International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00288
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

Web scraping with excel

Manikandaprabhu T (22bcs138)
Department of Computer Science , Sri Krishna Arts and Science College, Coimbatore

ABSTRACT

Web scraping is a process that is particularly important in fields such as Business Intelligence in the modern age. Web scrapping is a technology that allow us to extract structured data from text such as HTML. Web scrapping is extremely useful in situations where doesisn't provided in readable format such as JSON or XML. The use of web scrapping to gather data allows us to gather prices in near real time from retail store sites and provide further details, web scrapping can also be used to gather intelligence of illicit businesses such as drug marketplaces in the darknet to provide law enforcement and researchers valuable data such as drug prices and varieties that would be unavailable with conventional methods. It has been found that using a web scraping program would yield data that is far more thorough, accurate, and consistent than manual entry Web scraping with Excel is the process of automatically extracting data from websites and importing it into Excel for analysis and reporting purposes. Excel's built-in web scraping tool, Power Query, allows users to connect to a website, select the data they want to extract, and transform the data into a usable format. However, it's important to be aware of the potential legal and ethical issues surrounding web scraping, and to use discretion when extracting data from websites

KEYWORDS

Scraping, Excel, Data, structure.

INTRODUCTION

Microsoft Excel is undoubtedly one of the best scraping tools Python to manage information in a structured form. Excel is like the Swiss army knife of data, with its great features and capabilities. Here is how MS Excel can be used as a basic web scraping tool to extract web data directly into a worksheet. We will be using Excel web queries to make this happen.

Web queries feature of MS Excel is used to fetch website data to excel and can be extracted into the worksheet easily. It can automatically find tables on the webpage and would let you pick the particular table you need the data from. Web queries can also be handy in situations where an ODBC connection is impossible to maintain, apart from just extracting data from the web pages. Let's see how web queries work and how you can crawl HTML tables and use MS Excel as the best scraping tool in python.

Web Scraping

Web scraping is the process of extracting data from websites. While it is possible to do web scraping with Excel, it is not the most efficient tool for the job. Excel is primarily designed for managing and analyzing data stored in spreadsheets. While it does have some web scraping capabilities, it is limited and requires a lot of manual effort. Additionally, web scraping with Excel can be slow and error-prone, especially when dealing with large amounts of data or complex websites.If you need to do web scraping, there are more powerful and specialized tools available, such as Python's BeautifulSoup or Scrapy. These tools allow you to automate the process of extracting data from websites and can handle complex data structures and large volumes of data.That being said, if you only need to extract a small amount of data from a simple website, you can

ISSN: 2583-6129
www.isjem.com

use Excel's built-infunctionality to import data from a web page. To do this, go to the Data tab in the ribbon, select "From Web" and follow the prompts to specify the URL and data you want to import Web scraping involves extracting data from websites, and it can be a powerful tool for gathering information or conducting research. While there are many specialized tools for web scraping, Excel can also be used to extract data from websites When web scraping, it is important to respect the website's terms of service and to avoid violating any legal or ethical boundaries. It is also important to be aware of potential issues such as data privacy and security concerns.In summary, web scraping is a powerful tool for extracting data from websites. It can be done manually or using specialized software or programming tools, and can be used for a variety of purposes. However, it is important to use web scraping responsibly and to be aware of potential legal and ethical issues.

Types of web scrapers

Self-built Web Scrapers that requires advanced knowledge of programming. And if you want more features in your Web Scraper, then you need even more knowledge. On the other hand, pre-built Web Scrapers are previously created scrapers that you can download and run easily. These also have more advanced options that you can customize Self-built web scrapers are programs or scripts that are developed by users to extract data from websites. These scrapers are typically customized to meet specific needs and can be built using various programming languages, such as Python, JavaScript, or Ruby. The process of building a web scraper involves identifying the target website and analyzing its structure to determine the best way to extract the desired data. This may involve analyzing the HTML and CSS of the website, identifying specific tags and classes that contain the data, and using techniques like regular expressions to extract the data

Browser extensions Web Scrapers are extensions that can be added to your browser. These are easy to run as they are integrated with your browser, but at the same time, they are also limited because of this. Any advanced features that are outside the scope of your browser are impossible to run on Browser extension Web Scrape Browser extensions are add-ons that users can install on their web browsers to add new features and functionality. Some browser extensions are designed for web scraping, allowing users to extract data from websites without needing to write code or use external software.
There are many browser extensions available for web scraping, such as Web Scraper, Data Miner, and Scraper. These extensions typically work by allowing users to select the data they want to extract using a visual interface, and then exporting the data to a spreadsheet or other file format

Cloud Web Scrapers run on the cloud, which is an off-site server mostly provided by the company that you buy the scraper from. These allow your computer to focus on other tasks as the computer resources are not required to scrape data from websites Cloud web scrapers are web-based services that allow users to extract data from websites using a remote server or cloud infrastructure. With cloud web scrapers, users can avoid the need to install and maintain software on their local machines, and can easily scale up or down to accommodate larger or smaller scraping projects.
Cloud web scrapers typically provide a user-friendly interface for creating scraping tasks, specifying the websites to scrape, and selecting the data to extract. Some cloud web scrapers also provide features like scheduling, data visualization, and data transformation to help users make the most of their extracted data.

ISSN: 2583-6129
DOI: 10.55041/ISJEM00288
www.isjem.com

Conclusion

In conclusion, web scraping with Excel can be a powerful tool for extracting data from websites and importing it into a format that is usable for analysis and reporting. Excel's built-in web scraping tool, Power Query, provides a user-friendly interface for extracting and transforming data, and can be used to create custom scraping solutions.

However, it's important to be aware of the potential legal and ethical issues surrounding web scraping, as some websites may prohibit or restrict scraping of their data. In addition, users should exercise caution when using self-built web scrapers or third-party tools like browser extensions and cloud web scrapers, and ensure that they are adhering to the terms of service of the websites they are scraping.

Overall, web scraping with Excel can be a valuable skill for professionals in a variety of fields, including marketing, finance, and research, allowing them to gather and analyze data more efficiently and effectively

Reference

effectively Get external data from a Web page" - Microsoft Support: https://support.microsoft.com/en-us/office/get-external-data-from-a-web-page-3a557ddb-00b3-4f7a-9ecf-72a300816947
This is a tutorial on using Power Query in Excel to extract data from websites.

"Web Scraping with Excel Power Query" - Excel Campus: https://www.excelcampus.com/powerquery/web-scraping-with-power-query/
This is a comprehensive guide on web scraping with Excel, covering topics such as connecting to websites, selecting data to extract, and transforming the data.

"Web Scraping Tutorial with Excel VBA" - WiseOwl: https://www.youtube.com/watch?v=4PwGwA5ybh0
This is a video tutorial on using Excel VBA to create a custom web scraper.

"Web Scraping using Excel VBA" - DataCamp: https://www.datacamp.com/community/tutorials/web-scraping-using-excel-vba
This is a tutorial on using Excel VBA to extract data from websites.

"Web Scraping with Excel: A Beginner's Guide" - JotForm: https://www.jotform.com/blog/web-scraping-with-excel/
This is a beginner's guide to web scraping with Excel, covering topics such as finding data sources, using Power Query, and avoiding legal issues