via a button or dropdown
- Search for word frequency in given version (may be limited to book/chapter if looping over the whole bible is too fast for bible.com)
- Ability to specify the version to search through
- Ability to specify the book, but not required to. If not selected entire version would be searched for search word
- Ability to specify the chapter to search through, if none selected, entire book would be searched
- The GUI displays the number of times the given search word is found within the given search parameters
- Allow user to search for any word case insensitive

UI sketch:



Architecture:

The GUI will have buttons that allow the user to choose their search parameters. These buttons will be used as input to decide what the requests module should search for. This is all contained within one class. One method inside the class will be using the requests module to get the raw html from bible.com. Another method will be taking the data from requests and using beautiful soup to get the useful information from the html code string. Another method will be used to display the stats I get after searching for the word within the user's specified book and chapter.

Minimum Viable Product:

The minimum would be a functioning GUI with connected buttons/drop-downs that connect to methods that gets the html using the requests module. And then loops over the string to return the number of search word matches.

Interesting Challenges:

I think the main thing I am worried about in this project is learning to use the web scraping tools available to get the bible text. I am also not entirely certain how I will be using requests to allow the user to customize the search URL. Maybe an f-string with variables connected to the drop-down options on the GUI?

Stretch:

- Add ability to compare versions/books/chapters for the frequency of search word
- Add a histogram to the GUI displaying in a visual manner the frequency of the word searched
- Have different languages of the bible available