

Q1:- Web scraping is a term used to describe the use of a program or algorithm to extract and process large amounts of data from the web. Whether you are a data scientist, engineer, or anybody who analyzes large amounts of datasets, the ability to scrape data from the web is a useful skill to have.

Web scraping is commonly used to retrieve the most updated data about properties, sale prices, monthly rental income, amenities, property agents, and other data points. Web scraped data also informs property value appraisals, rental yield estimates, and real estate market trends analysis.

3 areas where web scrapping is used to get data is:- 1).Market Research 2).News Monitoring 3).Price Monitoring

Q2:- 1)Human copy-and-paste. 2)Text pattern matching. 3)HTTP programming. 4)HTML parsing. 5)DOM parsing. 6)Vertical aggregation. 7)Semantic annotation recognizing. 8)Computer vision web-page analysis.

Q3 BeautifulSoup is a Python library that makes it easy to scrape information from web pages. It sits atop an HTML or XML parser and provides Pythonic idioms for iterating, searching, and modifying the parse tree.

Scrapy, Selenium, import.io, ParseHub, and Octoparse are the most popular alternatives and competitors to BeautifulSoup.

As BeautifulSoup is not a standard python library, we need to install it first. We are going to install the BeautifulSoup 4 library (also known as BS4), which is the latest one. To isolate our working environment so as not to disturb the existing setup.

Q4

Flask is a lightweight framework to build websites. We'll use this to parse our collected data and display it as HTML in a new HTML file. The requests module allows us to send http requests to the website we want to scrape. The first line imports the Flask class and the render_template method from the flask library.

Q5 AWS Services are:-

1. Amazon EC2 (Elastic Compute Cloud):-EC2 is a cloud platform provided by Amazon that offers secure, and resizable compute capacity. Its purpose is to enable easy access and usability to developers for web-scale cloud computing, while allowing for total control of your compute resources.
2. Amazon RDS (Relational Database Services):-Amazon Relational Database Service (Amazon RDS) makes database configuration, management, and scaling easy in the

cloud. Automate tedious tasks such as hardware provisioning, database arrangement, patching, and backups – cost-effectively and proportionate to your needs.

3. Amazon S3 (Simple Storage Service):- Amazon S3, at its core, facilitates object storage, providing leading scalability, data availability, security, and performance. Businesses of vast sizes can leverage S3 for storage and protect large sums of data for various use cases, such as websites, applications, backup, and more.
4. D. Amazon Lambda:-Lambda permits you to run code without owning or managing servers. Users only pay for the compute time consumed.
5. Amazon Cognito:- AWS Cognito administers a control access dashboard for on-boarding users through sign-up, and sign-in features to their web and mobile apps. AWS Cognito scales to millions of users and offers sign-in support with social identity providers including Facebook, Google, and Amazon, along with enterprise identity providers via SAML 2.0.
6. Amazon Glacier:-AWS Glacier services are secure, flexible, and affordable Amazon S3 cloud storage classes for data caching and prolonged backup.
7. Amazon SNS:-Amazon SNS is a fully managed messaging solution that provides low-cost infrastructure for bulk message delivery, primarily to mobile users.
8. Amazon Lightsail:- Amazon Lightsail is an easy on-ramp for users who are getting started with AWS and just need virtual private servers. Lightsail has the tools required for a speedy product launch at a low and predictable price.

In []: