

**Паулін О.М.**

Одеський національний політехнічний університет

**Комлева Г.О.**

Одеський національний політехнічний університет

**Улізко Г.В.**

Одеський національний політехнічний університет

## РЕКОМЕНДАЦІЙНА СИСТЕМА ДЛЯ ДОПОМОГИ У ВИВЧЕННІ МУЗИЧНИХ ТВОРІВ

*Роботи у сфері створення рекомендаційних систем можуть мати як комерційний, так і науково-дослідний характер і потребують вирішення ряду питань. Перше питання полягає в усвідомленні місця рекомендаційної системи у бізнес-процесі та формалізації функціоналу, виконання якого вона повинна забезпечувати. Це вимагає чіткого розуміння принципів роботи кожного етапу бізнес-завдання. Друге питання пов'язане з визначенням принципів створення рекомендацій. Традиційно для побудови рекомендаційних систем використовуються контент-орієнтовані методи, колаборативна фільтрація та кластеризація. Особливий інтерес становить використання рекомендаційних систем як помічника при самостійному опануванні нового пізнавального напрямку чи поглиблення власних навичок у будь-якій сфері навчання. Метою роботи є підвищення якості процесу вивчення музичних творів шляхом використання рекомендаційної системи, що дозволяє створювати власну траєкторію навчання з урахуванням поточного рівня гри та музичних уподобань без залучення музичного ментора. У роботі визначено методи надання рекомендацій і метрики для оцінювання релевантності цих рекомендацій вподобанням користувача. Згідно з типовим використанням програми користувачеві для власного опанування надаються нотні тексти, які він переглядає та зберігає у поточний плейлист, створюючи тим самим фрагмент траєкторії навчання на найближчий проміжок часу. За допомогою зворотного зв'язку перевіряється ступінь задоволення користувача наданим списком творів. Система використовує оцінки TF-IDF та метод косинусної подібності для обробки векторів, що визначають «профілі» для користувача та музичних творів. З архітектурного погляду розроблена система є вебсервісом, який містить серверну частину проекту на основі Apache, рекомендаційну модель на основі мови Python, бібліотек Sklearn і Pandas, сервер бази даних MySQL і базу для збереження даних.*

**Ключові слова:** рекомендаційна система, контент-орієнтований метод, метод косинусної подібності, машинне навчання, модель, метрика.

**Постановка проблеми.** Центральною проблемою навчання з метою оволодіння навичками в новій для себе предметній сфері є управління процесом навчання за допомогою ефективних методів і засобів. Зазвичай це управління виконується вчителем, або ментором, який має більший професійний досвід, ніж його підопічний. Ментор бере функції регулювання, мотивації, виявлення загальних слабких сторін і розвитку стратегії подальшого навчання. Відсутність ментора викликає комплекс труднощів. Складність організації навчально-пізнавального процесу без участі ментора впливає з недостатньої формалізації загальної мети, контролю та складових частин цього процесу. Потужним інструментом для підтримки пізнавального процесу без залучення ментора можуть стати рекомендаційні системи.

В останні роки рекомендаційні системи отримали поширення при вирішенні завдань із різних предметних сфер. Рекомендаційна система – це програмний інструмент, призначений для аналізу минулого досвіду користувача і надання списку пропозицій із великого обсягу інформації. Принципи роботи рекомендаційних систем залежать від особливостей конкретного бізнес-завдання та часто ґрунтуються на методах машинного навчання. Система рекомендацій – це метод, який використовується для надання користувачеві оптимальних пропозицій із вибору елемента. Ці пропозиції засновані на різних процесах прийняття рішень, наприклад, виборі предметів для покупки, перегляді фільму з набору фільмів та ін. Особливий інтерес становлять рекомендаційні системи, націлені на допомогу в навчанні залежно

від вхідного рівня користувача, особливостей навчального контенту та кінцевої мети навчання.

Різні типи алгоритмів рекомендацій працюють відповідно до їх сфер і знань, що використовуються користувачами; в кінці для створення рекомендацій використовуються різні типи алгоритмів прогнозування.

Для кожної рекомендаційної системи повинен бути визначений набір метрик для оцінки їх роботи. Відсутність чітких критеріїв робить неможливою оцінку успішності обраної моделі рекомендаційної системи та методів, які вона використовує. За наявності декількох варіантів правильний вибір метрик для поставленого завдання дозволяє отримати оптимальне рішення бізнес-завдання.

Найчастіше для побудови рекомендаційних систем використовуються контент-орієнтовані методи, колаборативна фільтрація та кластеризація. Серед метрик оцінки використовують статистичні метрики та метрики підтримки прийнятих рішень.

При розробці рекомендаційної системи потрібно визначитися з підходом, який буде покладено в її основу, таким як:

- *контент-орієнтована* фільтрація, заснована на зіставленні користувачів із тим контентом або товарами, котрим було віддано перевагу, з використанням методів косинусної подібності та міри Жаккара;

- *колаборативна* фільтрація, що використовує відомі уподобання групи користувачів щодо певного контенту або товару для прогнозування невідомих уподобань іншого користувача з використанням матриці корисності;

- *кластеризація*, яка є завданням машинного навчання без вчителя та дозволяє за відсутності чи малої кількості розмічених даних кластеризувати спостереження на основі набору ознак, а потім призначити рекомендації для кластерів на основі міток, якими володіють об'єкти в певному кластері.

Кожен із цих підходів дозволяє отримати доступні, індивідуальні, а головне – високоякісні рекомендації для певних завдань прийняття рішень.

### **Аналіз останніх досліджень і публікацій.**

Нині доступна велика кількість рекомендаційних систем, спроектованих відповідно до певного підходу. Формування системи рекомендацій починається з інформаційного аналізу елементів і користувачів, за яким виконується створення моделі користувача. У цієї моделі зберігається інформація, оброблена шляхом аналізу інформації, після чого модель використовується для генерації рекомендацій [1, с. 113].

Рекомендуючи елементи користувачам, важливо враховувати багато показників продуктивності, а не тільки точність прогнозів рекомендацій. Існують різні метрики для оцінки рекомендацій [2, с. 187].

Метрика різноманітності – це включення різних типів елементів, поданих у рекомендації для користувача, що відрізняються від попередніх уподобань. Різноманітність обчислюється за допомогою міри подібності стосовно списку елементів. Різноманітність у системах рекомендацій запроваджується для вирішення проблеми надмірності, яка протягом останніх кількох років стає темою, котру багато дослідників обговорюють в інформаційних публікаціях [3, с. 19; 4, с. 2445]. Крім того, застосування метрики різноманітності дозволяє задовольнити користувачів рекомендаціями із використанням диверсифікації [5, с. 176; 6, с. 154].

Міра випадковості – це міра того, як для користувача формуються несподівані або відповідні рекомендації. Випадковість обчислюється як різниця ймовірності елемента  $i$ , рекомендованої для користувача  $u$ , і ймовірності того, що елемент  $i$  буде рекомендований для будь-якого іншого користувача [7, с. 273; 8, с. 84].

Метрика новизни – це показник якості системи рекомендацій, завдяки якому оцінюється ефективність додавання нового елемента до списку рекомендацій. Недостатнє значення цього показника веде до низького ступеня диверсифікації [9, с. 1].

Велика кількість розроблених рекомендаційних систем є контент-орієнтованими [10, с. 176]. Тематична фільтрація формує рекомендацію на основі поведінки користувача, тобто ретроспективної інформації про те, які елементи обирав користувач, і характеристики цих елементів. Цей контент може бути визначений у ручному режимі або витягнутий автоматично на основі інших методів подібності [11, с. 3; 12, с. 1; 13, с. 524].

Таким чином, як показав аналіз, існує багато продуманих методів і метрик якості для реалізації й оцінювання контент-орієнтованих рекомендаційних систем для пошуку інформаційних елементів. Тому доцільне використання цього підходу для розробки методів і програмних засобів, націлених на підтримку самостійної навчально-пізнавальної діяльності.

**Постановка завдання.** Метою роботи є підвищення якості процесу вивчення музичних творів шляхом використання рекомендаційної системи, що дозволяє створювати власну траєкторію навчання з урахуванням поточного рівня гри

та музичних уподобань без залучення музичного ментора.

**Виклад основного матеріалу дослідження.** Розроблювана рекомендаційна програмна система заснована на контент-орієнтованій фільтрації. Кожен користувач зіставляється з тим контентом, тобто музичними творами, які йому подобалися та були обрані для вивчення. Мета контент-орієнтованих методів – створити «профіль» для кожного користувача і кожного твору.

Найважливішу роль віграють атрибути творів (метадані), що використовуються для побудови списку рекомендацій. Серед таких атрибутів можна виділити такі: назву, жанр, композитора, складність, тривалість, національність стилю, ритм, часовий діапазон та ін. Крім того, можливо витягувати такі характеристики, як оцінка настроїв і оцінки TF-IDF з описів творів. TF-IDF (TF – term frequency, IDF – inverse document frequency) – це статистичний показник, що використовується для оцінки важливості слів у контексті документа, який є частиною колекції документів, у нашому випадку – описів музичних творів. Вага (значимість) слова пропорційна кількості вживань цього слова у документі й обернено пропорційна частоті вживання

слова в інших документах колекції. Оцінка TF-IDF відображає, наскільки важливе слово для опису твору в наборі описів. Попередньо машинний класифікатор навчається на колекції заздалегідь розмічених описів, а потім отримана модель використовується для аналізу нових описів. Для рекомендаційної системи це означає необхідність отримання та розмітки певної кількості описів для створення моделі аналізу, яка потім буде використовуватися для нових описів і визначати, чи рекомендувати той чи інший твір. Дані, що використовуються для навчання моделі, повинні бути повними та коректними [14, с. 91]. Якщо немає потрібної кількості таких даних, вирішується задача кластеризації [15, с. 376].

На рис. 1 показана узагальнена схема роботи системи. Вона передбачає: 1) типове використання програми, за якого зареєстрований користувач може отримувати нотні тексти згідно з його уподобаннями від рекомендаційної системи; 2) оцінювання роботи програми з наданням зворотної реакції щодо релевантності наданого рекомендаційного списку уподобанням користувача; 3) оформлення підписки на програмні оновлення для подальшого просування на ринку та монетизації програмного продукту; 4) відстеження про-



Рис. 1. Узагальнена схема роботи системи

дуктивності для забезпечення роботи системи з великою кількістю користувачів.

За типового використання програмної системи користувач обирає вікову категорію – дорослу чи дитячу – та переглядає список музичних творів. Щоб допомогти користувачеві орієнтуватися серед великої кількості творів, були розроблені фільтри, що дозволяють шукати твори за певними ознаками, наприклад, за жанром, автором (композитором), розділом тощо. Алгоритм фільтрації починається з формування запити до бази даних на вибірку всіх наявних у ній творів. Далі відбувається читання даних і відображення їх відповідно до обраних фільтрів. За бажанням користувача йому надається рекомендація щодо твору (рис. 2).

На рис. 3 показана діаграма роботи рекомендаційної системи. Спочатку користувач робить запит на рекомендацію відповідно до заданих ним вимог. Контролер перевіряє метадані й описи творів, наявні у базі даних, обчислює оцінки TF-IDF та створює відповідні вектори для користувача та творів. Далі система з використанням методу косинусної подібності обчислює подібності між користувачами та творами. Твори повинні бути рекомендовані користувачеві, якщо: 1) вони мають найбільшу схожість із користувачем або

2) мають велику схожість з іншими елементами, прочитаними користувачем.

Архітектура системи наведена на рис. 4. Вона містить: 1) сторінки сайту та їхні складові частини – компоненти; 2) серверну частину проекту на основі Apache, що опрацьовує запити клієнта; 3) рекомендаційну модель, яка кастомізується завдяки засобам машинного навчання у Python; 4) сервер бази даних MySQL і саму базу.

**Висновки.** У роботі проведено проектування рекомендаційної системи, що дозволяє обирати власну траєкторію навчання за самостійного опанування музичних творів. Система використовує принцип контент-орієнтованої фільтрації для побудови рекомендацій. Для оцінки точності методу фільтрації шляхом порівняння прогнозованих рейтингів безпосередньо з фактичним рейтингом користувачів використані такі статистичні метрики, як середня абсолютна помилка та середньоквадратична помилка. Застосування системи дозволить підвищити якість процесу навчання з наданням персональних онлайн-рекомендацій. Зворотній зв'язок дозволяє робити додаткове налаштування з метою відстеження відповідності встановленим метрикам для оцінки рекомендацій: різноманітності, випадковості, новизни.

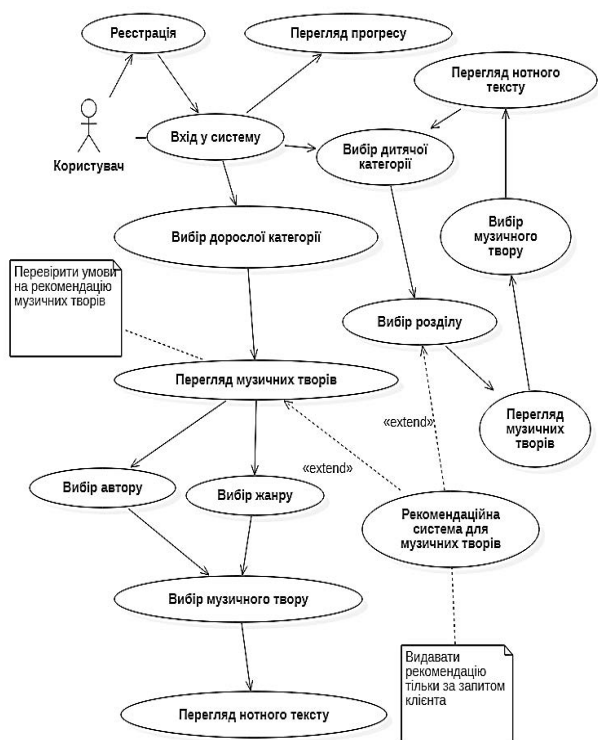


Рис. 2. Деталізація функціоналу рекомендаційної системи

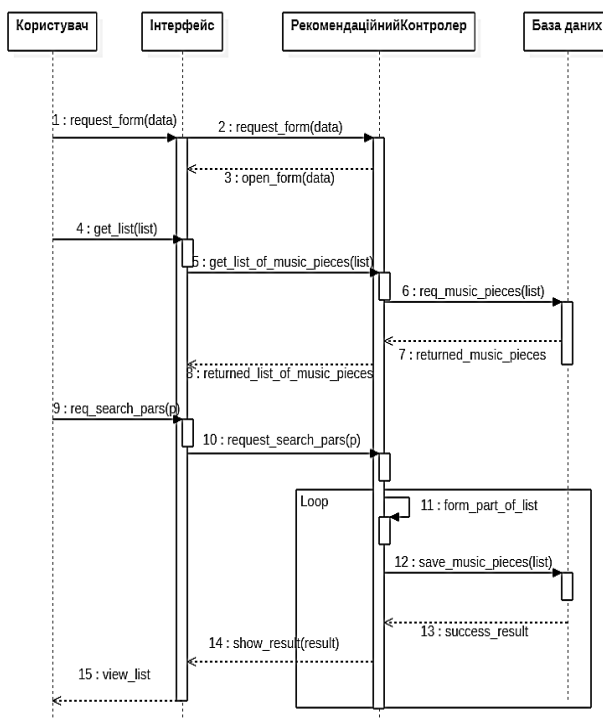


Рис. 3. Діаграма послідовності роботи рекомендаційної системи



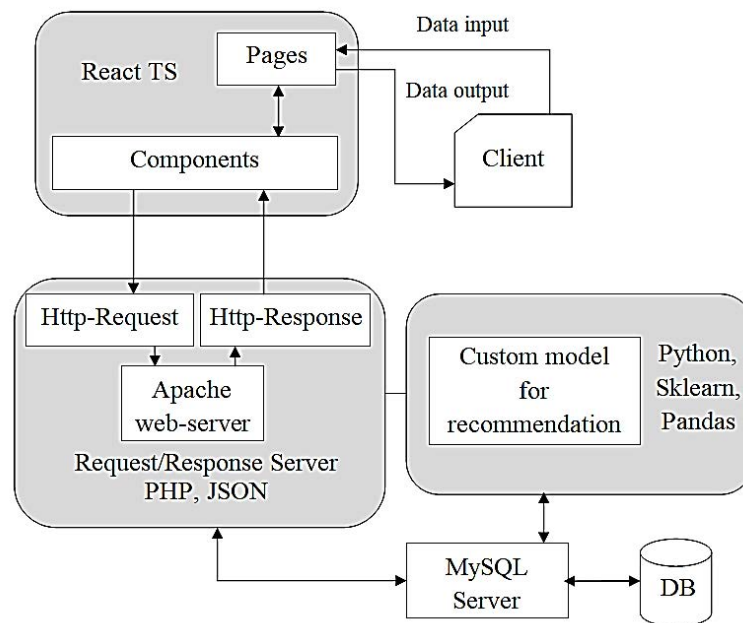


Рис. 4. Архітектура програмної системи

#### Список літератури:

1. Yadav N., Kumar R., Singh A., Pal S. Diversity in Recommendation System: Cluster Based Approach. *Hybrid Intelligent Systems*. 2020. P. 113–122.
2. Bag S., Abhijeet G., Manoj K.T. An integrated recommender system for improved accuracy and aggregate diversity. *Computers Industrial Engineering*. 2019. P. 187–197.
3. Yuan B. One-class Field-aware Factorization Machines for Recommender Systems with Implicit Feedbacks. *Technical Report. National Taiwan University*. 2019. 216 p.
4. Cho-Jui H., Nagarajan N., Inderjit D. PU Learning for Matrix Completion. *In Proceedings of the 32nd International Conference on Machine Learning (ICML)*. 2015. P. 2445–2453.
5. Gantner Z., Drumond L., Freudenthaler C., Rendle S., Schmidt-Thieme L. Learning attribute-to-feature mappings for cold-start recommendations. *In Proceedings of the IEEE International Conference on Data Mining (ICDM)*. 2010. P. 176–185.
6. Kunaver M., Požrl T. Diversity in recommender systems – A survey. *Knowledge-Based Systems*. 2017. P. 154–162.
7. Twardowski B. Modelling contextual information in session-aware recommender systems with neural networks. *In Proceedings of the 10th ACM Conference on Recommender Systems*, ACM. 2016. P. 273–276.
8. Adomavicius G., Bockstedt J.C., Curley S.P., Zhang, J. Effects of online recommendations on consumers' willingness to pay. *Information Systems Research*. 2017. V. 29 (1). P. 84–102.
9. Javari A., Jalili M. A probabilistic model to resolve diversity – accuracy challenge of recommendation systems. *Knowledge and Information Systems*. 2014. P. 1–19.
10. Глибовец Н.Н., Сидоренко М.О. Создание рекомендационной системы учебного типа с использованием фреймворка. *Проблемы интеллектуализации компьютера : сборник статей. Институт кибернетики им. В.М. Глушкова НАН Украины*, 2012. С. 176–181.
11. Nath S., Chakrabartty S.N. Cosine similarity approaches to reliability of Likert scale and items. *Romanian journal of psychological studies*. 2018. P. 3–16.
12. Fletcher S., Zahidul I. Comparing sets of patterns with the Jaccard index. *Australasian Journal of Information Systems*. 2018. V. 22. P. 1–17.
13. Sama R., Thamarai L. A Survey on Recommender System in Learning Analytics. *Education Data mining*. 2020. P. 524–541.
14. Крисилов В.А., Комлевая Н.О. Анализ и оценка компетентности источников информации в задачах интеллектуальной обработки данных. *Международная научно-практическая конференция «Электротехнические и компьютерные системы: теория и практика» ELTECS-2019. Problemele energeticii regionale*. Вып. 1–1 (40). 2019. С. 91–104.
15. Komleva N., Liubchenko V., Zinovatnaya S. Methodology of information monitoring and diagnostics of objects represented by quantitative estimates based on cluster analysis. *Applied Aspects of Information Technology*. 2020. V. 3. P. 376–392.

**Paulin O.M., Komleva H.O., Ulizko H.V. RECOMMENDATION SYSTEM FOR ASSISTANCE IN THE STUDY OF MUSICAL WORKS**

*Work in the field of creating recommendation systems can be both commercial and research in nature, and need to address a number of issues. The first question is to understand the place of the recommendation system in the business process and formalize the functionality that it must provide. This requires a clear understanding of the principles of each stage of the business task. The second issue is to define the principles for making recommendations. Traditionally, content-oriented methods, collaborative filtering and clustering are used to build recommendation systems. Of particular interest is the use of recommendation systems as an assistant in the independent acquisition of a new cognitive direction or the deepening of their own skills in any field of study. The aim of the work is to improve the quality of the process of studying musical works by using a recommendation system that allows you to create your own learning trajectory taking into account the current level of play and musical preferences without the involvement of a music mentor. The paper identifies methods for providing recommendations and metrics for assessing the relevance of these recommendations to the user's preferences. According to the typical use of the program, the user is provided with musical texts for his own mastery, which he views and saves in the current playlist, thus creating a fragment of the learning trajectory for the near future. Feedback is used to check the user's satisfaction with the provided list of works. The system uses TF-IDF estimates and the cosine similarity method to process vectors that define "profiles" for the user and music. From an architectural point of view, the developed system is web service that contains the server part of the project based on Apache, recommendation model based on Python language, Sklearn and Pandas libraries, MySQL database server and database for storing data.*

**Key words:** recommendation system, content-oriented method, cosine similarity method, machine learning, model, metrics.