ABSTRACT

Main objective of Web Scraping is to extract information from one or many websites and process it into simple structures such as spreadsheets, database or CSV file. However, in addition to be a very complicated task, Web Scraping is resource and time consuming, mainly when it is carried out manually. Previous studies have developed several automated solutions. The purpose of this article is to revisit the different existing Web Scraping approaches, categories, and tools, but also its areas of application.

1.WHAT IS WEB SCRAPING?

[Web scraping](https://parsehub.com/) refers to the **extraction of data from a website**. This information is collected and then exported into a format that is more useful for the user. Be it a spreadsheet or an API.



Although [web scraping can be done manually](https://www.upwork.com/search/profiles/?nbs=1&q=web%20scraping), in most cases, automated tools are preferred when scraping web data as they can be less costly and work at a faster rate.

But in most cases, web scraping is not a simple task. [Websites come in many shapes and forms](https://www.expertmarket.co.uk/web-design/different-types-of-websites), as a result, web scrapers vary in functionality and features.

This project will show what can be done with web scraping:

* This will show how to scrap a news or data from another site with the help of web scraping.
* We have to enter the URL manually.
* Then there is button scraping, That’s what we have to choose.
* And here we go!! We are done.

2.SCOPE

In this document we have explained the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. This project will run with help of python language.

3.Definition

* URL- Universal resource locator
* HTML-Hypertext markup language
* JSON-JavaScript object notation
* CSV-Comma separated values

GENERAL DESCRIPTION

1.Product perspective

In this project we have done a web scraping that scraps message from the site and gives information on the site. It is an interesting project that we have done.

2. Problem statement

Here we got a problem with the web scraping and gathering a information from the site we didn’t familiar with.

3. Problem solution

We have successfully created a code with the help of python and its dictionary beautiful soup.

4. Further improvement

By this we can use the method of scraping for news gathering, information gathering, Article reading and so on…

This will make things easier for the people and for the Businessman.

5.Tools used

* We have used python programming language as the major tool.
* And we have used dictionary to run the code for crawling .
* The dictionary is” PYTHON BEAUTIFUL SOUP DICTIONARY”.
* We have successfully tested it in the Linux and we compile and debug in the PYCHARM IDE.

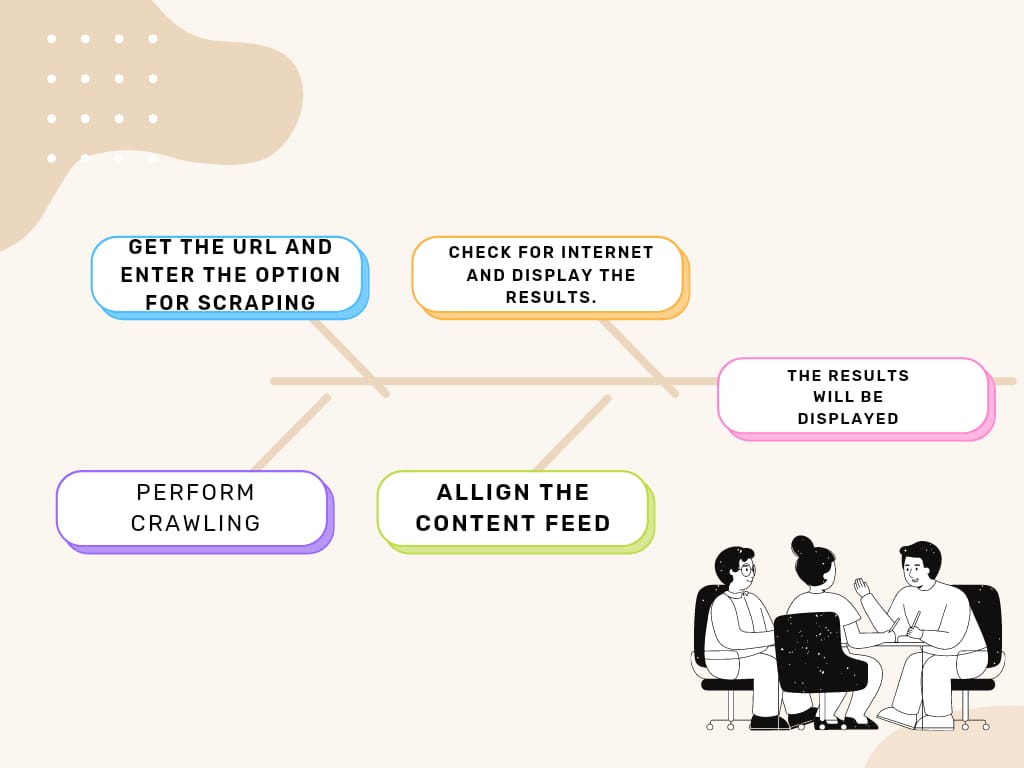
7.Constaraints

This will help the normal people to the high-level people. This will makes the news easier to throughout the large circumstance.

DESEIGN DETAILS

1.Process workflow

Here we can see the workflow of the project we done here. It is an simple flowchart explains our project on behalf of web scraping.



2. Error Handling

Eventually we have some error with the dictionary we have worked with there is some error happened. We have fixed that error by the maximum effort.

PERFORMANCE

1.Reusability

The method and we used dictionary method are fair and easy to use so it stands on reusability.

2.Application Compatibility

Since we are using python and it is compatible with any platform, we follow Application compatibility.

3. Deployment

Our code is deployed in github. And we have done a demo video for this.

CONCLUSION

It is important to recognize that in certain circumstances web scraping can be illegal. If the terms and conditions of the web site we are scraping specifically prohibit downloading and copying its content, then we could be in trouble for scraping it.