**WEB SCRAPPER**

**SYNOPSIS**

*Submitted to Department of Computer Science Engineering, University Institute of Engineering and Technology,*

*M. D. University, Rohtak*

**Submitted By: Harsh Kumar**

**Roll No - 23617**

**B.Tech (Computer Science and Engineering)**

**Under the Supervision of:**

**Dr. KAMALDEEP**

**Assistant Professor,**

**Department of Computer Science and Engineering, Maharshi Dayanand University, Rohtak-124001**



**UIET, MAHARSHI DAYANAND UNIVERSITY, ROHTAK-124001 (INDIA)**

**October 2022**

**CONTENTS**

|  |  |
| --- | --- |
| **1.** | **ABSTRACT** |
| **2.** | **INTRODUCTION** |
| **3.** | **OBJECTIVES** |
| **4.** | **ADVANTAGES** |
| **5.** | **REFRENCES** |
|  |  |

**DECLARATION**

**I hereby certify that the work which is being presented in the report entitle “WEB SCRAPPER” by Harsh Kumar in partial fulfillment of requirement for the award of degree of B.Tech submitted to department of Computer Science and Engineering of UIET at MDU, Rohtak is an authentic record of my own work carried out during a period from August 2022 to December 2022.**

**Signature of Student**

**This is to certify that the above statement made by the candidate is correct to the best of my knowledge.**

**Signature of Supervisors**

1. **ABSTRACT**

From the advancement of the World Wide Web, the situation of the web client and information trade has fastly changed. As average citizens join the web and begin to utilize it, heaps of new systems are elevated to help up the system. Simultaneously to improve PCs and system office, new innovations were brought which results in a consequent reduction in the cost of equipment and site's connected expenses. Business, academician, scientists all share their data on the web with the goal that they can be associated with individuals fastly with no problem at all. Because of exchange, share, and storage options for the information on the web, another issue emerged that how to deal with such information overload and how the client will get or get to the best data in the least endeavours. To settle these issues, specialist spot out a new procedure called Web Scraping. Web scraping, or web scratching, is a procedure which is utilized to create organized information based on accessible unstructured information on the web. Created organized information at that point is then put away in focal database to be dissected in spreadsheets. Presently, there are heaps of tools accessible in the market for web scratching. This paper is centered around the overview of the data extraction method and how to implement it using python.

# ACKNOWLEDGEMENT

We are highly grateful to the Dr. Yudhvir Singh, Director, University Institute of Engineering and Technology (UIET) for providing this opportunity to carry out the major project work at UIET, MDU

The constant guidance and encouragement received from Dr. Kamaldeep, CSE Department, UIET, MDU, Rohtak has been of great help in carrying out the project work and is acknowledged with reverential thanks.

We would like to express gratitude to other faculty members of Computer Science and Engineering Department of UIET, MDU, Rohtak for their intellectual support throughout the course of this work.

Finally, we are indebted to all whosoever have contributed in this report work.

Student name: Harsh Kumar

Roll No: 23617

7th sem, B.tech(CSE B), UIET

# INTRODUCTION

**Description of the project:**

In today's time of data science & engineering, it is entirely expected to gather information from sites for examination purposes. Realizing how to scrap site pages will set aside your time and money. A few organizations like Twitter do provide APIs to get their data in a progressively composed manner while we need to scratch different sites to get information in an organized configuration.

The general thought behind web scratching is to recover information that exists on a site and convert it into a configuration that is usable for analysis. Python is one of the most normally utilized programming dialects for data science ventures. Utilizing python with BeautifulSoup makes web scrapping simpler. Through this paper, we will be experiencing a detail however a basic clarification of how to scratch information in Python using BeautifulSoup. This will help information researchers gather and store information from site pages effortlessly without investing an excess of energy in getting ready datasets.

1. **OBJECTIVE**
2. **ADVANTAGES**

**1. Time Efficient**

* The advantage of web Scraping is its time-efficient and low maintenance. For example, downloading big data may take hours, and then analyzing every single row manually at a time is worth spending your entire month.

**2. Complete Automation**

* Many web scraping services can be automated through Big Data Analytics and [Machine Learning](https://optisol.com.au/services/).
* Some advantages of automation are that it doesn’t get bored or tiring, does not require any breaks, and never gets distracted they follow the given instructions.

**3. Cost Efficiency**

* Web scraping services provide essential services at a competitive cost because it’s much cheaper than hiring a company to perform the same task.

**4. Track product performance**

* Better product analysis is yet another perk of website scraping.
* By monitoring listings and sales data, it allows you to see how well different products are performing.

**5. Data Accuracy**

* There are no humans are involved in this process, Simple errors in data extraction may lead to major issues.
* Web scraping is not only a fast process, but it’s also very accurate too. [Data scraping](https://datalabs.optisolbusiness.com/) is a way of extracting information from various websites and extracting data for other purposes.

1. **REFERENCES**

* <https://www.google.com/>
* <https://github.com/>
* wikipedia.com
* https://[www.youtube.com/](http://www.youtube.com/)