Marwadi University
Marwadi Chandarana Group

MARWADI EDUCATION FOUNDATION'S GROUP OF INSTITUTIONS
GAURIDAD CAMPUS, RAJKOT
DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

Web Scrapping And Data Analytics

Date: 28-09-2024

Event Report

WEB SCRAPPING AND DATA ANALYTICS was an event conducted by Data Science Cub at MA112 at 10:00 AM. The event was mainly focused on what is web scrapping and how it is done on different platforms. Flow of the event was:

1. About DS club:

The session starts with the introduction of DS Club, developed by students of ICT department and guided by Professor Nishith Kotak of ICT department, aims to explore vastness of Data Science and Data visualization. The club also aims to teach their members that how data can be used to find conclusions and make correct decision after correct analysis of data.

2. What is Data Science?

Data science is a field that uses data to create insights and strategies for businesses and industries. It's a multidisciplinary field that combines principles and practices from many other areas, including: mathematics, statistics, business, artificial intelligence, and computer engineering.

3. What is Web Scrapping?

Web Scrapping is simply a process of extracting the data from websites. Python, JavaScript, PHP, C#,etc. languages are used for Web Scrapping. Python is widely used for web scrapping due its libraries Like BeautifulSoup, Scrapy, and Selenium.

4. Google Colab

Colab notebooks allow you to combine executable code and rich text in a single document, along with images, HTML, LaTeX and more, also it has many libraries already install. So we are using google colab rather than vs code.

The process of web scraping in Python involves the following steps:

1. Set up your environment: Install Python and the BeautifulSoup and requests libraries. You can also create a virtual environment to keep your Python environment organized.
2. Understand the website: Identify the website you want to scrape and understand its structure.
3. Fetch the web page: Use the requests library to fetch the content of the web page.
4. Parse the HTML: Use BeautifulSoup to parse the HTML and navigate the document tree.
5. Extract data: Use BeautifulSoup methods to find and extract data.
6. Handle pagination: If the data is split across multiple pages, you need to handle pagination by iterating over the pages.
7. Store the data: Store the extracted data in various formats, such as CSV or JSON.
8. Respect the website's guidelines: Follow the website's robots.txt and legal guidelines.

Web scraping is the process of automatically gathering data from websites. It can be used for a variety of purposes, such as collecting product prices, gathering research data, or monitoring social media trends.

5. Web Scrapping using BeautifulSoup

First of all import the requests library. Then, specify the URL of the webpage you want to scrape. Send a HTTP request to the specified URL and save the response from server in a response object called r. Now, as print r.content to get the raw HTML content of the webpage. It is of 'string' type.

Step by Step process:

1. Set Up Google Colab:

- ✓ Create a new notebook in Google Colab.
- ✓ Install the necessary libraries:

2. Import Required Libraries:

3. Specify the Target URL:

4. Send a GET Request:

5. Check for Successful Response:

6. Parse the HTML Content:

7. Find Specific Elements:

Use soup.find() or soup.find_all() to locate elements based on tags, attributes, or text content.

8. Extract Data:

Extract the desired data from the found elements using appropriate methods:

- ✓ *element.text for text content*
- ✓ *element.get('href') for link attributes*
- ✓ *element.attrs for all attributes*
- ✓ *element.find() for nested element*

9. Process and Store Data:

- ✓ *Process the extracted data as needed (e.g., clean, format, or transform).*
- ✓ *Store the data in a suitable format (e.g., CSV, JSON, database).*

6. Web scrapping using Selenium

It is the advance version of BeautifulSoup. It works similarly like BeautifulSoup but there are some steps where it differs.

The different steps are as follows:

3. Create a WebDriver Instance:

4. Navigate to the Target URL:

5. Wait for Elements to Load (Optional):

- ✓ Use WebDriverWait and expected_conditions to wait for specific elements to appear or become visible before interacting with them.

6. Interact with Elements (Optional):

- ✓ Use driver.find_element_by_* or driver.find_elements_by_* methods to locate elements by different criteria (e.g., ID, class name, XPath).
- ✓ Perform actions on elements, such as clicking, sending text, or submitting forms.

List of Participants:

GR NO	EN Roll No	Participate Name
117715	92200133008	ISHIKA KIRIT SHETH
119732	92200133043	PRIYANSHI HIREN MADANI
119487	92200133027	RISHIT KAMLESHBHAI RATHOD
119519	92200133041	DHRUVI BHALODIYA
119144	92200133016	MALHARKRISHNA SAHILKRISHNA SHAH
119680	92200133035	KRISHNA HARESHBHAI DIXIT
115019	92100133014	AJAY DINESHBHAI RANK
112309	92100133007	ANIKET PARESHBHAI SOLANKI
113640	92100133021	JATAN ASHISHBHAI SANGHVI
114699	92100133071	VENISHA DINESHBHAI PARMAR
113367	92100133016	VIVEK SUNILBHAI PATEL
114319	92100133037	MIHIR BHARATBHAI RUPAPARA
114670	92100133066	ARPIT DIPAKKUMAR AGRAVAT
113821	92100133029	MEET BHARATBHAI KALATHIYA
110402	92000133018	PUSHTI SUBHASHBHAI DEPANI
109229	92000133006	TAPAN VIPULBHAI KHOKHARIYA
110473	92000133021	HUZEFA ALIASGAR DHANKOT
109107	92000133003	CHARMI SALIMBHAI GANGANI
111810	92110133014	HARSH KIRITBHAI SAPARA
109549	92000133009	RENISH VIMALBHAI SURANI

PHOTOS:

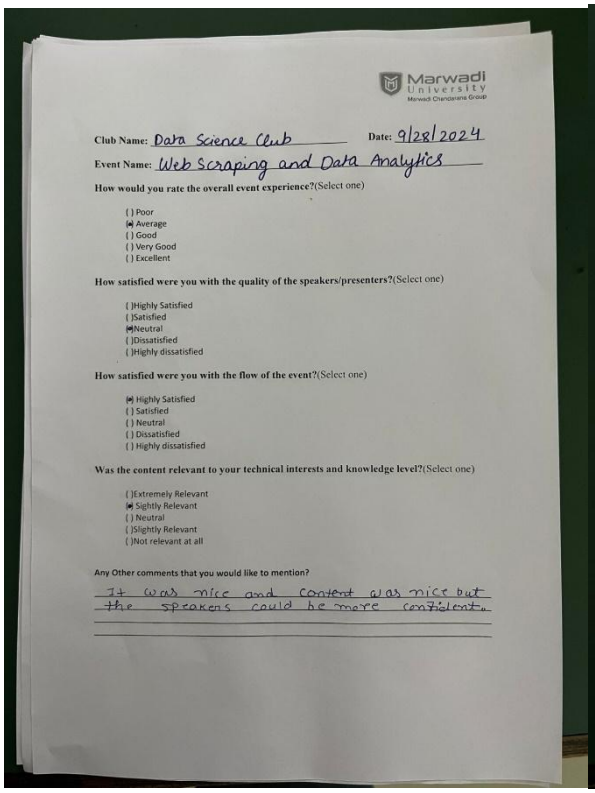

ZURU
GPS Map Camera
Rajkot,GJ,India
Rajkot, Rajkot, 360003, GJ, India
Lat 22.367143, Long 70.796960
09/28/2024 10:23 AM GMT+05:30
Note : Captured by GPS Map Camera


GPS Map Camera
Rajkot,GJ,India
Rajkot, Rajkot, 360003, GJ, India
Lat 22.367181, Long 70.796857
09/28/2024 11:57 AM GMT+05:30
Note : Captured by GPS Map Camera


GPS Map Camera
Rajkot,GJ,India
Rajkot, Rajkot, 360003, GJ, India
Lat 22.367259, Long 70.796942
09/28/2024 11:58 AM GMT+05:30
Note : Captured by GPS Map Camera


GPS Map Camera
Rajkot,GJ,India
Rajkot, Rajkot, 360003, GJ, India
Lat 22.367177, Long 70.796848
09/28/2024 12:46 PM GMT+05:30
Note : Captured by GPS Map Camera

Feedback form for Participants:



Marwadi University
Marwadi Chandarana Group

Club Name: Data Science Club Date: 9/28/2024
Event Name: Web Scraping and Data Analytics

How would you rate the overall event experience?(Select one)

( ) Poor
(•) Average
( ) Good
( ) Very Good
( ) Excellent

How satisfied were you with the quality of the speakers/presenters?(Select one)

( )Highly Satisfied
( )Satisfied
(•)Neutral
( )Dissatisfied
( )Highly dissatisfied

How satisfied were you with the flow of the event?(Select one)

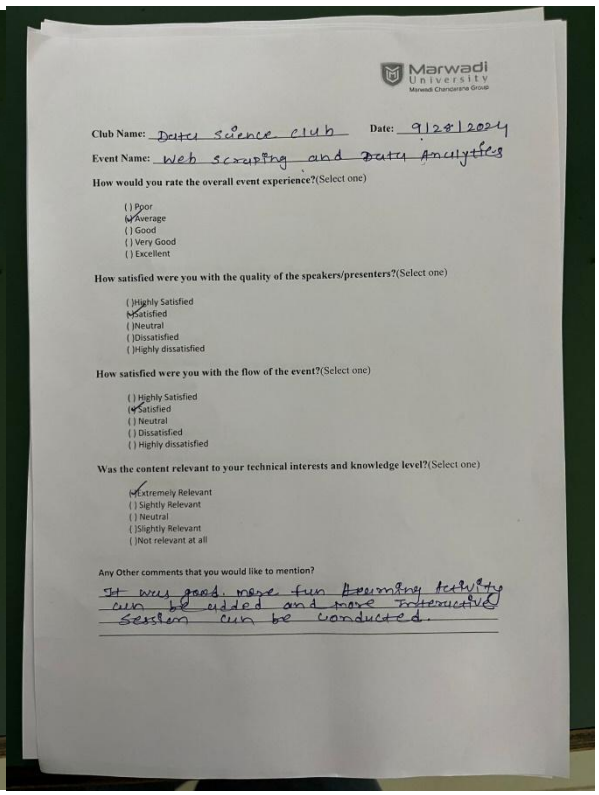
(•) Highly Satisfied
( ) Satisfied
( ) Neutral
( ) Dissatisfied
( ) Highly dissatisfied

Was the content relevant to your technical interests and knowledge level?(Select one)

( )Extremely Relevant
(•) Sightly Relevant
( ) Neutral
( )Slightly Relevant
( )Not relevant at all

Any Other comments that you would like to mention?

It was nice and content was nice but the speakers could be more confident.



Marwadi University
Marwadi Chandarana Group

Club Name: Data Science club Date: 9/28/2024
Event Name: Web scraping and Data Analytics

How would you rate the overall event experience?(Select one)

( ) Poor
(✓) Average
( ) Good
( ) Very Good
( ) Excellent

How satisfied were you with the quality of the speakers/presenters?(Select one)

( )Highly Satisfied
(✓)Satisfied
( )Neutral
( )Dissatisfied
( )Highly dissatisfied

How satisfied were you with the flow of the event?(Select one)

( ) Highly Satisfied
(✓)Satisfied
( ) Neutral
( ) Dissatisfied
( ) Highly dissatisfied

Was the content relevant to your technical interests and knowledge level?(Select one)

(✓)Extremely Relevant
( ) Sightly Relevant
( ) Neutral
( )Slightly Relevant
( )Not relevant at all

Any Other comments that you would like to mention?

It was good. more fun learning activity can be added and more interactive session can be conducted.