

УДК 004.738.5:004.415.2

АНАЛІЗ АРХІТЕКТУРНИХ ТА ТЕХНОЛОГІЧНИХ ОСОБЛИВОСТЕЙ ВЕБСАЙТІВ ПРОДУКТОВИХ МЕРЕЖ АТБ ТА СІЛЬПО

Козинець Б.С. – гр. БІП-21, бакалавр, bohthankozinets@gmail.com

Новак Д.С. – к.т.н., доцент, novak.ds@knutd.edu.ua

Скідан В.В. – к.т.н., завідувач кафедри, skidan.vv@knutd.edu.ua

Київський національний університет технологій та дизайну

Метою роботи є проведення аналізу архітектурних і технологічних особливостей вебсайтів продуктових мереж АТБ та Сільпо з метою розробки ефективних, гнучких і масштабованих парсерів для збору товарної інформації, а також побудови архітектури програмної системи, що забезпечує надійну обробку даних із різних джерел.

Виявлено суттєві відмінності у структурі сайтів: якщо atbmarket.com має відносно просту, статичну HTML-архітектуру з чіткою ієрархією категорій, традиційною пагінацією та мінімальним використанням JavaScript, то silpo.ua базується на сучасних фронтенд-фреймворках, використовує динамічне завантаження контенту, нескінченний скролінг та багаторівневі спливаючі меню. Ці фактори суттєво впливають на підходи до парсингу [1].

На основі аналізу структур сайтів був розроблений універсальний базовий клас BaseParser, який інкапсулює спільну логіку парсингу: взаємодію з браузером, очікування елементів, логування, обробку помилок та реалізацію шаблонного методу (template method). Такий підхід забезпечує розширюваність і повторне використання коду в конкретних реалізаціях, як-от ATBParser і SilpoParser. Реалізовані механізми очікування наявності елементів використовують комбінацію явних і неявних затримок, що особливо критично для сайтів із динамічним завантаженням контенту.

Для сайту АТБ розроблено спеціалізований парсер, який ефективно взаємодіє з навігаційною ієрархією сайту, обробляє сторінки з пагінацією, аналізує формати цін (за одиницю або вагу), та фільтрує акційні пропозиції. Окремий акцент зроблено на метод `_try_selectors`, який дає змогу підвищити надійність парсингу за рахунок використання альтернативних CSS-селекторів при зміні DOM-структури.

Парсер для Сільпо враховує складніший інтерфейс: реалізовано механізм емуляції нескінченного скролінгу, що відстежує повне завантаження товарів у категоріях. Особливу увагу приділено динамічним атрибутам товарних карток, які містять інформацію про акції, бонуси програми лояльності «Власний рахунок» та наявність товарів у магазинах. Для обробки складних сценаріїв впроваджено поведінкові шаблони: ротація User-Agent, випадкові кліки та імітація дій користувача.

Для організації архітектури системи розроблено UML-діаграму класів (рис.1), що демонструє взаємозв'язки між основними компонентами:

Платформа: ІНФОРМАЦІЙНІ СИСТЕМИ. КОМП'ЮТЕРНІ СИСТЕМИ ТА МЕРЕЖІ. ТЕХНОЛОГІЇ INTERNET OF THINGS TA SMART-СИСТЕМИ

BaseParser, ATBParser, SilpoParser, DataManager, ParserManager, ParserFactory, PriceAnalyzer та Visualization.

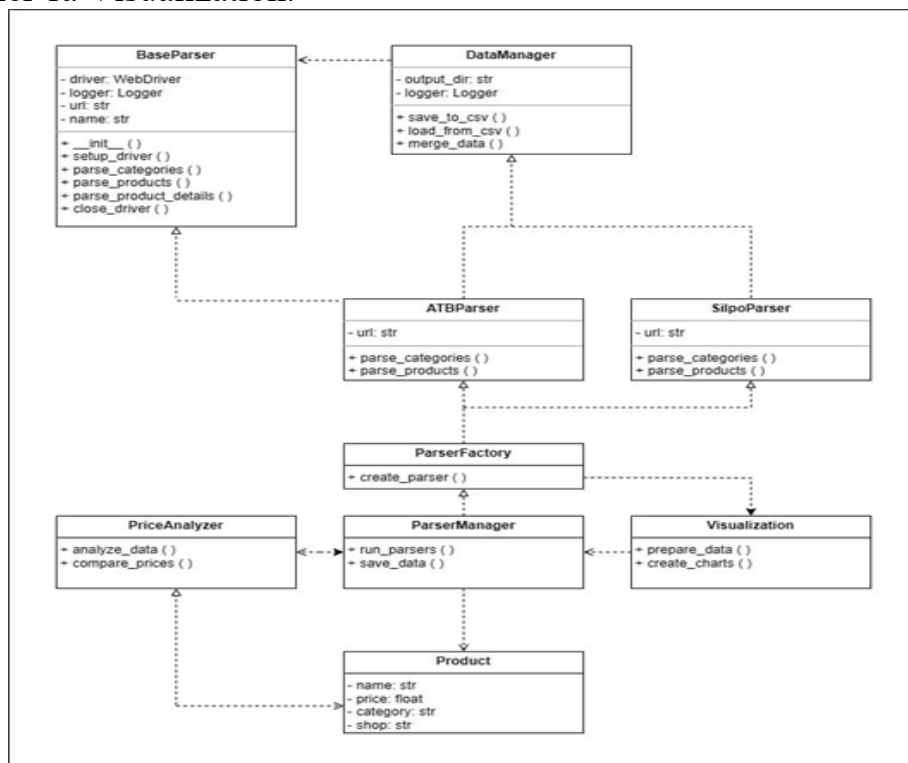


Рисунок 1 – UML-діаграма класів

Уся система побудована з урахуванням принципів поліморфізму, інверсії залежностей і розділення відповідальностей. Кожен парсер зберігає незалежність від конкретної реалізації менеджера даних, а створення екземплярів реалізовано через ParserFactory. Модуль Visualization використовує дані, зібрані парсерами через ParserManager, що спрощує інтеграцію результатів з іншими компонентами.

Висновок. Проведений аналіз показав суттєві відмінності у підходах до побудови вебсайтів: від статичної HTML-структури АТБ до динамічного контенту на основі сучасних фронтенд-фреймворків у Сільпо. Це обумовило необхідність розробки адаптивних механізмів парсингу, включаючи емуляцію нескінченного скролінгу, обробку динамічних атрибутів і ротацію User-Agent.

Запропонована архітектура системи з використанням патернів проектування та принципів ООП забезпечує модульність, гнучкість і зручність подальшої інтеграції. Результати дослідження можуть бути використані для побудови універсальних парсерів інших e-commerce платформ, що використовують різні технології представлення контенту.

Л і т е р а т у р а

1. Volkov, A. S., & Chernenky, M. V. (2024, February). Generalized algorithm for website parsing. In 2024 6th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE) (pp. 1-5). IEEE.