

УДК 629.341

DOI: 10.30977/BUL.2219-5548.2022.96.0.40

## РЕАЛІЗАЦІЯ ПОШУКУ АНАЛОГІЧНИХ ТОВАРІВ НА ОСНОВІ FAISS

Знахур С. В., Знахур Л. В.

Харківський національний економічний університет ім. С. Кузнеця

**Анотація.** Проблема отримання повної множини аналогічних товарів різних постачальників з урахуванням інформації щодо зображення та опису товару є актуальною для e-commerce та практично цікавою. Робота містить алгоритм побудови загальної архітектури рішення засобами GCP та створення AI компонент на базі FAISS та Deep Learning для реалізації Product Quantization пошуку товарів на основі їхніх векторів текстових ознак та зображень. Запропонований підхід пошуку аналогічних товарів використовує мережу DNN для виділення features-зображень, текстових тегів та формування на їхній основі векторів, що застосовуються для FAISS-кластеризації.

**Ключові слова:** AI, ML, Deep Learning, FAISS, GCP, e-commerce, кластеризація.

### Вступ

Сучасні e-commerce, що працюють за схемами дропшипінгу чи маркетплейсу, на відміну від звичайних магазинів, не закуповують товар у постачальників, вони використовують зовнішню інформацію (відповідно до своєї системи) про їхні товари для організації бізнесу B2B або B2C. Організація інформаційних потоків e-commerce базується на достовірній та максимально повній інформації щодо товарів різних постачальників. Тому один із необхідних елементів автоматизації процесів e-commerce – це автоматичне створення опису товарів та їхнє зіставлення. Автоматичне зіставлення товарів допомагає виявити схожі товари й вказати ступінь схожості. Це особливо важливо для товарів, що складно піддаються порівнянню, наприклад одяг, аксесуари. Отже, у роботі визначимо базовими завданнями функціоналу маркетплейсу такі: пошук релевантного товару серед аналогічних товарів на маркетплейсі; отримання повної множини аналогічних товарів різних виробників і постачальників з урахуванням нечіткої або неповної інформації опису товару.

### Аналіз публікацій

Проведений аналіз показав напрями використання AI, які є комерційними перевагами сучасних e-commerce магазинів та маркетплейсів [1, 2]:

1. Тегування товарів на основі їхнього зображення та тегування товарів на основі NLP-аналізу їхнього текстового опису.

2. Пошук товарів на основі їхнього зображення або текстового опису.

3. Удосконалення рекомендаційної системи на основі масштабування, використання

гібридних підходів DSSM, Attentive Collaborative Filtering.

4. Підбір набору товарів та комбінації товарів для покупця.

5. Персоналізація пошуку та вибору товарів на основі поведінкової моделі користувача.

6. Використання онлайн-моделювання в процесі вибору або підбору товарів (зокрема 3D-моделювання).

Аналіз показав, що більшість AI рішень та сервісів є типовими для конкуруючих компаній у сфері e-commerce. Зазвичай це такі групи сервісів: тегування товарів на основі їхнього зображення та тегування товарів на основі NLP-аналізу їхнього текстового опису; пошук товарів на основі зображення; рекомендаційні системи та інструменти мапінгу (зіставлення зображень товарів або опису). Точність результатів рішень у конкурентів кардинально не відрізняється, оскільки вони використовують спеціалізовані DNN (Deep Neural Network), які мають точність від 90 до 97 % [1, 5, 6]. Особливістю є застосування власних наборів даних та розмітки для тегування товарів, постійна актуалізація та тюнінг DNN-мереж. Найбільш складною та унікальною розробкою є рекомендаційна система для користувачів щодо підбору товарів. Якість її роботи визначає точність підбору та ступінь персоналізації результатів для споживача. Рекомендаційну систему можливо побудувати на основі моделей DNN, що показали найбільшу точність для пошуку товарів на основі визначених атрибутів та ознак features (наприклад ResNet, EfficientNet) та на основі алгоритмів і мереж для рекомендації товарів на базі поведінкової моделі Deep Learning Recommender

System [2, 3]. Але необхідно зазначити, що базовим функціоналом кожної рекомендаційної системи є пошук схожих товарів.

### Мета та постановка завдання

Мета роботи – побудова алгоритму та архітектурного рішення пошуку аналогічних товарів маркетплейсу на основі використання алгоритму FAISS та Deep Learning. Робота передбачає огляд напрямів використання AI та передумов дослідження інтелектуального пошуку товарів для e-commerce, загальний алгоритм пошуку схожих товарів, реалізацію AI-компонент на базі FAISS і DeepLearning для Product Quantization пошуку товарів на основі використання їхніх текстових ознак та векторів зображень.

### Виклад основного матеріалу

Ключові передумови дослідження інтелектуального пошуку товарів для e-commerce такі:

1. Наявність тренду smart-покупки. Користувачі прагнуть купувати брендові або якісні речі за максимально низькими цінами, тому здійснюють пошук аналогічних товарів у різних постачальників.

2. Користувачі все більше здійснюють покупки, застосовуючи мобільні пристрої в режимі онлайн, що потребує швидкого та інтуїтивно зручного пошуку.

3. Наявність високої конкуренції серед маркетплейсів та електронних магазинів. Ефективний та інтелектуальний пошук товарів є значною перевагою в e-commerce.

4. Використання AI та Big Data у маркетплейсі, як обов'язкового компоненту.

5. Наявність значної кількості провайдерів AI та Big Data сервісів для маркетплейсів, агрегаторів, магазинів.

6. Доступність сучасних технологій, архітектур та технічних ресурсів для побудови або оренди AI-рішень. Існують відкриті рішення, навчені мережі для розпізнавання зображень товарів, аксесуарів на фото, тегування, кластеризації об'єктів (наприклад YOLO v3).

Розглянемо три напрями розвитку AI та Big Data рішень, які є актуальними для e-commerce:

1. Рішення, що оптимізують процеси (flow) покупки та надають додаткові цінності для покупця. Наприклад, це рішення, які побудовані на основі використання AI, Big Data, WEB-аналітики, карт поведінки, кліків, статистики поведінки та вибору користува-

чів, прогнозування на основі моделі поведінки, профілю користувачів тощо. Ці рішення дозволяють у режимі реального часу пропонувати (рекомендувати) товари певної групи (категорії), зменшити кількість пропозицій. Такі рішення становлять 80 % трудомісткості робіт із розроблення AI, Big Data та операційних витрат BI сучасних маркетплейсів.

2. AI та Big Data рішення, що оптимізують процеси (flow) продавця щодо зручності завантаження, контролю та моніторингу даних. Наприклад, завантаження даних і валідація щодо повноти опису товару, якості фото, наявності описових тегів. Для процесів (flow) продажів продавця – це аналіз даних продажів, пошук закономірностей продажів за товарами та їхніми групами, прогнозування та рекомендації щодо товарів.

3. AI та Big Data рішення, які забезпечують та оптимізують роботу адміністраторів маркетплейсу. Наприклад, оптимізація зберігання даних, оптимізація ETL, оптимізація запитів, пошук та виявлення інцидентів, передбачення інцидентів, оптимізація виконання запитів для агрегацій та звітів.

Найбільш цікавим є рішення AI, які оптимізують процеси (flow) покупки. Більшість AI та Big Data рішень e-commerce недоступні для детального аналізу на рівні архітектури, але можливо дослідити моделі, які покладені в їхню основу. Найбільш поширені моделі AI – це моделі класифікації та кластеризації товарів або запитів. Результати аналізу публікацій щодо методів та моделей класифікації або кластеризації товарів, що використовуються для пошуку та мапінгу, такі:

1. Точність алгоритмів та методів ґрунтується на використанні спеціалізованих, розмічених наборів даних (наприклад, MNIST [5])

2. Тегування зображень товарів (ідентифікація та класифікація) вже мають максимальну точність 92–96 % [4, 5], що дозволяє суттєво знизити помилки пошуку.

3. Пошук аналогічних об'єктів (товарів) ефективно реалізовувати в системах інтеграції на основі кластеризації (найбільша оперативність в умовах онлайн-пошуку).

4. DNN більш ефективні, ніж класичні методи ML для визначення ознак товарів та вирішення завдання класифікації [5].

Отже, для виконання завдань пошуку аналогічних товарів використаємо сучасний підхід на основі кластеризації, що передбачає вилучення n-вимірної вектора ознак товару (embeddings vector); зіставлення товарів за

векторами ознак (на основі кластеризації). Для реалізації цих кроків пропонується використати бібліотеку TensorFlow. Пошук товарів за векторами можливо реалізувати на основі бібліотеки FAISS [6, 8]. FAISS – це бібліотека для ефективного пошуку схожості і кластеризації векторів. Вона містить алгоритми пошуку в наборах векторів будь-якого розміру. Бібліотека має підтримку GPU, що значно пришвидшує процеси пошуку.

Загальний алгоритм пошуку схожих товарів наведено нижче та містить такі блоки:

1. Угрупування товарів за категоріями. Для практичного використання рішення рекомендується угрупувати товари тільки верхнього рівня (наприклад, групи верхнього одягу, взуття, аксесуари), оскільки категорії можуть надалі додаватися й глибина декомпозиції може збільшуватися. Ця процедура попереднього угруповання необхідна для зменшення часу побудови кластерів у пам'яті. Що менше товарів використовується в групі товарів, то швидше можна побудувати кластери та розрахувати відстані між ними.

2. На етапі ETL-фідів товарів формуються текстові теги на основі текстового опису товару. Для зручності кластеризації пропонується використовувати попередньо отримані теги (текстові атрибути) товарів, а не повний текст опису товарів. На основі набору тегів товару формуємо вектор певної довжини (наприклад 512-Flat). Для розрахунку вектора пропонується використати навчену модель нейронної мережі з архітектурою ResNet-18. Робота з нейронною мережею може бути здійснена на базі фреймворку TensorFlow.

3. На основі Product Quantization вектори розмірності (наприклад 512) розбиваються на  $n$ -частин.

4. Необхідно завантажити для всіх товарів вектори індексів (IVF-індекси). Для оптимізації необхідно завантажити вектор лише певної категорії товарів.

5. Використовуємо алгоритм кластеризації faiss.Clustering. Простір векторів розбивається методом  $k$ -means на  $k$  кластерів. Налаштовуємо кількість кластерів (наприклад, квадратний корінь із кількості товарів у категорії).

6. Створюємо Inverted Lists IVF простору векторів на диску. Навчаємо IVF-індекс із параметрами максимально можливого обсягу даних, який може бути завантажений у пам'ять сервера.

7. У навчений індекс частинами додаємо вектори, що були в навчанні, записуємо на диск IVF-індекс для кожної з частин (зберігаємо їх у БД).

8. Етап експлуатації для векторів на основі використання текстового опису такої:

У запиті (API) надсилаємо (id) товару та ідентифікуємо його вектор (у пам'яті простору IVF). Для пошуку схожих товарів використовуємо отриманий IVF-індекс і визначаємо найближчих сусідів (на відстані) для обраного товару. Ранжуємо знайдених найближчих сусідів (не більше ніж 10) за їхньою віддаленістю від нашого товару.

9. Аналогічно виконуємо процедуру кластеризації для векторів, які отримуємо на основі перетворення зображення на вектор. Етапи 3–7 аналогічні. Результатом є  $m$  найближчих сусідів для обраного вектора зображення товару.

10. Зіставляємо результати кластеризації (товари – найближчі сусіди). Вибираємо результати, які мають перетин товарів як для векторів за текстовими ознаками, так і для векторів за зображеннями (це необхідно для зменшення ймовірності помилки. За наявності великого набору товарів та постачальників можливо для кожного способу пошуку найближчих сусідів встановлювати вагу, наприклад, більшу вагу встановлюємо результатам за векторами зображення, меншу вагу – за текстовим описом).

11. Надаємо отриманий набір схожих товарів (id) як відповідь на запит пошуку (за допомогою API).

Фрагмент архітектури рішення щодо розв'язання завдання кластеризації наведено на рис. 1.

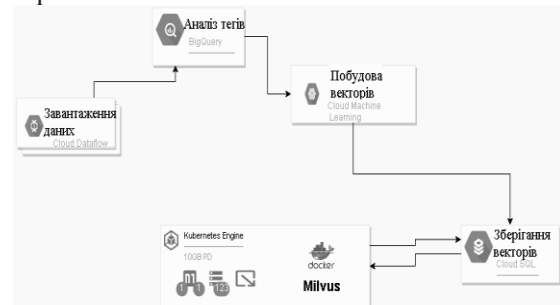


Рис. 1. Фрагмент архітектури рішення

Загальна архітектура базується на використанні сервісів GCP. Платформа Google Cloud (GCP) – це сукупність обчислювальних ресурсів Google. Для рішення пропонується такі сервіси:

1. Google Dataflow надає аналітичне рішення для отримання, оброблення й аналізу потоків подій в інфраструктурі GCP та є альтернативою пакетному ETL (Execute-Transform-Load). Потокова аналітика в GCP спрощує конвеєри ETL за допомогою API Java або Python. В архітектурі використовуємо Google Cloud Dataflow Template Pipelines для трансформації JSON даних товарів (фід) в дані BigQuery.

2. BigQuery – це безсерверне сховище даних, що використовується для зберігання та аналізу наборів даних фідів товарів (після препроцесингу – тегів). BigQuery дозволяє: будувати власні схеми, які організують дані в набори даних та таблиці; завантажувати дані з різних джерел, зокрема потокові дані; використовувати запити SQL; використовувати WEB-інтерфейс або API; здійснювати завантаження, експорт та копіювання даних за допомогою завдань.

3. Платформа GCP AI використовується для машинного навчання (ML) та побудови векторів зображень та тегів. Існує можливість використовувати API, які надають змогу застосовувати заздалегідь підготовлені моделі DNN. Також можна створити й навчити власні масштабні, складні моделі, використовуючи TensorFlow. Реалізацію завдань embedding та побудови векторів пропонується реалізувати в GCP AI на основі ResNet-18 у фреймворку TensorFlow.

4. Cloud SQL – служба реляційних баз даних для MySQL, PostgreSQL та SQL Server. Для інтеграції з рішенням MILVUS пропонується використовувати PostgreSQL для зберігання даних щодо векторів.

5. Kubernetes Engine (GKE) – це середовище для розгортання контейнерних застосунків. GKE забезпечує швидке розроблення та ітерацію застосунків, полегшуючи розгортання, оновлення та управління програмами та послугами. Підтримка апаратних прискорювачів дозволяє запустити машинне навчання, GPU загального призначення, високо-ефективні обчислення та інші робочі навантаження [9]. Реалізацію завдання кластеризації (FAISS) можливо здійснити на основі вже розробленого докеру проєкту MILVUS. Масштабування рішення здійснюється за рахунок налаштування GKE.

5. Хмарні функції GCP (FaaS) забезпечують середовище виконання без сервера для побудови та підключення інших хмарних сервісів. Хмарні функції в роботі використовуються для: попереднього оброблення да-

них фідів товарів та завантаження зображень у Cloud Storage; відповіді на тригери HTTP; роботи з API; реалізації WEB-бекенда.

### Висновки

Сучасні рішення AI можуть бути максимально практичні для e-commerce, вони допомагають утримувати або повертати клієнтів, рекомендувати товари або визначати цільову аудиторію тощо. Унаслідок дослідження було запропоновано алгоритм щодо реалізації рішення на базі GCP, розроблено, протестовано ML-компоненти задачі пошуку аналогічних товарів на основі використання алгоритмів FAISS та Deep Learning в GCP. Але залишилось багато практичних питань щодо оптимізації рішення, наприклад, вибір моделей векторизації даних щодо товару або препроцесингу фідів, які суттєво впливають на швидкість, точність та оперативність рішення.

### Література

1. 2021: A Year Full of Amazing AI papers. URL: [https://github.com/louisfb01/best\\_AI\\_papers\\_2021](https://github.com/louisfb01/best_AI_papers_2021)
2. Marketplace for AI Models. URL: <https://arxiv.org/pdf/2003.01593.pdf>.
3. A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems. URL: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/frp1159-songA.pdf>.
4. GAN-based-Rec recommender-System. URL: <https://github.com/jihoo-kim/GAN-based-Rec-recommender-System>.
5. Image classification on fashion-MNIST. URL: <https://github.com/cmasch/zalando-fashion-mnist>
6. Benchmark Fashion MNIST. URL: <http://fashion-mnist.s3-website.eu-central-1.amazonaws.com/#>
7. FAISS. URL: <https://github.com/facebookresearch/faiss>
8. MILVUS. URL: <https://github.com/milvus-io/pymilvus/tree/0.2.14>.
9. Product quantization for nearest neighbor search. URL: [http://lear.inrialpes.fr/pubs/2011/JDS11/jegou\\_searching\\_with\\_quantization.pdf](http://lear.inrialpes.fr/pubs/2011/JDS11/jegou_searching_with_quantization.pdf).

### References

1. 2021: A Year Full of Amazing AI papers. URL: [https://github.com/louisfb01/best\\_AI\\_papers\\_2021](https://github.com/louisfb01/best_AI_papers_2021)
2. Marketplace for AI Models. URL: <https://arxiv.org/pdf/2003.01593.pdf>.
3. A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems. URL: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/frp1159-songA.pdf>.

- songA.pdf.
4. GAN-based-Recommender-System. URL: <https://github.com/jihoo-kim/GAN-based-Recommender-System>.
  5. Image classification on fashion-MNIST. URL: <https://github.com/cmasch/zalando-fashion-mnist>.
  6. Benchmark Fashion MNIST. URL: <http://fashion-mnist.s3-website.eu-central-1.amazonaws.com/#>
  7. FAISS. URL: <https://github.com/facebookresearch/faiss>.
  8. MILVUS. URL: <https://github.com/milvus-io/pymilvus/tree/0.2.14>.
  9. Product quantization for nearest neighbor search. URL: [http://lear.inrialpes.fr/pubs/2011/JDS11/jegou\\_searching\\_with\\_quantization.pdf](http://lear.inrialpes.fr/pubs/2011/JDS11/jegou_searching_with_quantization.pdf).

**Знахур Сергій Вікторович**, к.е.н., доц. каф. інформаційних систем, Харківський національний економічний університет ім. С. Кузнеця, тел. +38 050-30-04-390, [serhii.znakhur@gmail.com](mailto:serhii.znakhur@gmail.com),  
**Знахур Людмила Володимирівна**, ст. викл. каф. інформаційних систем, Харківський національний економічний університет ім. С. Кузнеця, тел. +38 050-65-11-282, [razinalv@gmail.com](mailto:razinalv@gmail.com).

#### Similar goods search based on FAISS

**Abstract.** *The problem of obtaining a complete set of similar goods from different manufacturers based on the image and description of the product is relevant and interesting. The article includes description of algo and architecture solution for online goods clustering. Goal.* *The goal of the work is to create*

*and explore models of e-commerce ML (online goods clustering). The work includes developing AI components based on FAISS and Deep Learning to implement Product Quantization of goods searches based on their embedding vectors and vectors of images. The proposed approach to the search for similar goods uses the DNN for features detection and embedding vectors, which are used for FAISS clustering. Methodology.* *The analytical and empirical methods of research based on the development and DNN are used, ML methods to determine features of goods and solve classification problems. Results.* *The architecture of the solution is based on the use of GCP services. Practical value.* *AI solutions have practical value for e-commerce, they help retain or return customers, and recommend products. As a result of the work, a basic solution to the problem of matching similar products based on the use of FAISS and Deep Learning algorithms in GCP was developed and tested.*

**Key words:** *GCP, ML, Deep Learning, FAISS, images, tags, clustering.*

**Znakhur Serhii**, Assoc. Prof. Information Systems Department, Simon Kuznets Kharkiv National University of Economics, tel. +38 050-300-43-90, [serhii.znakhur@gmail.com](mailto:serhii.znakhur@gmail.com),

**Znakhur Liudmyla**, Lecturer of Information Systems Department, Simon Kuznets Kharkiv National University of Economics, tel. +38 050-65-11-282, [razinalv@gmail.com](mailto:razinalv@gmail.com).