Assignment For IT Business Analytics Internship

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

Solution –

1. Use a CAPTCHA solver service:

There are a number of CAPTCHA solver services available, such as **Anti-Captcha, 2Captcha**, and **CaptchaMonkey.** These services use human workers to solve CAPTCHAs, and they can be integrated into your scraping script.

2. Rotate IP addresses:

hCAPTCHAs can detect and block repeated requests from the same IP address. To avoid this, you can use a proxy pool to rotate your IP address every request.

3. Use a browser emulation tool:

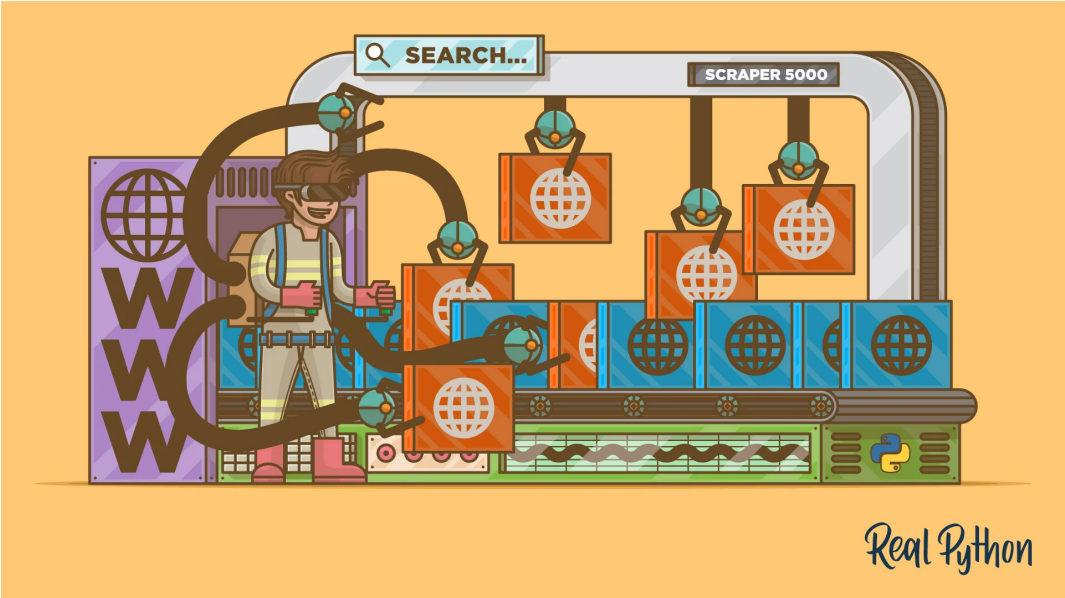
A browser emulation tool can help you to simulate human behavior when scraping. This can include things like varying the speed of your requests, waiting between requests, and using different User-Agent strings.

4. Avoid hidden traps:

Some hCAPTCHAs include hidden traps, such as asking you to click on a certain part of the image or to type a certain word. You can avoid these traps by using a CAPTCHA solver service that is specifically designed to solve hCAPTCHAs.

5. Save cookies:

hCAPTCHAs can also use cookies to track your activity. To avoid this, you can save the cookies from your first request and use them in subsequent requests.



6. Hide automation indicators:
hCAPTCHAs can also detect automation indicators, such as the presence of headless browsers or certain types of web drivers.

- Start with a small number of requests and gradually increase the number of requests over time.
- Avoid scraping too many pages from the same website in a short period of time.
- Be respectful of the website's robots.txt file.
- If you are blocked, stop scraping for a while and try again later.

Here are some additional suggestions for the developer:

- Use a cloud-based scraping platform.
- Use a multithreaded scraping script. This will allow you to scrape multiple pages at the same time, which can speed up your scraping process.
- Split the directory site into multiple parts and scrape each part separately.

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?

Solution–

Here are some ways to estimate the income range of 10k LinkedIn people profiles:
Use LinkedIn Salary Insights:

LinkedIn Salary Insights is a tool for estimating base salary and total compensation:

- Create a LinkedIn Talent Insights account to access the tool.
- Generate a Talent Pool Report for your client's dataset, which includes compensation summaries.
- The report provides the average total compensation and compensation range for the profiles in the dataset.

Use a Third-party Salary Calculator:

- Various third-party salary calculators like Salary.com, PayScale, and Glassdoor are available.
- Input job titles, company names, locations, and experience details for each profile.
- The calculator estimates the salary range for each job profile based on the provided data.

Use a Combination of LinkedIn Salary Insights and Third-party Calculators:

- Begin with LinkedIn Salary Insights for a general income estimate for the profiles.
- Then, use third-party salary calculators to obtain more precise income estimates for each individual profile.
- This approach offers a balance between general insights and detailed estimations.

Use a machine learning model:
You can also use a machine learning model to estimate the income range of the LinkedIn profiles in your client's dataset. To do this, you will need to train the model on a dataset of LinkedIn profiles with known income ranges. Once the model is trained, you can use it to estimate the income range of the LinkedIn profiles in your client's dataset.

Contact the LinkedIn profiles directly:
If you need a more accurate estimate of the income range for a particular LinkedIn profile, you can contact the profile owner directly. However, it is important to note that not everyone is comfortable sharing their salary information.

- Which method you choose will depend on your budget and the level of accuracy you need.
- If you need a general estimate of the income range for the LinkedIn profiles in your client's dataset, then using LinkedIn Salary Insights or a third-party salary calculator is a good option.
- If you need a more accurate estimate for a particular LinkedIn profile, then you may need to contact the profile owner directly.

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

Solution–

Use a LinkedIn Company Finder Tool:
- Consider using tools like Botster, Phantombuster, or Derrick App.
- These tools allow you to scrape LinkedIn for company pages based on a list of company names.

Utilize Browser Extensions:
- Explore browser extensions like LinkedIn Sales Navigator and LinkedIn Sales Solutions.
- Install the extension and navigate to the company's website.
- The extension will automatically detect the company's LinkedIn page and provide you with a link.

Manual LinkedIn Search (for a Small List):
- For a small number of company names, you can search for the company pages manually on LinkedIn.
- Go to the LinkedIn search bar and type in the company name to find the company's page.

Efficiency Consideration:
If you have a large list of company names, it's recommended to use a LinkedIn company finder tool or a browser extension, as it will save you a significant amount of time and effort.

Exporting LinkedIn Company Links:
- Once you find the LinkedIn company pages, you can export the links to a spreadsheet or CSV file.
- This file can be used for automating tasks like sending connection requests or messages to the companies.

Important Note:

- LinkedIn has policies against scraping and automating tasks, and violating these policies can lead to account suspension or banning.
- It's crucial to use LinkedIn company finder tools and browser extensions responsibly and in compliance with LinkedIn's terms and conditions.

There are a number of AI tools that can be used to find LinkedIn company links from a list of company names. Here are a few examples:

Phantombuster:
Phantombuster is a cloud-based automation platform that can be used to scrape LinkedIn for company pages. It offers a number of different templates for scraping LinkedIn, including a template for scraping company pages.

Expandi:
Expandi is a browser extension that can be used to automate tasks on LinkedIn, such as sending connection requests and messages. It also offers a feature for scraping LinkedIn company pages.

Crystal:
Crystal is an AI-powered tool that can be used to analyze LinkedIn profiles and companies. It also offers a feature for finding LinkedIn company links from a list of company names.

It is important to note that some AI tools may charge a fee for their services.

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

Solution–

1. **Search for companies that use Python on their website:**
   Many companies list the programming languages they use on their website's career page or in their blog posts. You can use a search engine to find companies that use Python on their website.

2. **Use a tech stack analysis tool:**
   There are a number of tech stack analysis tools available, such as BuiltWith and StackShare. These tools can be used to identify the programming languages and technologies that a company uses.

Here are 5 companies whose tech stack is built on Python:

- Google
- Netflix
- Spotify
- Instagram
- Pinterest

Here is an example of how to identify a list of companies whose tech stack is built on Python using an AI tool:

AI tool used: Crystal

Crystal is an AI-powered tool that can be used to analyze LinkedIn profiles and companies. It also offers a feature for finding companies whose tech stack is built on Python.

Note: Crystal is a paid service. However, it offers a free trial, so you can try it out before you commit to a subscription.

Example:
To find a list of companies whose tech stack is built on Python using **BuiltWith**, you can follow these step:

Go to the BuiltWith website and enter "Python" in the search bar.

Example:
To find a list of companies whose tech stack is built on Python using **StackShare**.

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

Solution–

- There is no official LinkedIn API for sending messages to other LinkedIn users, but there are unofficial LinkedIn APIs that can be used for this purpose.
- Examples of unofficial LinkedIn APIs include the LinkedIn Messaging API from Beeper and the LinkedIn API from Phantombuster.
- To use an unofficial LinkedIn API to send messages to other LinkedIn users, you will typically need to create an account with the API provider and obtain an API key.
- Using unofficial LinkedIn APIs can be risky, so it is important to use them responsibly and to carefully read the terms of service of the API provider before using their API.

Here are some examples of how to send a message to another LinkedIn user using unofficial LinkedIn APIs:

- Using the LinkedIn Messaging API from Beeper:
- Get your API key from Beeper.
- Create a request to send a message.
- Check the response status code.

It is important to note that using unofficial LinkedIn APIs can be risky. LinkedIn may ban your account if they detect that you are using an unofficial API. It is important to use unofficial LinkedIn APIs responsibly and to carefully read the terms of service of the API provider before using their API.

Here is an example of how to send a message to another LinkedIn user using the LinkedIn Messaging API from Beeper:

Python

```
import requests

# Get your API key from Beeper
API_KEY = "YOUR_API_KEY"

# Create a request to send a message
request = requests.post(
    "https://api.beeper.com/v1/linkedin/messages/send",
    headers={
        "Authorization": "Bearer {}".format(API_KEY),
    },
    json={
        "recipient": "RECIPIENT_LINKEDIN_URL",
        "subject": "Message subject",
        "body": "Message body",
    },
)

# Check the response status code
if request.status_code == 200:
```

```
    print("Message sent successfully")
else:
    print("Error sending message: {}".format(request.content))
```