

Hemalatha M

Assignment: -

1. A developer is assigned a task to scrape 1 lakh website pages from a directory site, while scrapping he is facing such captcha, which are placed to stop people from scrapping As a project Coordinator suggest ways to solve this problem

In here the 1st option is to use CAPTCHA solving services where some are like 2Captcha, Anti-Captcha where it can also be expensive if you want to scrape in large number of pages. And another choice is headless browser is a web browser that can run without a graphical user interface (GUI). This makes it ideal for web scraping, as you can control the browser programmatically. There are a number of headless browsers available, such as Selenium and Puppeteer, then another option is to use a proxy server where have to be carefully use it if its illegal where this proxy server is used as a mediator where there are many free servers available.

2. Our client has around 10k linkedin people profiles, he wants to know the estimated income range of these profiles. Suggest ways on how to do this?

The first option is analyse the job titles, companies and industries to estimate income level, then another option is to look at the pattern of there education and analyse based on that.

3. We have a list of 1L company names, need to find linkedin company links of these profiles, how to go about this?

Utilize the LinkedIn API to search for company profiles based on company names, Implement web scraping techniques to extract company links from search engine results or business directories, Leverage third-party data providers like Crunchbase or ZoomInfo for comprehensive company information.

4. How to identify list of companies whose tech stack is built on Python. Give names of 5 companies if possible, by your suggested approach

Utilize tools like BuiltWith or Wappalyzer to identify technologies used by companies, Explore GitHub repositories of these companies to determine the use of Python in their tech stack, Look for company job postings or technical blogs mentioning the use of Python.

5. Need to find an API, through which we can send linkedin messages to other linkedin users

Utilize the LinkedIn Messaging API, if available, to send messages programmatically, Explore third-party LinkedIn automation tools that offer messaging functionalities, Develop a custom script using the LinkedIn API to send messages to other users.