

Q1. What is Web Scraping? Why is it Used? Give three areas where Web Scraping is used to get data.

Answer :

Web Scraping is data scraping used to scrap or extract data from websites, It's actually a technique to gather data from internet. For the first time people think that it's illegal but no, that's not illegal but legal. For example : we can extract product data from ecommerce websites for the product development

3 areas where Web Scraping is used to get data :

Market Research : In market, We can collect competitors product data using web scraping, like pricing, quality and etc.

Lead Generations : Using Web Scraping we can collect potential customers information, which can help to grow a company

Analysis : We can use Web Scraping to gather any information for analysis

Q2. What are the different methods used for Web Scraping?

Answer :

1. HTTP Programming :

This method involves sending http requests to the websites and parsing the html content of the websites

2. DOM Programming :

In this method, to inspect a webpage or dynamically modify it, client-side scripts parse the contents of webpage into DOM tree

Copy-pasting :

The manual human examination and copy-pasting method may sometimes prove

Q3. What is BeautifulSoup? Why is it used?

Answer :

Beautiful Soup is a python library used for web scraping. it's used to pull out the data from HTML and XML files. It creates a parse tree from the source code, which can be further used to extract data from the website in more readable format.

Q4. Why is flask used in this Web Scraping project?

Answer :

Flask is used for web scraping coz it provides a user friendly and easy to use interface to create a web application that can scrap data from websites

Q5. Write the names of AWS services used in this project. Also, explain the use of each service.

Answer :

Code Pipeline and Beanstalk are the two services, which we used in this project.

As much as we are told, Code Pipeline works as a moderator/broker between our GitHub Code and the beanstalk

Beanstalk is used to quickly deploy and manage our web applications

