Muhammad Shoaib Amin

Data Engineer

R-1333/2, F.B. area,
Karachi,
(+92) 331-2111-050
shoaibsivany@gmail.com

TOOLS

Beautiful Soup, Scrapy, Selenium, Puppeteer, Fetch, Requests, APIs, Multiprocessing, Scheduling

WORKS

ETL of Ecommerce Marketplaces. *- Python - Selenium*

- Bot that scrapes data from 20 Giant Ecommerce Stores.
- Scraping over 55 Products Daily.
- Scraping 50,000 - 100,000 Records Daily
- XAMPP Server, Advanced Queries, Multiprocessing, Operating Systems, Crontab, and Task Scheduler.

Pakistan's largest Ecommerce site Bot *- Python - Selenium*

- A Python bot that extracts the products information of the user-defined category.
- This program extracts the products information along with description, price, comments, and ratings.
- It also scrapes pictures of e-commerce products and stores them as files with relevant names.
- The entire scraped data is output in CSV, JSON, or SQLite format.
- Includes pagination.

Grocery Scraper *- Python - Selenium*

- Scrapes grocery items in multiple sites.
- The scraped data contains names, prices, descriptions, and ratings.
- Also, the scraper downloads the images of groceries for quality assurance.
- The entire scraped data is output in CSV, or SQLite format.
- Includes pagination.

Linkedin Bot *- Javascript ES6 - Puppeteer*

- This bot render's LinkedIn login page, enters username and password, and hit submit
- Then it scrapes Linkedin's user's information such as Name, Tagline, Description, Education, and Experience.

Pakistan's Largest Urdu News Scraper *- Python - Scrapy*

- Fetch news along with the featured content and complete article.
- The scraped data contains headlines, articles, featured content, and ratings.
- It has the tendency to get all news from the year 2020.
- The entire scraped data is output in CSV, or SQLite format.
- Includes pagination.

Facebook Posts Scraper *- Python - Scrapy*

- Fetch Facebook posts along with the post text and comments.
- It has the tendency to get all comments of the post along with a number of replies and likes.
- The entire scraped data is output in CSV, and SQLite format.
- Includes comments pagination.

Twitter Bot *- Python - Scrapy*

- The bot logs into Twitter by navigating to the login page from the main Twitter site.
- It inserts log in fields and submit login.

Article Rewriting Application *- Python - Beautiful Soup*

See output at
https://docs.google.com/document/d/1jtOcrUfN51lwKRI0k2VOVQfejXLm8KtQYChWW
QhjidA/edit?usp=sharing
AUGUST 2020 - SEPTEMBER 2020

- This is a python command-line application that generates articles based on the unique articles published on real websites.
- The application takes a Google search query as input and automatically hits www.google.com with the search query and it scraps all the search results links of the recent articles. Then it iterates to every link and scraps all the content inside that website link.
- Which means it gets all recent results of the search query (the recent articles in this case)

- The application itself identifies the introduction, body, and conclusion of all articles with the help of article text length.
- Then by using the concept of Natural Language Processing, which is tokenization, context-free grammar, and POS tagging, it generates a rough draft of an article with an introduction, body, and conclusion.

Dawn News Headline Fetcher - *Python - Beautiful Soup*

See output at
https://docs.google.com/document/d/15D18SRsI-BQc2BCXrgmKyGOKCPKQoQxFLDSC-izcuDc/edit?usp=sharing

MARCH 2020

- This application fetches the headlines of Dawn News.
- It takes categories as input such as sports, education, politics, etc.
- Based on the input, it creates the URL that hits the specific category and fetches all headlines with the published date, and generates a dynamic CSV file.

Twitter Scraping Bot - *Javascript ES6 - Puppeteer*

Code and output images will be provided upon request

JANUARY 2020 - FEBRUARY 2020

- This application fetches the last 10 tweets and keeps the log of all the followers of a particular user of Twitter.
- It hits the URL every 5-7 seconds and gets the followers of the user.
- Then it generates a dynamic line graph that specifies the time and no. of followers of the Twitter user.

EDUCATION

NED University, Karachi- *Master in Data Science*

MAY 2021 - PRESENT

Sindh Madrassatul Islam University, Karachi- *Bachelor in Computer Science*

JANUARY 2016- SEPTEMBER 2020

Secured a good score in my 4-years bachelor's degree program.