

ІНТЕГРАЦІЯ СИСТЕМ АНАЛІЗУ РИНКУ З ВІДКРИТИМИ ДЖЕРЕЛАМИ ДАНИХ

Вінницький національний технічний університет

Анотація. Розглянуто підхід до інтеграції систем аналізу ринку з відкритими джерелами статистичних і економічних даних. Запропоновано архітектуру, що поєднує адаптери зовнішніх API, процеси вилучення, перетворення та завантаження даних, централізоване сховище й аналітичний модуль. Визначено основні проблеми сумісності форматів, семантичної узгодженості, актуальності та якості даних. Запропонований підхід дає змогу автоматизувати формування показників привабливості й ризику ринку та забезпечити повторне використання інтеграційних компонентів.

Ключові слова: інтеграція IT-систем, відкриті дані, аналіз ринку, REST API, SDMX, ETL, сховище даних.

Abstract. The paper considers an approach to integrating market analysis systems with open sources of statistical and economic data. An architecture combining external API adapters, extract-transform-load processes, a centralized data repository, and an analytical module is proposed. The main issues of format interoperability, semantic consistency, timeliness, and data quality are identified. The proposed approach enables automated calculation of market attractiveness and risk indicators and supports the reuse of integration components.

Keywords: IT systems integration, open data, market analysis, REST API, SDMX, ETL, data warehouse.

Вступ

Прийняття рішень щодо виходу компанії на новий ринок, розширення асортименту або запуску продукту потребує комплексного аналізу макроекономічних, демографічних, галузевих і конкурентних показників. Значна частина таких відомостей доступна у відкритих джерелах: на порталах державної статистики, міжнародних організацій, органів Європейського Союзу та у відкритих реєстрах. Зокрема, портал data.euroora.eu забезпечує централізований доступ до наборів відкритих даних європейських установ і держав [1], World Bank Indicators API надає програмний доступ до великої кількості часових рядів [2], а Eurostat та OECD підтримують автоматизоване отримання статистичних даних через API і стандарт SDMX [3, 4].

Проблема полягає в тому, що джерела використовують різні протоколи, формати, структури показників, одиниці вимірювання, часові інтервали та правила оновлення. Тому без спеціального інтеграційного шару аналітична система не може стабільно отримувати й порівнювати дані. Метою роботи є формування архітектурного підходу до інтеграції системи аналізу ринку з відкритими джерелами даних, який забезпечує автоматизацію збору, уніфікацію, перевірку якості та подальше аналітичне використання інформації.

Архітектура інтеграційної системи

Доцільно використовувати багаторівневу архітектуру, в якій взаємодія із зовнішніми постачальниками відокремлена від бізнес-логіки системи. Для кожного джерела створюється адаптер, що інкапсулює адресу кінцевої точки, параметри запиту, автентифікацію, обмеження частоти звернень та перетворення відповіді у внутрішню модель. Такий підхід зменшує залежність системи від змін конкретного API та відповідає принципам слабкого зв'язування інтегрованих компонентів [5].

Загальну схему запропонованої інтеграції наведено на рис. 1. Вхідними даними можуть бути відповіді REST API у форматах JSON або XML, файли CSV, статистичні повідомлення SDMX та метадані каталогів відкритих даних. Після вилучення дані проходять очищення, нормалізацію й семантичне зіставлення, зберігаються у централізованому сховищі, після чого використовуються алгоритмами оцінювання ринку та засобами візуалізації.

Інтеграційний контур системи аналізу ринку

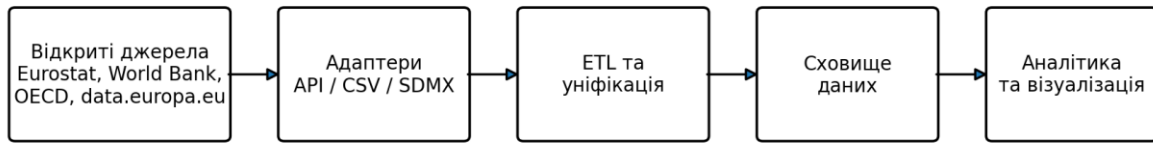


Рис. 1 – Архітектура інтеграції системи аналізу ринку з відкритими джерелами даних

Інтеграційний конвеєр доцільно реалізувати у вигляді послідовності ETL- або ELT-процесів. На етапі вилучення система отримує нові або змінені записи. На етапі перетворення узгоджуються назви країн і регіонів, коди валют, періоди спостережень, одиниці вимірювання та структури класифікаторів. На етапі завантаження зберігаються як нормалізовані значення, так і службові атрибути: джерело, дата отримання, версія набору, ліцензія, рівень довіри та ознака повноти. Це забезпечує простежуваність походження даних і можливість повторного розрахунку показників.

Джерела та узгодження даних

Для аналізу ринку система може використовувати показники валового внутрішнього продукту, інфляції, доходів населення, чисельності цільових груп, зовнішньої торгівлі, зайнятості, інвестицій, цифровізації та динаміки цін. Різні постачальники можуть публікувати близькі за змістом показники з різними кодами й методиками. Тому необхідний внутрішній каталог показників, у якому кожному бізнес-поняттю відповідають джерельні коди, опис методології, одиниця вимірювання та допустимі правила перетворення.

Таблиця 1 – Приклади відкритих джерел для інтеграції

Джерело	Спосіб доступу	Типові дані	Особливості інтеграції
data.europa.eu	Каталог, API, RDF/SPARQL	Державні та галузеві набори даних	Різні ліцензії та формати ресурсів
World Bank	REST API, JSON/XML	Макроекономічні та соціальні індикатори	Часові ряди, коди країн та індикаторів
Eurostat	REST, JSON-stat, SDMX	Статистика ЄС, торгівля, населення, ціни	Багатовимірні набори та класифікатори
OECD	REST API на основі SDMX	Економічні, галузеві й регуляторні показники	Необхідність роботи зі структурами SDMX

Для запобігання помилковим порівнянням необхідно контролювати не лише технічний формат, а й семантику. Наприклад, показник може бути поданий у поточних або постійних цінах, у національній валюті або доларах США, за календарний чи фінансовий рік. У внутрішній моделі потрібно зберігати код методики та виконувати перетворення лише за визначеними правилами. Під час обробки пропусків система має позначати відсутні значення, а не автоматично прирівнювати їх до нуля.

Оцінювання ринкової привабливості та ризику

Після інтеграції даних аналітичний модуль може формувати узагальнений профіль ринку. Для цього показники групуються за напрямками: місткість і темпи зростання, платоспроможність попиту, конкурентне середовище, регуляторні обмеження, макроекономічна стабільність та доступність інфраструктури. Значення нормалізуються до спільної шкали, а ваги задаються відповідно до виду

продукту та стратегії компанії. Важливо, щоб система показувала не лише підсумковий рейтинг, а й внесок кожного показника, джерело та дату його оновлення.

Ризик інтеграційної помилки також повинен враховуватися. Якщо ключовий показник застарілий, отриманий лише з одного джерела або має значну частку пропусків, рівень довіри до висновку зменшується. Тому разом із бізнес-оцінкою доцільно обчислювати індикатор якості даних, що враховує актуальність, повноту, узгодженість і надійність джерела. Це дозволяє відокремити ризик самого ринку від ризику недостатньої інформаційної основи.

Надійність, безпека та супровід

Відкритий характер даних не усуває вимог до безпечної інтеграції. Ключі доступу до зовнішніх сервісів слід зберігати у захищеній конфігурації, а всі запити — журналювати без розкриття секретів. Необхідно реалізувати повторні спроби з експоненційною затримкою, локальне кешування, контроль лімітів API та черги завдань. У разі недоступності джерела аналітичний модуль має використовувати останню підтверджену версію даних із зазначенням дати її актуальності.

Для супроводу системи важливими є автоматизовані тести контрактів API, моніторинг змін схем і перевірка якості після кожного завантаження. Доцільно застосовувати версіонування адаптерів та внутрішньої схеми даних. Якщо постачальник змінює назву поля або структуру відповіді, помилка повинна виявлятися на інтеграційному рівні, не впливаючи безпосередньо на користувацьку частину системи.

Висновки

Інтеграція систем аналізу ринку з відкритими джерелами даних дає змогу автоматизувати отримання інформації, розширити доказову основу управлінських рішень та скоротити витрати на ручне збирання статистики. Ключовими складовими запропонованого підходу є адаптери джерел, уніфікована внутрішня модель, ETL/ELT-конвеєр, централізоване сховище, каталог показників і контроль якості даних. Подальший розвиток системи може передбачати використання машинного навчання для прогнозування попиту, автоматичне виявлення аномалій та формування рекомендацій щодо вибору ринку для конкретного виду продукції.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. The European Data Portal [Електронний ресурс]. – Режим доступу: <https://data.europa.eu/en> (дата звернення: 07.06.2026).
2. About the Indicators API Documentation / The World Bank [Електронний ресурс]. – Режим доступу: <https://datahelpdesk.worldbank.org/knowledgebase/articles/889392-about-the-indicators-api-documentation> (дата звернення: 07.06.2026).
3. API – Introduction / Eurostat User Guides [Електронний ресурс]. – Режим доступу: <https://ec.europa.eu/eurostat/web/user-guides/data-browser/api-data-access/api-introduction> (дата звернення: 07.06.2026).
4. OECD Data via API / Organisation for Economic Co-operation and Development [Електронний ресурс]. – Режим доступу: <https://www.oecd.org/en/data/insights/data-explainers/2024/09/api.html> (дата звернення: 07.06.2026).
5. Hohpe G., Woolf B. Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions. – Boston : Addison-Wesley, 2004. – 736 p.
6. Kimball R., Ross M. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. – 3rd ed. – Indianapolis : Wiley, 2013. – 600 p.

Бондар Владислав Олександрович – студент групи ІПІ-25м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: vladbondart11@gmail.com.

Науковий керівник: Ліщинська Людмила Броніславівна, Доктор технічних наук, професор, Вінницький національний технічний університет, Вінниця.

Bondar Vladyslav Oleksandrovich – student, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: vladbondart11@gmail.com.

Scientific supervisor: Lishchynska Lyudmila Bronislavivna, Doctor of Technical Sciences, Professor, Vinnytsia National Technical University, Vinnytsia.