Python Advance Assignment - 6

Name:K.Durga Surya Prasad

Roll No: 19B21A0549

**1.Explain why selenium is important in web scraping.**

Ans: Selenium is an automation testing framework for web applications/websites which can also control the browser to navigate the website just like a human. Selenium uses a web-driver package that can take control of the browser and mimic user-oriented actions to trigger desired events.

**2. What's the difference between scraping images and scraping websites?**

**Use an example to demonstrate your point.**

Ans: Scraping Images: Image scraping is a subset of the web scraping technology. While web scraping deals with all forms of web data extraction, image scraping only focuses on the media side – images, videos, audio, and so on.

Scraping websites: Web scraping is an automatic method to obtain large amounts of data from websites. Most of this data is unstructured data in an HTML format which is then converted into structured data in a spreadsheet or a database so that it can be used in various applications. There are many different ways to perform web scraping to obtain data from websites. These include using online services, particular API’s or even creating your code for web scraping from scratch. Many large websites, like Google, Twitter, Facebook, StackOverflow, etc. have API’s that allow you to access their data in a structured format. This is the best option, but there are other sites that don’t allow users to access large amounts of data in a structured form or they are simply not that technologically advanced. In that situation, it’s best to use Web Scraping to scrape the website for data.

1. Price Monitoring

Web Scraping can be used by companies to scrap the product data for their products and competing products as well to see how it impacts their pricing strategies. Companies can use this data to fix the optimal pricing for their products so that they can obtain maximum revenue.

2. Market Research

Web scraping can be used for market research by companies. High-quality web scraped data obtained in large volumes can be very helpful for companies in analyzing consumer trends and understanding which direction the company should move in the future.

**3. Explain how MongoDB indexes data.**

Ans: MongoDB uses multikey indexes to index the content stored in arrays. If you index a field that holds an array value, MongoDB creates separate index entries for every element of the array. These multikey indexes allow queries to select documents that contain arrays by matching on element or elements of the arrays.

**4.What is the significance of the SET modifier?**

Ans: A set modifier modifies the selection of the preceding set identifier. If no set identifier is referenced, the current selection state is implicit.

**5. Explain the MongoDB aggregation framework.**

Ans: In MongoDB, aggregation operations process the data records/documents and return computed results. It collects values from various documents and groups them together and then performs different types of operations on that grouped data like sum, average, minimum, maximum, etc to return a computed result.