

УДК 004.7

*Васянович О.А., здобувач,
Морозов А.В., к.т.н., доцент
Державний університет «Житомирська політехніка»*

АРХІТЕКТУРА РЕКОМЕНДАЦІЙНИХ СИСТЕМ В УМОВАХ ВІДСУТНОСТІ СТАТИСТИКИ ВЗАЄМОДІЙ

Базовою проблемою інформаційних платформ, орієнтованих на розміщення нового користувачького контенту, є неможливість його початкового ранжування. Традиційні алгоритми рекомендацій, побудовані на методах колаборативної фільтрації, вимагають накопиченої історії взаємодій користувачів з об'єктом (кількість переглядів, прослуховувань, додавань до обраного). Щойно завантажений медіафайл не має такої статистики, тому алгоритм не здатний рекомендувати його цільовій аудиторії. Цей стан системи визначається як проблема «холодного старту» [1].

Для забезпечення просування нових публікацій під час розробки інформаційної системи організації прослуховування аудіоконтенту (проект «IndieStream») реалізовано контентно-орієнтовану архітектуру рекомендацій (content-based filtering) [2]. Алгоритм базується на аналізі внутрішніх метаданих самого об'єкта, що дозволяє формувати видачу незалежно від поведінки користувачів.

Під час завантаження аудіотреку в базу даних система фіксує його текстові характеристики: жанр, настрій, темп та інструментальні теги. Для обчислення релевантності між поточним треком та іншими записами ці характеристики перетворюються на багатовимірні числові вектори за допомогою алгоритму TF-IDF (Term Frequency-Inverse Document Frequency) [3].

Пошук подібних композицій здійснюється через розрахунок косинусної подібності між вектором цільового треку A та векторами інших записів B . Математична модель розрахунку має вигляд:

$$\text{sim}(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Значення функції наближається до одиниці за умови високого збігу векторних характеристик композицій.

На програмному рівні цю архітектуру імплементовано мовою Python з використанням математичної бібліотеки scikit-learn. При запиті до сторінки конкретного треку бекенд динамічно обчислює матрицю подібності для тегів і формує JSON-відповідь із вибіркою композицій, що мають найвищий коефіцієнт збігу.

```
recommendation.py ×
1 from sklearn.feature_extraction.text import TfidfVectorizer
2 from sklearn.metrics.pairwise import cosine_similarity
3
4 # Метадані треків у базі (жанр, настрій, темп)
5 tracks = [
6     {"id": 1, "name": "Ocean Drive", "tags": "indie rock instrumental upbeat"},
7     {"id": 2, "name": "Silent City", "tags": "dark post-punk heavy bass"},
8     {"id": 3, "name": "Neon Runner", "tags": "indie rock electronic driving"},
9     {"id": 4, "name": "Late Night", "tags": "lo-fi chill study"},
10 ]
11
12 # Формування векторів TF-IDF
13 vectorizer = TfidfVectorizer()
14 tfidf_matrix = vectorizer.fit_transform([t["tags"] for t in tracks])
15
16 # Розрахунок косинусної подібності для треку Ocean Drive
17 target_index = 0
18 cosine_sim = cosine_similarity(tfidf_matrix[target_index], tfidf_matrix).flatten()
19
20 # Формування результату (сортування за релевантністю)
21 similar_indices = cosine_sim.argsort()[::-1]
```

Рис. 1. Реалізація косинусної подібності на стороні бекенду.

Впровадження контентно-орієнтованого алгоритму з розрахунком векторної подібності вирішує проблему «холодного старту». Архітектура забезпечує авторам отримання органічного трафіку відразу після публікації матеріалу в інформаційній системі, спираючись виключно на початкові характеристики контенту.

Список використаних джерел:

1. Isinkaye F. O., Folocunso Y. O., Ojokoh B. A. Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal. 2015. Vol. 16, P. 261-273. URL: <https://doi.org/10.1016/j.eij.2015.06.005> (дата звернення: 23.03.2026).
2. Pengshi Alvin. Content-Based Recommender for E-Commerce Web Store. Towards Data Science. 2020. URL: <https://towardsdatascience.com/a-content-based-recommender-for-e-commerce-web-store-7554b5b73eac> (дата звернення: 23.03.2026).
3. Офіційна документація бібліотеки scikit-learn. URL: <https://scikit-learn.org/> (дата звернення: 23.03.2026).