

Web Crawler Assignment – Time Line Schedule

Name: Vedant Kesarinath Gharat

College: Vidyalkar Institute of Technology, Mumbai

GitHub repository link:

Start of the project: 9:30 AM, February 28, 2021

- Started to scrape Amazon's Website but the access was blocked for scraping thus had to use User Agents to avoid getting blocked from amazon.
- Started to manually parse the code and implement it using beautiful soup.
- Since amazon is a complex dynamic website it took so many trial and errors to extract data from the website

Started to work on flipkart.com

- Tried to Work on flipkart.com but couldn't complete the entire part.

Instructions to run the code from GitHub repository:

- Take the file amazon.py and run in on your machine.
- Once executed the program will ask for a keyword to search on AMAZON's website. Enter the keyword.
- The program will take some time, once the execution is completed a excel sheet will be created will the data scarped from amazon.com.
- The excel sheet will be created in in same location where amzon.py is downloaded.