

УДК 004.77:[339:780.6]

РОЗГЛЯД ПИТАННЯ ОБ'ЄДНАННЯ ТОВАРІВ З РІЗНИХ ПЛАТФОРМ НА ПРИКЛАДІ ОНЛАЙН МАГАЗИНІВ МУЗИЧНИХ ІНСТРУМЕНТІВ

Топчій М.А.

email: mykyta.topchii@nure.ua

Науковий керівник – к.т.н., доц. Яковлева О.В.

Харківський національний університет радіоелектроніки, каф. ІНФ
м. Харків, Україна

This study examines the architecture, analysis, and design of a web platform for aggregating products in the music equipment industry. The purpose of the work is to develop a web-application for proving information and advertising services using the web scraping method to collect valuable data about musical instruments. In the process of work, web scraping technologies based on Node.js library called Puppeteer were analyzed. Other web scraping libraries were considered and possible architectural solutions for the system of combining product offers implementation were investigated. As a result of the work, a comprehensive software system was created for searching, storing and displaying a compilation of data from online stores of musical instruments.

У сучасному світі спостерігається стрімке зростання обсягів даних та онлайн-пропозицій, зокрема у сфері торгівлі музичними інструментами. Велику кількість інтернет-крамниць не завжди зручно оглядати користувачеві, оскільки інформація про однакові товари може бути розпорошена по різних платформах.

Робота спрямована на створення інтегрованого веб-сервісу, що об'єднуватиме дані про товари з різних онлайн-магазинів музичних інструментів.

Мета роботи – розробити веб-платформу, яка виконуватиме функції автоматичного пошуку та порівняння музичних інструментів із різних онлайн-крамниць, допомагатиме потенційним клієнтам швидко орієнтуватися в асортименті та знаходити найкращі пропозиції за ціною й характеристиками.

Для досягнення поставленої мети було сформульовано такі основні завдання:

- дослідити та порівняти існуючі методи вебскрейпінгу для збирання даних про товари (DOM parsing, API integration, тощо);
- розробити методику обробки та порівняння товарних пропозицій (product matching) з метою виявлення дублювання або відмінностей у назвах та характеристиках;
- спроектувати архітектуру та реалізувати клієнт-серверний застосунок;
- забезпечити можливість пагінації, фільтрації та сортування товарів за низкою характеристик (бренд, тип інструмента, цінові межі тощо);

– перевірити працездатність та ефективність створеного сервісу на тестовому наборі реальних даних.

Для вирішення задачі вебскрейпінгу, потрібно обрати інструмент автоматичного парсингу дерева DOM.

DOM – це стандарт W3C (World Wide Consortium) [1]. DOM визначає стандарт для доступу до документів: «Модель об'єкта документа W3C (DOM) – це платформа та нейтральний інтерфейс, який дозволяє програмам та сценаріям динамічно отримувати доступ та оновлювати вміст, структуру та стиль документа».

Для зберігання та швидкої обробки отриманих даних використовується хмарна MongoDB. Така схема дає змогу масштабувати застосунок без втрати швидкодії, адже NoSQL-база гнучко підтримує динамічну структуру документів.

Нижче наведено схематичний приклад алгоритму автоматичного вебскрейпінгу (рис. 1).

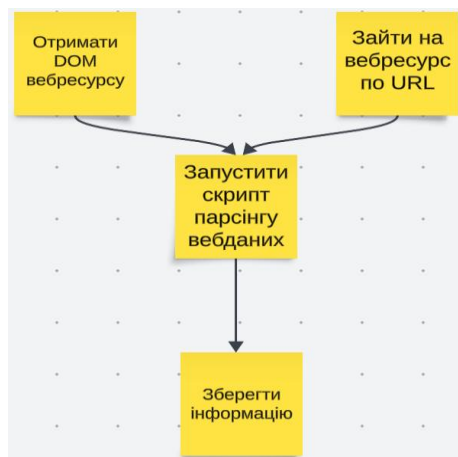


Рисунок 1 – Алгоритм отримання даних за допомогою вебскрейпінгу

Product matching, або визначення однакових товарів, є критичною задачею у багатьох системах e-commerce. Це завдання полягає у виявленні однакових або подібних товарів, представлених на різних онлайн-платформах для їх подальшого об'єднання та використання інформації в різних цілях бізнес-логіки. В контексті вебплатформи для об'єднання товарних пропозицій від онлайн-крамниць музичних інструментів ефективно вирішення цієї проблеми є надзвичайно важливим для забезпечення точності даних, що відображаються користувачам.

Для вирішення проблеми product matching, можна виділити три основних методи:

1. Текстові методи на основі евристичного підходу – це такі методи, які порівнюють будь-яку текстову інформацію на основі спеціалізованого алгоритму в контексті певної задачі. Суть алгоритму полягає у різноманітних методах порівнянь символів, слів, комбінацій.

2. Текстові методи на основі порівняння ембедінгів – це такі методи, що використовують ембедінги для порівняння текстової інформації [2]. Ембедінги – це векторне чисельне уявлення тексту для його подальшої обробки комп'ютером.

3. Методи на основі порівняння візуального контенту – це така група методів, що визначає схожість двох зображень на основі схожості їх ознак [3, 4], наприклад, ембедінгів цих зображень.

4. Гібридні методи – це поєднання всіх або декількох методів для визначення більш детального результату. Наприклад, за допомоги коефіцієнту, що визначатиме вагу кожного з методу складової гібридного підходу.

Для того, щоб обрати найефективніший метод вирішення проблеми збігання однакових товарів з характерними рисами, що відрізняються, потрібно дослідити кожний з наведених вище методів на тому самому датасеті. Для подальшої роботи планується створення застосунку на базі мікросервісної архітектури [5], що буде складатися з парсеру, текстової моделі машинного навчання, API та клієнта.

Список використаних джерел:

1. Документація W3. URL: <https://www.w3schools.com/> (дата звернення: 04.03.2025).

2. Application a Committee of Kohonen Neural Networks to Training of Image Classifier Based on Description of Descriptors Set / V. Gorokhovatskyi et al. IEEE Access. 2024. vol. 12, pp. 73376-73385. <https://doi.org/10.1109/access.2024.3404371>.

3. Gorokhovatskyi O., Yakovleva O. Medoids as a packing of ORB image descriptors. *Advanced Information Systems*. 2024. Vol. 8, no. 2. P. 5–11. URL: <https://doi.org/10.20998/2522-9052.2024.2.01>.

4. Yakovleva O., Matúšová S., Koshel V. Implementation of AI approaches in current tools for managing image collections to improve the search capabilities // *Proceedings of the IV Correspondence International Scientific and Practical Conference «Science in motion: classic and modern tools and methods in scientific investigations» in Periodical International scientific journal «Grail of science»*. February 21, 2025. Vinnytsia, Ukraine - Vienna, Austria. 2025. Vol. 49. P. 752–755. <https://doi.org/10.36074/grail-of-science.21.02.2025.096>.

5. Shelest V., Yakovleva O. Research on selecting web application architecture based on the analysis of applied requirements. *Computer integrated technologies of automation of technological processes* : Proceedings of the X International Scientific and Practical Conference. November 5-8, 2024. Hamburg, Germany, 2024. P. 46-54. <https://doi.org/10.46299/ISG.2024.2.10>.