DevelopUS Guidelines

For Fast-Access Linkedin Data-Scraper

Overview: This guide will walk you through the process of setting up and using a Scrapy spider to scrape job postings from LinkedIn. The spider is designed to extract job titles, company names, job details, and other relevant information from LinkedIn's job search results.

Requirements:

- Python (3.6 or higher)
- Scrapy (2.5 or higher)

Setting the Project Up:

1. Install Scrapy using pip: < pip install scrapy >
2. Create a new directory for your Scrapy project.
3. Inside the project directory, create a new Scrapy project using : < scrapy startproject linkedin_jobs >
4. Navigate to the project directory using command : < cd root_file_directory >

Customizing the Spider:

1. Open the linkedin_jobs/spiders/linkedjobs_spider.py file.
2. Customize the api_url variable to specify your LinkedIn job search parameters, such as keywords, location, etc.
3. Customize the spider's behavior in the parse_job method according to your scraping requirements.
4. Save the changes.

Configuring Output Settings:

1. Open the linkedin_jobs/settings.py file.
2. Customize the FEED_URI variable to specify the output file path and format (e.g., CSV).
3. Optionally, customize other settings such as ROBOTSTXT_OBEY, ITEM_PIPELINES, etc.
4. Save the changes.

Running the Spider:

1. Open a terminal or command prompt.
2. Navigate to the project directory (linkedin_jobs).
3. Run the spider using the following command: < scrapy crawl linkedin_jobs >

Using Custom Search Parameters:

1. To customize the search parameters (e.g., keywords, location), modify the api_url variable in the spider file (linkedjobs_spider.py).
2. Update the URL query parameters in the api_url variable to reflect your desired search criteria.
3. Save the changes.
4. Run the spider again using the scrapy crawl linkedin_jobs command to apply the updated search parameters.
5. You can also use the custom parameter and line command like : < scrapy crawl linkedin_jobs -a keywords='data science' -a location='San Francisco, California' -a geoId='101184761' -a file_number=1 >

Viewing the Output:

- After running the spider, check the specified output directory for the scraped data file (e.g., output/jobs.csv).
- Open the output file using a text editor or spreadsheet software to view the scraped job postings.

The DevelopUS Team wishes everyone

All the Best

DevelopUS

GROW FASTER THAN EVER