Data Mining / Scraping LIVE



Who am I

Recruiter for over 20 years

Sourcing for over 15 years

Scraping for over 10 years

Since last year Registered OSINT Expert and Licenced PI



PLEASE NOTE: "Is it legal to scrape a website?"

Often, **websites** will allow third party **scraping**. For example, most **websites** give Google the express or implied permission to index their web pages. ... Although **scraping** is ubiquitous, it's not clearly **legal**. A variety of **laws** may apply to unauthorized **scraping**, including contract, copyright and trespass to chattels **laws**.

What are we talking about in the next hour?

- Data mining vs Data scraping
- Why
- Tooling
- How does it work
- Q&A



Data mining vs Data scraping

Data Mining

is a process of identifying or discovering patterns from large data sets. Data mining consists of 3 areas: content mining, usage mining, structure mining.

Usually we scrape the data so we can actually start working the data: mine it.

What we are going to use

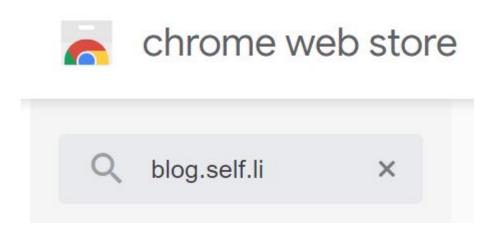


https://chrome.google.com/webstore/search/DATAMINER

Step 1 Download the extension

Step 2 Login with your Google account

Extra tooling: Auto Scroll







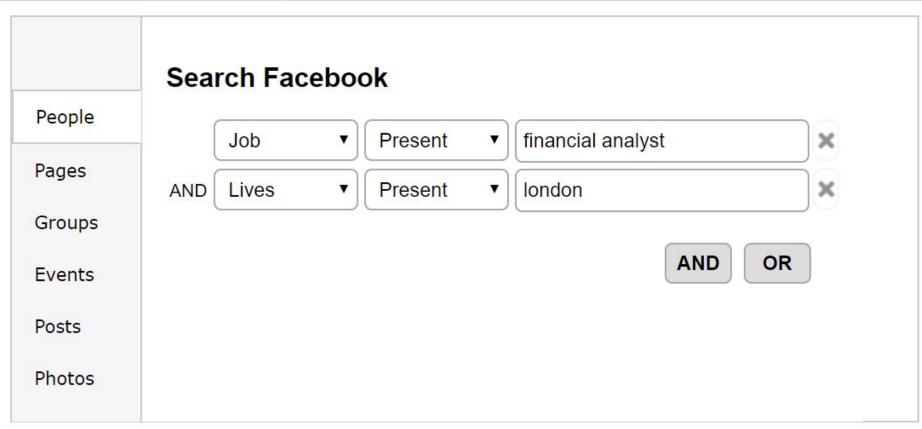
Facebook tooling



Q intelligence search ×



Facebook



Next Linkedin

STEPS in Data miner

Create a new RECIPE:

Start Choose type of page: List Page

Rows Select the part of the page that will become 1 row of your output file.

Cols Specify, name each column for the csv output file.

Nav To scrape multiple paginated pages specify how Data Miner should

navigate to next page.

Actions Data Miner can automatically click on elements on the page or scroll

down before scraping the data from that page

JS You need to be a bit more techie than I am to work on this one....

Save Name your recipe and save it for the next time

Data enrichment

Step 1 Upload your csv or xls or txt file Step 2 Pipl will tell you what to do



PIPL.COM / FILES

This is it for today!

Want more information?

Gordon @

Booleanresults.com

