LinkedIn Employee Scraper

**How to call**

python3 bridgekeeper.py

**What it does**

The script does the following.

1. Retrieves list of search queries (company name, keyword, location) from a Google Spreadsheet
2. Searches for them on google/bing/yahoo
3. Inserts results in Google Spreadsheet

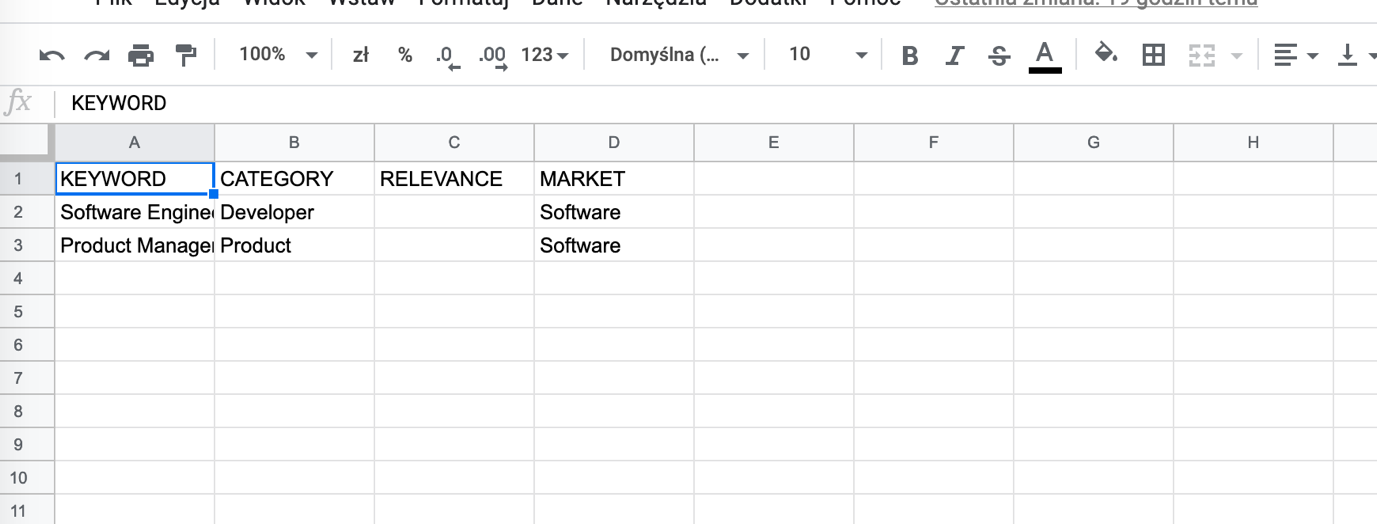
**What it uses**

The script mainly uses two Python libraries – Selenium (for fetching data from browser) and pygsheets and pandas for inserting data into google spreadsheet.

**Algorithm steps:**

**How it fetches keywords for searching**

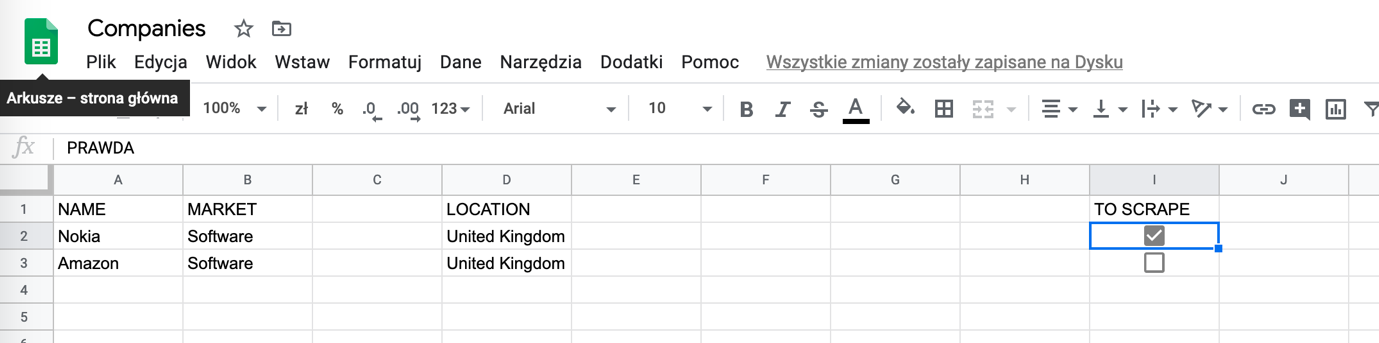
It accesses the keywords spreadsheet and scrapes keyword + market for keywords pairs for it, eg.



With this config for each company that is in ‘Software’ market we will search for employees that have roles ‘Software Engineer’ and ‘Project Manager’ listed as their current LinkedIn job.

**How it fetches companies for scraping**

It accesses companies sheet and fetches only company names that have TO SCRAPE set to TRUE.



In this example we will only search over company ‘Nokia’.

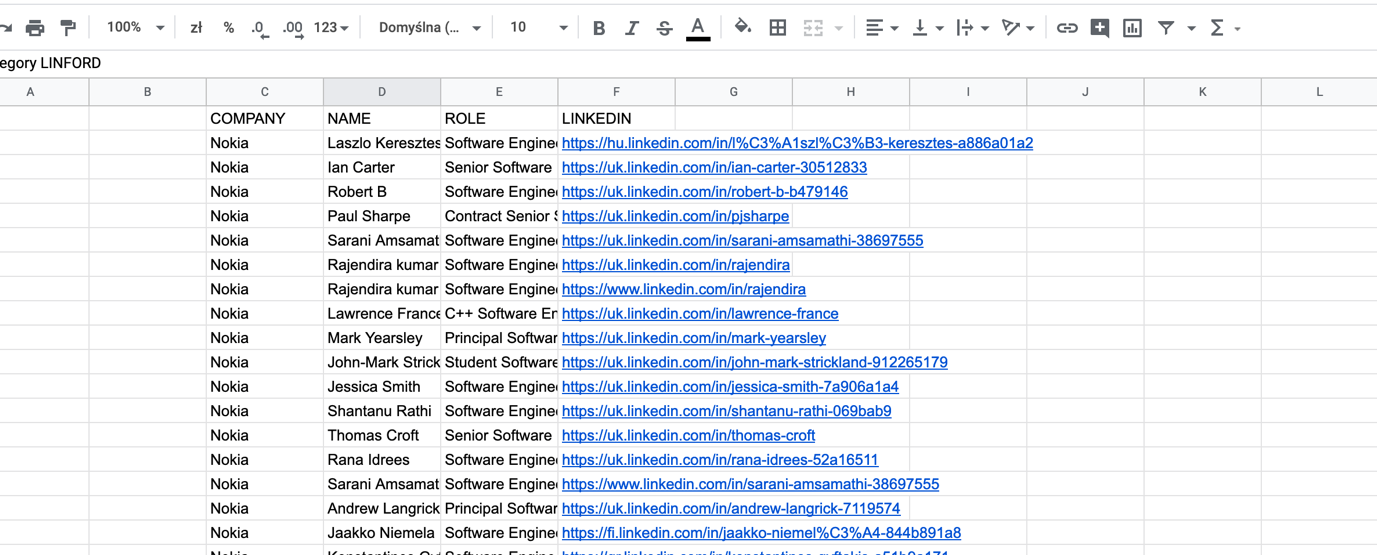
**Actual search**

For every keyword + market and for every company we search over for profiles by querying Google/Yahoo/Bing for example with this query <https://www.google.com/search?q=site%3Alinkedin.com%2Fin%2F+%22at+Amazon+%22+%22Software+Development+Engineer+%22+United+Kingdom+%22%22%22&start=10>

We fetch results until we run out of them.

**Result**

Fetched profiles are inserted into Contacts sheet.



**Setup**

To make the script use your Google Spreadsheets you need to replace creds.json file in root of script package to one that is from your google account, create it following your guide - <https://www.twilio.com/blog/2017/02/an-easy-way-to-read-and-write-to-a-google-spreadsheet-in-python.html>

The script needs three spreadsheets to exists:

**Contacts**

Required fields

|  |  |  |  |
| --- | --- | --- | --- |
| COMPANY | NAME | ROLE | LINKEDIN |

They can be in any order.

**Companies**

Required fields

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NAME | MARKET |  | LOCATION |  |  |  |  | TO SCRAPE |

They can be in any order.

**Keywords**

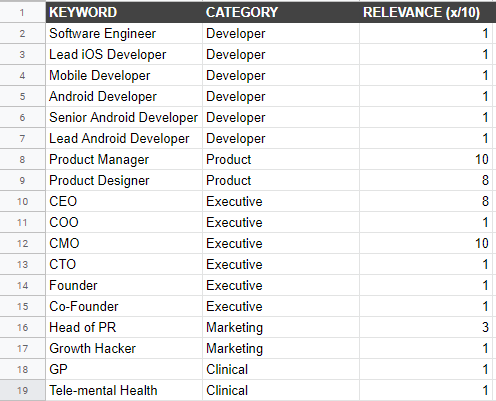
Required fields

**KEYWORD CATEGORY RELEVANCE MARKET**

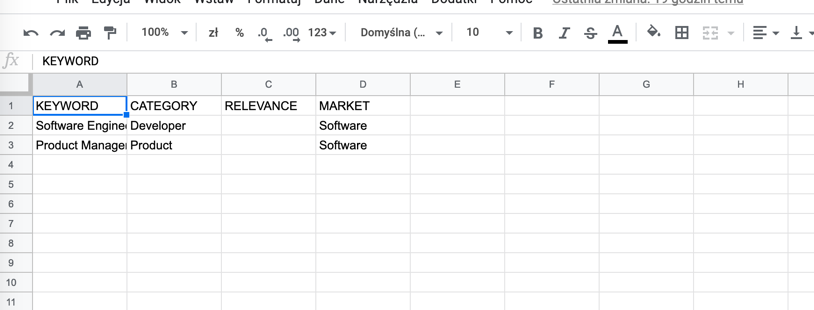
They can be in any order.

**Required sheet change**

**Keywords table in example was as follows:**

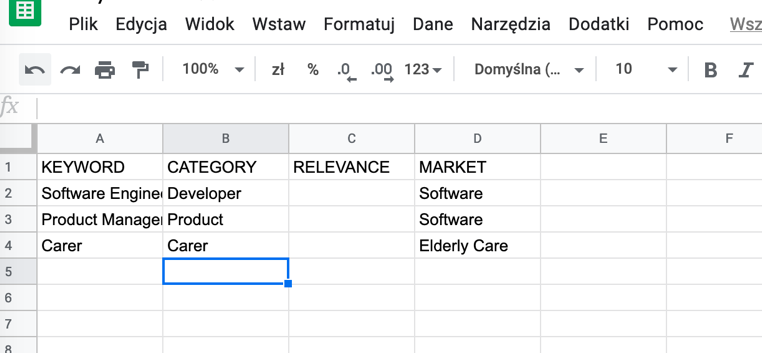


The script requires adding a new column to this sheet – called market.

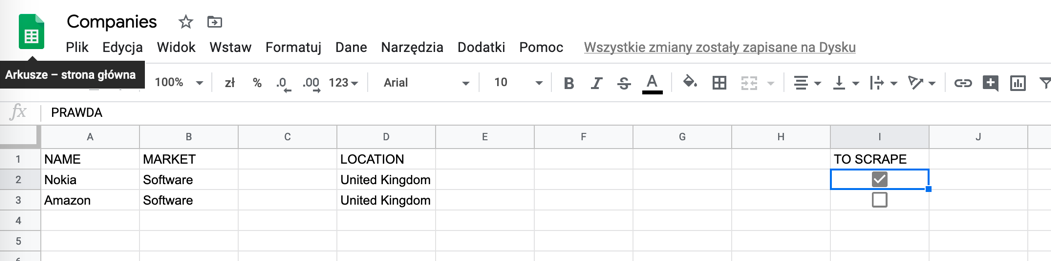


This is so that we search for given keyword only in company on given market.

For example if keywords were like this



And companies looked like this:



We wouldn’t look for Carers in Nokia and Amazon, we wouldn’t find anything there. We only search for Carer for companies whose market is Elderly Care.