

# ПРОГРАМНИЙ ЗАСІБ ПОШУКУ ТА ВІДБОРУ ДАНИХ З ОНЛАЙН-ПЛАТФОРМ ОРЕНДИ НЕРУХОМОСТІ

Вінницький національний технічний університет

## **Анотація**

*Розроблено програмний засіб автоматичного збору та аналізу даних по ринок оренди нерухомості України.*

**Ключові слова:** веб-додаток, парсинг даних, оренда нерухомості.

## **Abstract**

*A software tool for automatic collection and analysis of data on the real estate rental market of Ukraine has been developed.*

**Keywords:** web application, data parsing, real estate rental.

## **Вступ**

У сучасному світі, де величезний обсяг інформації доступний онлайн, завдання автоматичного збору та аналізу даних з веб-ресурсів стає все більш важливим і актуальним завданням. За оцінками центру Statista [1] у 2020 році обсяг згенерованої інформації у світі досяг 47 зетабайт і це число збільшиться до 175 зетабайт до кінця 2025 року.

З кожним днем кількість даних про різні об'єкти, у томі числі і для оренди, зростає. За допомогою програмного засобу, який може агрегувати дані з різних платформ, користувачі можуть швидше та зручніше отримувати доступ до широкого вибору пропозицій. Метою роботи є вдосконалення методу автоматичного збору та аналізу даних про ринок оренди нерухомості шляхом використання сучасних веб-технологій.

## **Результати дослідження**

У створеному програмному засобі використано автоматичний збір даних з різних веб-ресурсів та здійснюється оптимізація процесу пошуку нерухомості шляхом застосування сучасних веб-технологій та фреймворків, таких як Java, JavaScript, Spring та React [2]. Основною задачею є розробка архітектури програми збору даних із відкритих джерел у мережі Інтернет, їх зберігання та подальша обробка для створення системи, яка забезпечує користувачам можливість знайти бажану інформацію про оренду нерухомості на основі їхніх вимог та вподобань. Для збору та початкової обробки даних були використані збирач даних, обробник даних та модуль зберігання даних.

В роботі збирача даних, який є парсером, виконано процес надсилання HTTP-запитів до зовнішніх джерел інформації, таких як інтернет-магазини, для отримання даних. Для отримання цих даних парсер використовує зовнішні API інтернет-магазинів або системи аналізу вмісту веб-сторінок [3]. Після успішного збору необхідної інформації, отриманий набір даних передається до модуля зберігання у системі для подальшого зберігання та обробки.

Модуль зберігання даних - це компонент системи, що відповідає за зберігання всієї необхідної для системи інформації. Цей модуль працює з різноманітними форматами даних, включаючи як чітко структуровані формати, наприклад, JSON та табличні дані, так і менш структуровані дані, такі як фотографії товарів. У своїй роботі модуль зберігання даних забезпечує інтерфейс для внутрішніх модулів, наприклад, обробника даних, а також для зовнішніх компонентів системи. Цей модуль забезпечує зручний доступ до даних, які зберігаються в ньому, іншим компонентам системи.

Обробник даних та класифікатор об'єктів виконує важливу роль у системі. Основна мета цих модулів полягає у перетворенні різноманітних слабо структурованих даних в однорідний формат. Окрім цього, обробник даних та класифікатор займаються ідентифікацією однакових об'єктів. Для виконання цих завдань використовуються необроблені дані, які зберігаються в data lake, та після їх аналізу створюють структуровану інформацію про об'єкти. Такий підхід дозволяє системі зручно та

ефективно опрацьовувати дані, роблячи їх доступними для подальшого використання у єдиному форматі.

Модуль геокодингу потрібен для обробки інформації про адреси фізичних магазинів та оцінки варіантів доставки або самовивозу з них. Для своєї роботи він використовує публічні геокодингові сервіси та записує отриману інформацію в озеро даних.

Перед тим як була виконана розробка структури кубів для аналізу ринку оренди нерухомості, була створена схема бази даних, у якій накопичуються мета-дані, на основі яких створюються куби. Для збереження даних про об'єкти оренди нерухомості було сформовано чотири таблиці. У першій таблиці зберігаються дані про міста, у яких знаходяться об'єкти нерухомості. У другій таблиці зберігаються дані про будівлі, у яких знаходяться квартири. У третій таблиці зберігаються дані про квартири. У четвертій таблиці зберігаються дані про усі операції, що проводились над певним об'єктом оренди.

В результаті розробки було створено програмний продукт, призначений для автоматичного збору даних та аналізу ринку оренди нерухомості. Його функціональність включає збір даних через веб-скрапінг, обробку цих даних з внесенням до бази даних. Він також відображає актуальні оголошення, створює OLAP куби для аналізу даних та візуалізує статистичні дані з використанням цих кубів [4]. Цей програмний засіб спрямований на забезпечення користувачів зручним та інформативним способом отримання інформації про ринок оренди нерухомості для подальшого аналізу та прийняття рішень.

При розробці програмного застосунку були використані такі програмні засоби, як Spring, бібліотека Jsooup, набір програмного забезпечення ХАМРР, MySQL для управління базами даних та JPA, що спрощує роботу з базами даних.

Бібліотека Jsooup була використана як потужний інструмент у Java для парсингу HTML та обробки даних на основі HTML-документів. Jsooup було обрано для розробки парсера даних з кількох причин. Вона має простий та легкий у використанні API, що спрощує витягування інформації з HTML-сторінок. Також вона дозволяє легко маніпулювати структурою та елементами HTML-документу через уніфікований DOM-інтерфейс. Крім цього, Jsooup надає можливість вибирати елементи з HTML-документу за допомогою CSS-подібного синтаксису. Було також використано набір програмного забезпечення ХАМРР, який включає в себе всі необхідні компоненти для створення локального веб-сервера або локального сервера для розробки веб-додатків.

## Висновок

В результаті було розроблено систему збору та обробки інформації, яка дозволяє користувачам приймати рішення щодо вибору оптимальної пропозиції за їх критеріями та досліджено можливості поліпшення досвіду користувача під час пошуку даних про оренду нерухомості шляхом інтеграції інтуїтивних інтерфейсів та використання рекомендаційних систем.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Chart: Global Data Creation is About to Explode | Statista [Електронний ресурс], – Режим доступу: <https://www.statista.com/chart/17727/global-data-creation-forecasts/>
2. The Ultimate Guide to Web Scraping with Python [Електронний ресурс], – Режим доступу: (<https://www.learn datasci.com/tutorials/ultimate-guide-web-scraping-w-python-requests-and-beautifulsoup/>).
3. Patel J. Introduction to Web Scraping / Jay Patel // Getting Structured Data from the Internet / Jay Patel. – Berkeley, 2020. – P. 1–30.
4. Dharmendra B. An overview of data warehousing and OLAP technology / Biswas Dharmendra // Data Warehouses and OLAP: Concepts, Architectures, and Solutions / Biswas Dharmendra. – Mumbai, 2020. – P. 65–74.

**Мартинів Павло Геннадійович** – студент 2KI-22m, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, email: [martynovpasha0000@gmail.com](mailto:martynovpasha0000@gmail.com).

**Очкуров Микола Андрійович** — старший викладач кафедри обчислювальної техніки, Вінницький національний технічний університет, м. Вінниця.

**Martynov Pavlo Hennadiiovych** - student of 2KI-22m, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, email: [martynovpasha0000@gmail.com](mailto:martynovpasha0000@gmail.com).

**Ochkurov Mykola A.** — Senior lecturer of the Computer Techniques Chair, Vinnytsia National Technical University, Vinnytsia.