

АВТОМАТИЗАЦІЯ ЗБОРУ ДАНИХ З ВИКОРИСТАННЯМ FASTAPI ТА SCRAPY НА САЙТАХ ПОШУКУ РОБОТИ

¹Вінницький національний технічний університет;

Анотація

Представлено проект до автоматизації збору даних з використанням FastAPI та Scrapy, вибрано оптимальні засоби для створення системи та розроблено архітектуру додатку. Проведено тестування з використанням Pytest.

Ключові слова: FastAPI, Scrapy, Pytest, збір даних, тестування.

Abstract

The project for automating data collection using FastAPI and Scrapy was introduced, the best tools were selected for system development, and the application architecture was developed. Testing was conducted using Pytest.

Keywords: FastAPI, Scrapy, Pytest, data collection, testing.

Вступ

Щодня по всьому світу надсилається понад 269 мільярдів електронних листів і приблизно 2,4 мільйона електронних листів за секунду. І в середньому бізнес-працівник надсилає 40 електронних листів і отримує 121 електронний лист щодня. Автоматизація такого рутинного процесу може обробляти великомасштабний збір даних із кількох джерел одночасно, і дозволить організації масштабувати зусилля зі збору даних без пропорційного збільшення робочої сили.

На сьогодні, досить багато людей стикаються з задачею збору даних, вони хочуть ефективно використовувати свій час та не гаяти його на рутинні процеси.

Збір даних передбачає збір структурованої або неструктурованої інформації з різноманітних джерел, таких як бази даних, веб-сайти, API, датчики, платформи, соціальні мережі тощо. Інструменти автоматизації можуть інтегруватися з різними джерелами даних, забезпечуючи безпроблемний пошук і консолідацію даних. Це дозволяє організаціям збирати дані з різних платформ і форматів, гармонізуючи дані для аналізу та звітності. Навіть працювати без нагляду, у неробочий час. Автоматизований процес може бути запланований для збору даних через певні проміжки часу або активований подіями, забезпечуючи безперервний збір даних без ручного втручання.

Результати дослідження

Було здійснено аналіз та порівняння різних технологій збору даних. Було встановлено, що Scrapy є одним з найефективніших інструментів для забезпечення гнучкого збору даних з веб-сайтів, а використання FastAPI забезпечує зручну та швидку взаємодію між клієнтською та серверною частинами, що є важливою умовою для створення зручного API.

Використання цих технологій дозволяє отримати доступ до великої кількості вакансій на різних сайтах для пошуку роботи, витрачаючи мінімум зусиль на збір даних.

FastAPI є одним з найбільш популярних веб-фреймворків для створення API на Python. Він має високу продуктивність завдяки використанню асинхронного Python коду, та вбудованого валідатора і серіалізатора даних Pydantic.

Крім того, FastAPI має вбудовану підтримку OpenAPI та відповідає всім сучасним стандартам безпеки, таким як автентифікація та авторизація з допомогою JWT.

Scrapy, з іншого боку, є високопродуктивним фреймворком для скрапінгу даних з веб-сайтів. Він дозволяє швидко та ефективно отримувати інформацію з веб-сторінок, обробляти їх та зберігати в зручному форматі для подальшого аналізу. Scrapy також має вбудовану підтримку асинхронного режиму роботи, що дозволяє прискорити процес збору даних.

В результаті роботи було розроблено серверну частину, яка базується на використанні FastAPI та Scrapy.

Було успішно реалізовано функціонал асинхронного режиму збору даних з різних веб-сайтів через FastAPI background tasks, що дозволяє нам регулювати навантаження на ресурси та API.

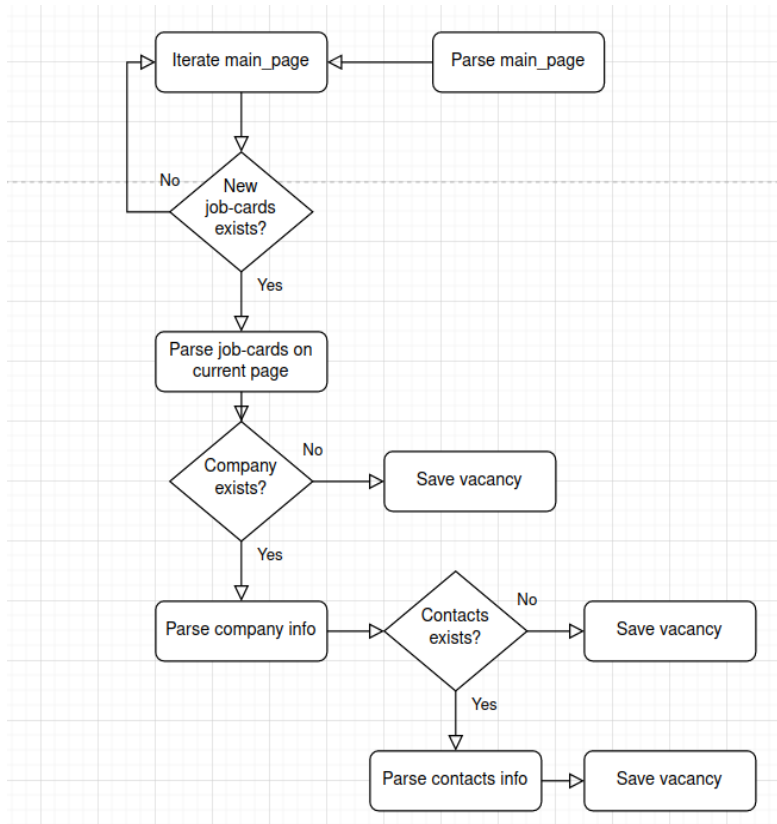


Рисунок 1 - Схема Scrapy-компонента для збирання даних

Було проведено тестування та аналіз роботи системи з використанням фреймворка Pytests. У процесі тестування було виявлено деякі помилки та неточності, які були виправлені. Загальний результат тестування був позитивним, що підтверджує стабільну та надійну роботу системи.

Висновки

Під час проведення дослідження було доведено, що використання FastAPI та Scrapy є ефективними інструментами для зменшення часу, необхідного для пошуку вакансій. Збільшення точності даних, та покращення конкурентноспроможності: використовуючи технології автоматизованого збору даних дозволяє швидко реагувати на зміни ринку праці.

Розроблене програмне забезпечення може бути корисним для кожного, хто шукає ефективні та інноваційні способи збору даних з мережі. Збір даних про різні вакансії допоможе швидше знайти собі роботу, оскільки користувачі матимуть доступ до широкого спектру вакансій з різних сайтів пошуку роботи. А також можливість зберегти їх в зручному форматі для подальшої обробки та аналізу. Це може значно зекономити час та зусилля, а також полегшити роботу рекрутерів та HR-фахівців, які шукають відповідних кандидатів для вакансій та пошукових систем.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. FastAPI [Електронний ресурс] – Режим доступу до ресурсу: <https://fastapi.tiangolo.com>.
2. Scrapy [Електронний ресурс] – Режим доступу до ресурсу: <https://docs.scrapy.org/en/latest/>.
3. Learning Python, Fifth Edition - Mark Lutz

Фариняк Дмитро Леонідович – студент групи ІАКІТ-19Б, кафедра автоматизації та інтелектуальних інформаційних технологій, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м.Вінниця, e-mail: dfarinyak@gmail.com

Богач Ілона Віталіївна – к.т.н., доцент кафедри автоматизації та інтелектуальних інформаційних технологій, факультет комп'ютерних систем і автоматики, Вінницький національний технічний університет, м.Вінниця, e-mail: ilona.bogach@gmail.com

Faryniak Dmytro L. – student of ІАСІТ-19В group, Department of Automation and Intelligent Information Technologies, Faculty of Intelligent Information Technology and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: dfarinyak@gmail.com

Bogach Ilona V. – PhD, Associate Professor of Automation and Intelligent Information Technologies, Faculty of Computer Systems and Automatics Vinnytsia National Technical University, Vinnytsia, e-mail: ilona.bogach@gmail.com