

УДК 005.932:004

Якуб'як Ю. – ст. гр. СНм-61

Тернопільський національний технічний університет ім. Івана Пулюя

РЕАЛІЗАЦІЯ АЛГОРИТМУ ОПТИМІЗАЦІЇ ПРОЦЕСУ ЗБОРУ ДАНИХ ПОТЕНЦІЙНИХ КЛІЄНТІВ

Науковий керівник: Б.Б. Млинко, канд.техн.наук, доц.

Yakubyak Y.

Ternopil Ivan Puluj National Technical University

REALISATION AN ALGORITHM FOR OPTIMIZING THE PROCESS OF POTENTIAL CUSTOMERS COLLECTING DATA

Supervisor: B.B. Mlynko Ph.D., Assoc. Prof.

Ключові слова: парсинг, оптимізація, алгоритм.

Keywords: parsing, optimization, algorithm.

Інформаційні системи на сьогоднішній день мають великі перспективи практично в кожній сфері сучасного життя, зокрема в логістиці. Якщо розглядати підприємство з точки зору організованої системи, то всі її складові частини потребують безперервного зв'язку між собою для створення інтегрованої та складної структури. Впровадження інформаційних технологій в логістиці та бізнес стратегіях, безпосередньо для збору даних про потенційних покупців з метою розширення клієнтської бази є актуальною задачею.

Метою роботи було, на основі сформованого алгоритму[1] реалізувати скрипт, який надасть можливість парсити дані з застосунку Google Maps.

Для розробки скрипта за основу використаємо наступний алгоритм (рисунок 1):

1. Початковим кроком є запуск консолі;
2. Далі вводиться команда для запуску скрипта, ключові слова в лапках, цільова країна, а також кількість карток, які будуть витягнуті за один запит;
3. Скрипт формує масив запитів з міст цільової країни та ключових слів;
4. Запити з масиву проганяються в Google Maps;
5. Перевіряється чи запит містить звичайні картки чи короткі картки;
6. Якщо запит містить звичайні картки, їх записується в окремий файл, посилання з якого потім будуть вигружені і прогнані в Google Maps, після чого отримані дані будуть записані в кінцевому файлі;
7. Якщо запит містить короткі картки, їх записується в кінцевий файл.

Розробляємо скрипт, який буде реалізовувати пошук компаній по заданих ключових словах та країні. Тобто користувач вводить команду `node parser.js`, після цього в лапках вводять ключові слова, за якими буде проводитися пошук компаній, а потім буквенний код країни, по якій буде проведено пошук. Дані формують масив запитів, які будуть введені в графі пошуку.

На виході отримуємо скрипт, який вирішує наступні задачі (рисунок 2):

1. Створення масиву запитів для перебору;
2. Перевірка даних на валідність;
3. Запуск роботи скрипта;

4. Автопрокручування сторінки;
5. Обробка посилань.

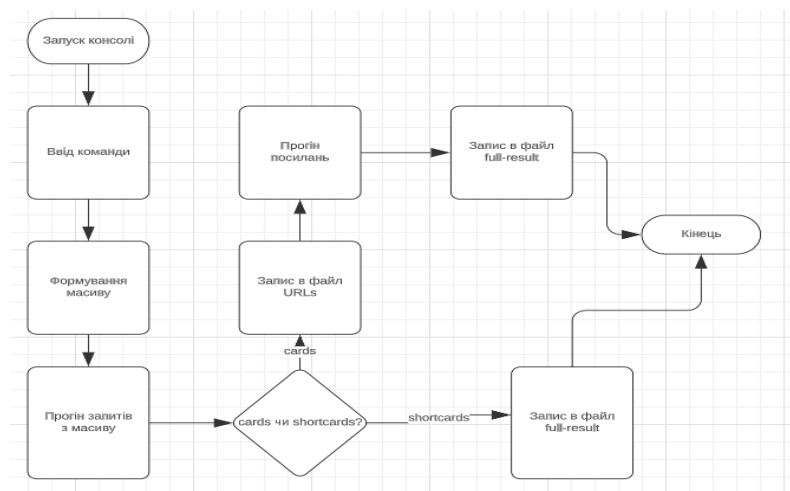


Рисунок 1 – Алгоритм роботи скрипта

Terminal output (top right):

```

[18:45:01] [
'Hypermarket, Dolnoslaskie, Poland',
'Hypermarket, Kujawsko-pomorskie, Poland',
'Hypermarket, Łódzkie, Poland',
'Hypermarket, Lubelskie, Poland',
'Hypermarket, Lubuskie, Poland',
'Hypermarket, Małopolskie, Poland',
'Hypermarket, Mazowieckie, Poland',
'Hypermarket, Opolskie, Poland',
'Hypermarket, Podkarpackie, Poland',
'Hypermarket, Podlaskie, Poland',
'Hypermarket, Pomorskie, Poland',
'Hypermarket, Śląskie, Poland',
'Hypermarket, Świętokrzyskie, Poland',
'Hypermarket, Warmińsko-mazurskie, Poland',
'Hypermarket, Wielkopolskie, Poland',
'Hypermarket, Zachodniopomorskie, Poland'
]
[18:45:17] too much places cards
[18:45:17] standart flow
[18:45:17] INFO done single
[18:45:17] Hypermarket, Dolnoslaskie, Poland
  
```

Terminal window (middle):

```

MINGW64: c:/Users/ulask/Desktop/napcmr/just-export-tools/gmap-count...
[18:46:43] DEBUG current line written
[18:46:44] 200
[18:46:44] DEBUG current line written
[18:46:44] 200
[18:46:44] DEBUG current line written
[18:46:45] 200
[18:46:45] DEBUG current line written
[18:46:45] 200
[18:46:45] DEBUG current line written
[18:46:46] 200
[18:46:46] DEBUG current line written
[18:46:46] 200
[18:46:46] DEBUG current line written
[18:46:47] 200
[18:46:47] DEBUG current line written
[18:46:47] 200
[18:46:47] DEBUG current line written
[18:46:48] 200
[18:46:48] DEBUG current line written
[18:46:48] 200
[18:46:48] DEBUG current line written
[18:46:48] 200
lines total: 320
lines done [155]
  
```

	name	activity	address	phone	site
1	Auchan	Hypermarket	Francuska 6 55-040 Bielany Wrocławskie, Poland	48224440222	null
2	Carrefour	Hypermarket	Czekoladowa 7/9 55-040 Bielany Wrocławskie, Poland	48713271650	https://www.carrefour.pl/
3	Carrefour	Hypermarket	Noworudzka 2 57-300 Kędziorów, Poland	48748627700	https://www.carrefour.pl/
4	Carrefour	Hypermarket	Podwale 22 47-400 Racibórz, Poland	48800300062	https://www.kaufland.pl/%3Fcid%3DF19054802C1200K01000W01030000
5	Kaufland Racibórz	Supermarket	al. Gen. Józefa Hallera 52 53-324 Wrocław, Poland	48713690100	https://www.carrefour.pl/
6	Carrefour	Hypermarket	aleja Piłsudskiego 84 59-220 Legnica, Poland	48767292000	https://www.carrefour.pl/
7	Carrefour	Hypermarket	Zakładowa 2-4 50-231 Wrocław, Poland	48717971700	https://www.kaufland.pl/
8	Leclerc	Hypermarket	Arkady Powstańców Warszawy 2-4, 53-333 Wrocław, Poland	48717242160	https://www.spar.pl/
9	EUROSPAR	Supermarket	Aleja Jana Pawła II 52 58-506 Jelenia Góra, Poland	48224440222	null
10	Jelenia Góra Auchan	Hypermarket	Szafarowa 55 46-100 Gliwice, Poland	48224440222	null
11	Auchan	Hypermarket	Marzaski 14 Józefa Piłsudskiego 7 48-303 Nysa, Poland	48618803000	https://www.lidl.pl/
12	Lidl	Discount supermarket	Ułpowa 144-100 Gliwice, Poland	48323027880	https://www.carrefour.pl/
13	Carrefour	Hypermarket	Plaskowa 17 50-359 Wrocław, Poland	48713438457	https://hala-targowa.pl/
14	Wrocław Market Hall	Produce market	Strzegomska 2 59-180 Ławdnica, Poland	48748627700	https://www.leclerc.pl/gazetki/
15	Leclerc Ławdnica	Hypermarket	Podhalna 36 46-335 Jastrzębie-Zdrój, Poland	48324781600	https://www.carrefour.pl/
16	Carrefour	Hypermarket	Legnicka 24 54-204 Wrocław, Poland	48713100900	https://www.carrefour.pl/
17	Carrefour	Hypermarket	Graniczna 58 54-610 Wrocław, Poland	48717710500	https://www.carrefour.pl/
18	Carrefour	Hypermarket	Bardzka 1A 50-516 Wrocław, Poland	48800300062	https://www.kaufland.pl/%3Fcid%3DF19054802C1200K01000W01030000
19	Kaufland Wrocław Huby	Supermarket	Kozelska 20 47-224 Głogów-Kościele, Poland	48717602000	https://www.carrefour.pl/
20	Carrefour	Hypermarket	OK Korona Bolesława Krzywoustego 126, 51-421 Wrocław, Poland	48224440222	null
21	Auchan	Hypermarket	Francuska 6 55-040 Bielany Wrocławskie, Poland	48224440222	null
22	Auchan	Hypermarket	Czekoladowa 7/9 55-040 Bielany Wrocławskie, Poland	48713271650	https://www.carrefour.pl/
23	Carrefour	Hypermarket	Noworudzka 2 57-300 Kędziorów, Poland	48748627700	https://www.carrefour.pl/
24	Carrefour	Hypermarket	Podwale 22 47-400 Racibórz, Poland	48800300062	https://www.kaufland.pl/%3Fcid%3DF19054802C1200K01000W01030000
25	Kaufland Racibórz	Supermarket	al. Gen. Józefa Hallera 52 53-324 Wrocław, Poland	48713690100	https://www.carrefour.pl/
26	Carrefour	Hypermarket	aleja Piłsudskiego 84 59-220 Legnica, Poland	48767292000	https://www.carrefour.pl/
27	Carrefour	Hypermarket	Zakładowa 2-4 50-231 Wrocław, Poland	48717971700	https://www.kaufland.pl/
28	Leclerc	Hypermarket	Arkady Powstańców Warszawy 2-4, 53-333 Wrocław, Poland	48717242160	https://www.spar.pl/
29	EUROSPAR	Supermarket			

Рисунки 2 – Процес роботи скрипта та зібрані дані

Висновок: в результаті дослідження, було реалізовано скрипт, який надає змогу парсити дані з сервісу Google Maps. Тобто, скрипт отримує чіткий запит для пошуку компаній в певній країні, перевіряє дані на валідність, проводить аналіз видачі Google Maps і відповідно до видачі:

1. записує вихідні дані в файл;
2. формує окремий файл з посиланнями, які необхідно опрацювати і використовує їх як вхідні дані для прогону на контактні дані потенційного клієнта.

Даний скрипт дозволяє прискорити вилучення даних з сервісу Google Maps, а також сприяє кращій конверсії, адже надає змогу обробити доволі великий масив даних

ЛІТЕРАТУРА

1. Матеріали XIII Міжнародної науково-практичної конференції молодих учених та студентів «Актуальні Задачі Сучасних Технологій». Режим доступу: https://tntu.edu.ua/storage/pages/00001070/TNTU_Aktualni_zadachi_suchasnyh_tehnologiy_2024.pdf [дата звернення: 15.03.2025].
2. Web Scraper - Free Web Scraping. Онлайн. Google Chrome - Download the Fast, Secure Browser from Google.. Режим доступу: <https://chrome.google.com/webstore/detail/web-scraper-free-web-scraper/jnhgnonknehp-ejjnehehlklplmbmhn>. [дата звернення 12.11.2024].
3. Scraping a site | Web Scraper Documentation. Онлайн. Web Scraper - The #1 web scraping extension. Режим доступу: <https://webscraper.io/documentation/scraping-a-site>. [дата звернення 12.11.2024].
4. Home | Import.io. Онлайн. Home | Import.io. [б. д.]. Режим доступу: <https://docs.import.io/>. [дата звернення 12.11.2024].
5. Wise, Alex. Netpeak Checker review from 2023: a tool for analyzing site SEO metrics. Online. Netpeak Journal is a magazine about internet marketing and online business in details. Режим доступу: <https://netpeak.net/ru/blog/obzor-netpeak-checker-2023-multifunktionalnogo-instrumenta-dlya-massovogo-analiza-i-sravneniya-saitov/>. [дата звернення 12.11.2024].
6. Інформаційні системи в логістиці : Державний університет телекомунікацій. Онлайн. Головна. Режим доступу: <https://dut.edu.ua/ua/lib/1/category/1137/view/1483>. [дата звернення 12.11.2024].
7. Парсинг даних з сайтів: що це та на чієму боці закон. Режим доступу: <https://highload.today/uk/parsing/>. [дата звернення 12.11.2024].
8. Парсинг сайтів: що це і навіщо він потрібен? Онлайн. Webpromo. Режим доступу: <https://web-promo.ua/ua/blog/parsing-sajtov-cto-eto-i-zachem-nuzhen/>. [дата звернення 12.11.2024].
9. Легальні інструменти для збору даних в інтернеті: що вони вміють та як ними користуватися. Режим доступу: <https://itc.ua/ua/partnerskij-proekt/legalni-instrumenty-dlya-zboru-danyh-v-interneti-shho-vony-vmiyut-ta-yak-nymy-korystuvatysya-detalnyj-rozbir/>. [дата звернення 12.11.2024].