

# ПОРІВНЯЛЬНИЙ АНАЛІЗ АСИНХРОННОГО ТА СИНХРОННОГО ЗБОРУ ДАНИХ НА PYTHON

<sup>1</sup>Вінницький національний технічний університет

## *Анотація*

*Представлено проект для оцінки переваг та недоліків автоматизованого збору даних з використанням асинхронного фреймворку FastAPI, синхронного фреймворку Django та Scrapy. Вибрано оптимальні засоби для створення системи та розроблено архітектуру додатків. Проведено тестування з використанням Pytest.*

**Ключові слова:** FastAPI, Django, Scrapy, Pytest, збір даних, асинхронність, синхронність.

## **Abstract**

*A project is presented to assess the advantages and disadvantages of automated data collection using the asynchronous FastAPI framework, the synchronous Django framework, and Scrapy, the optimal tools for creating the system are selected, and the application architecture is developed. Tested using Pytest.*

**Keywords:** FastAPI, Django Scrapy, Pytest, data collection, asynchronous, synchronous.

## **Вступ**

Збір даних є важливою складовою багатьох проектів на Python. Існує два основних методи збору даних: синхронний та асинхронний. 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.

Задача автоматизації збору даних стикається з багатьма людьми, які хочуть ефективно використовувати свій час та гроші, не витрачаючи їх на неефективні рішення. У даному проекті розглянуто порівняльний аналіз між використанням FastAPI та Scrapy і Django та Scrapy для автоматизації збору даних.

## **Результати досліджень**

В ході дослідження було проведено серію тестів для порівняння асинхронного та синхронного методів збору даних. Результати тестів показали, що асинхронний метод збирає дані значно швидше, ніж синхронний.

Під час розробки було здійснено порівняльний аналіз різних технологій збору даних. Було встановлено, що Scrapy є одним з найефективніших інструментів для забезпечення гнучкого збору даних та їх обробки з веб-сайтів. Використання FastAPI забезпечує зручну та швидку взаємодію між клієнтською та серверною частинами, що є важливою умовою для створення зручного API.

В рамках проекту було розроблено серверну частину, яка базується на використанні FastAPI та Scrapy і Django та Scrapy.

Django це високорівневий веб-фреймворк, що пропонує широкий спектр функцій, які роблять його ідеальним вибором для широкого кола проектів, від невеликих сайтів до масштабних веб-додатків, може легко масштабуватися для підтримки великої кількості користувачів і трафіку, Django пропонує "batteries included" підхід, що включає в себе готові до використання компоненти для аутентифікації, адміністрування, кешування та багато іншого. Це дозволяє розробникам зосередитися на створенні функціональних можливостей свого проекту, а не на написанні коду для типових завдань.

FastAPI є одним з найбільш популярних веб-фреймворків для створення API на Python. Він має високу продуктивність завдяки використанню асинхронного кодування та вбудованого валидатора даних Pydantic. Крім того, FastAPI має вбудовану підтримку OpenAPI та відповідає всім сучасним стандартам безпеки, таким як автентифікація та авторизація 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 її та зберігати в зручний формат для подальшого аналізу. Scrapy також має вбудовану підтримку асинхронного кодування, що дозволяє прискорити процес збору даних.

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

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 є більш ефективним інструментом для зменшення часу, необхідного для ефективного автоматизованого збору даних ніж Django за рахунок асинхронної роботи. Збільшення точності даних, та покращення конкурентноспроможності: використовуючи технології автоматизованого збору даних дозволяє швидко реагувати на цільову рекламу, зміни ринку праці, статистику.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. FastAPI [Електронний ресурс] – Режим доступу до ресурсу: <https://fastapi.tiangolo.com>.
2. Django [Електронний ресурс] – Режим доступу до ресурсу: <https://www.djangoproject.com/>.
3. DRF [Електронний ресурс] – Режим доступу до ресурсу: <https://www.django-rest-framework.org/>.
4. Scrapy [Електронний ресурс] – Режим доступу до ресурсу: <https://docs.scrapy.org/en/latest/>.
5. Using Asyncio in Python - Caleb Hattingh

***Фариняк Дмитро Леонідович** – студент групи ЗАКІТР-23м, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м.Вінниця, e-mail: dfarinyak@gmail.com*

***Бісікало Олег Володимирович** – д-р. техн. наук, професор кафедри автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет, м. Вінниця*

***Faryniak Dmytro Leonidovych** – student of 3ACIT-23m group, Faculty of Intelligent Information Technology and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: dfarinyak@gmail.com*

***Bisikalo Oleg Volodymyrovych** – Dr. of Technical Sciences and Professor of the Department of Automation and Intelligent Information Technologies at Vinnytsia National Technical University, Vinnytsia*