Web Crawler:

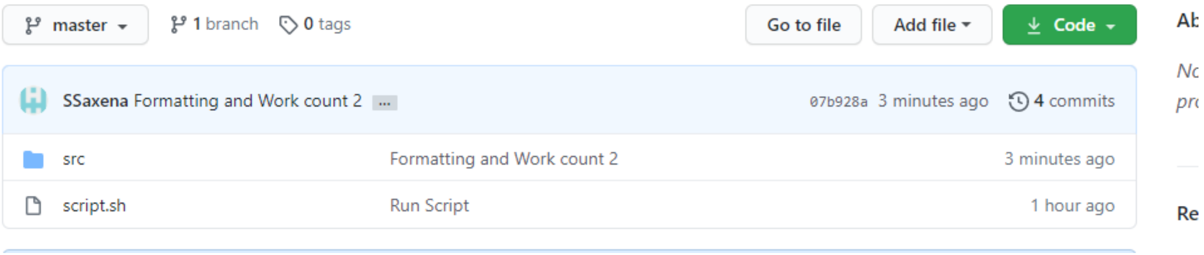
The script/app is developed for fetching the HTML pages from the targeted URL provided and process it to get the desired results.

For Web Crawling: we need to give limit the crawling depth to 4.

For Web Scrapping / HTML Processing:

- Need to fetch the top 10 words on each HTML page
- Need to fetch the top 10 combination of words together in the HTML Page

Repository has 2 parts:



SSaxena Formatting and Work count 2 ...	07b928a 3 minutes ago	4 commits
src	Formatting and Work count 2	3 minutes ago
script.sh	Run Script	1 hour ago

1. Script.sh: The script has following installation paths
   a. Installing the Python 3
   b. Installing the pip3 for installing the python modules
   c. Execution of the Scripts.
2. src folder with the source code. It has 3 files
   a. File to Start the execution
   b. Class to get the HTML page by web crawling
   c. Class to process each of the HTML and storing the records to CSV file
3. We can execute by

   Python3 Start.py

   This will execute with default parameters i.e.URL https://www.314e.com/ and Max Depth : 4

   If we need to execute by passing the parameters, then we can execute by

   Python3 Start.py https://www.314e.com/ 4

Few things which can be updated are:

- Fetching the URL HTMLs by using async await operation & creating tasks
- Naming convention can be improved
- Logging
- When writing to CSV or to DB, I could have used Model classes to have a uniform layer.