**Web Based scrapping Tools**

School

Course

Unit

**Abstract**

Web scraping, also known as data mining, is the process of collecting large amounts of data from the web and then placing it in databases for future analysis and later use. Web scraping offers insight into price data, market dynamics, prevailing trends, practices employed by your competitors and the challenges they face. This is readily available data if you know how to get it. What many marketers don't know is how useful it can be to them. In this article, we will explain the advantage of Web scraping technique and supply a practical example that can be beneficial for e-commerce businesses and online marketers. This research is based on two web based python scrapping tool and the main focus is on selenium and BeautifulSoup to scrap covid 19 data.

**KEYWORDS** web scraping, e-marketing, e-commerce, python.

**INTRODUCTION**

The internet is a bottomless well of knowledge that is open to everyone. Because of the latest technological developments, we've had to rethink our company practices. To conduct business in today's world, you must go online. One of the keys to success in e-commerce and e-marketing is knowing how to use the Internet and the various opportunities it can bring (Van den Brink, 2021). Data crawling, online scraping, and data harvesting have been around for as long as the internet itself has been around. However, at the beginning, it wasn't always used for web content extraction. It is possible to think of online scraping as the process of obtaining or extracting content from a website In web scraping, content is taken from a website and used in a manner that is not within the control of the website's owner.

**What is Web Scraping?**

Automated data extraction from webpages is known as web scraping. Mathew explains web scraping as the process of extracting data from a semi-structured document found on the internet, typically in the form of a web page, and using it for another purpose.

**BeautifulSoup**

HTML and XML parse trees may be easily navigated, searched, and customized using Beautiful Soup's easy methods. It takes a long-winded HTML page and turns it into a Python object tree. In addition, the page is immediately converted to Unicode, saving you the time and hassle of worrying about encodings. You may use this tool to not only scrape data, but also to clean it up. A number of Python parsers, including lxml and hml5lib, are supported by Beautiful Soup, in addition to the HTML parser included in Python's standard library.

**Selenium**

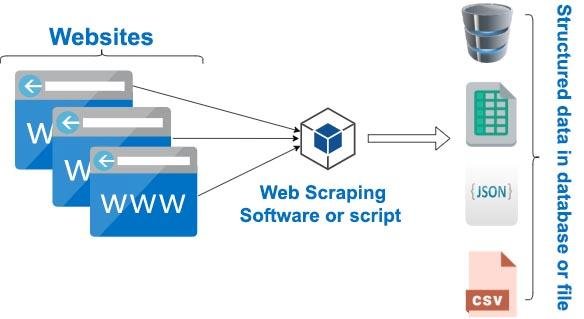
Selenium is a web application testing framework that may be used on any platform. It operates on Windows, Linux, and macOS and is open-source software provided under the Apache License 2.0. Selenium, despite its primary function, is also used for web scraping. WebDriver, one of Selenium's most helpful web scraping components, will be the emphasis of this article rather than the rest of the framework. In order to write and run tests, we can use Selenium WebDriver's programming interface to control a web browser.

Scraping data from websites will be our primary use for this tool. When a website uses JavaScripts to produce content, Selenium is a useful tool. The dynamic nature of these websites renders Scrapy's web scraping capabilities completely worthless. In this article, I hope to introduce you to Selenium and demonstrate some fundamental web scraping techniques.

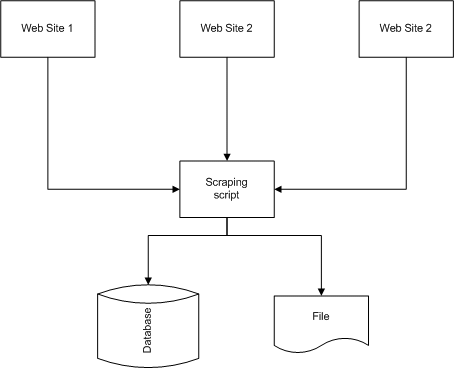
**HOW WEB SCRAPERS WORK**

To test web applications, you can use Selenium. Windows, Linux, and macOS can all use this Apache License 2.0-licensed open source application. Selenium is also used for web scraping, despite its primary function. We will dive into Selenium's components, therefore focusing on WebDriver, a component that's useful for web scraping. Selenium WebDriver gives us command over a web browser via a programming interface, allowing us to write and run tests.

Scraping data from websites is what we'll be doing with it in our case. Websites that use JavaScripts to display content can benefit from Selenium. With such dynamic websites, Scrapy, a well-known web scraping framework, is rendered ineffective (Patel, 2020). The purpose of this tutorial is to introduce you to Selenium and demonstrate how to scrape the web using Selenium. A web scrapper's architecture is depicted in the diagram below.



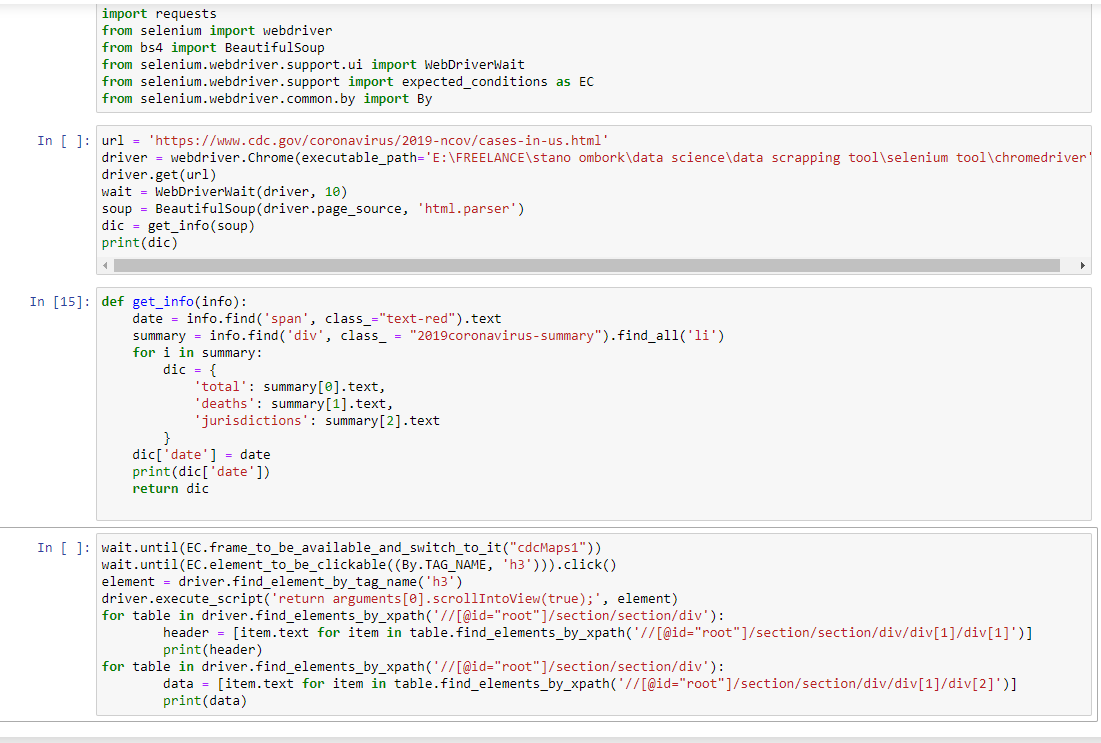
However, more powerful scrapers, such as those that may be used for an API rather than just a CSV or Excel spreadsheet will be able to export data in a variety of formats.



**CONCLUSION**

As a result of the Internet's availability, information-seeking activities grew common and time-consuming, making them among of the most prevalent and time-consuming activities. For the first time, we will be able to acquire large amounts of data without putting in a lot of effort or money. With web scraping, businesses may gain insight into their target market and make better decisions, whether they are in e-commerce or e-marketing.

**Appendix A – selenium code**



**Appendix B - BeautifulSoup code**



**REFFERENCES**

van den Brink, R. (2021). A Bottomless Well and Other Solutions. In *The End of an Antibiotic Era* (pp. 261-322). Springer, Cham.

Patel, J. M. (2020). Web Scraping in Python Using Beautiful Soup Library. In Getting Structured Data from the Internet (pp. 31-84). Apress, Berkeley, CA.