

РЕКОМЕНДАЦІЙНА ІНФОРМАЦІЙНА СИСТЕМА ДЛЯ ПОШУКУ ВІДЕОМАТЕРІАЛІВ**Ю. В. Парфененко, А. А. Ковтун, А. А. Вербицька**

Сумський державний університет

вул. Римського-Корсакова, 2, м. Суми, 40007, Україна. E-mail: yuliya_p@cs.sumdu.edu.ua

Зростання обсягів відеоматеріалів у мережі Інтернет та потреба користувачів у швидкому одержанні необхідної інформації зумовлює актуальність проблеми таргетованого пошуку відеоматеріалів. Стаття присвячена вирішенню проблеми автоматизації фільтрації відеоматеріалів з сервісу Youtube. Розроблена рекомендаційна інформаційна система у вигляді веб-додатку, основним завданням якої є підбір актуального відеоконтенту з урахуванням потреб користувача. Для надання рекомендацій у розробленій системі використано метод колаборативної фільтрації, що базується на колективній співпраці користувачів системи, які переглядають відеоматеріали та оцінюють їх. У результаті досліджень було встановлено, що використання розробленої рекомендаційної системи забезпечує точну фільтрацію матеріалів з YouTube. Дана робота є важливою при використанні в освітній діяльності, тому що полегшує пошук актуальних навчальних матеріалів за певною тематикою.

Ключові слова: інформаційна система, колаборативна фільтрація, пошук, відеоматеріали, YouTube.

РЕКОМЕНДАТЕЛЬНАЯ ИНФОРМАЦИОННАЯ СИСТЕМА ДЛЯ ПОИСКА ВИДЕОМАТЕРИАЛОВ**Ю. В. Парфененко, А. А. Ковтун, А. А. Вербицкая**

Сумской государственной университет

ул. Римского-Корсакова, 2, г. Сумы, 40007, Украина. E-mail: yuliya_p@cs.sumdu.edu.ua

Рост объемов видеоматериалов в сети Интернет и потребность пользователей в быстром получении необходимой информации обуславливает актуальность проблемы таргетированого поиска видеоматериалов. Статья посвящена решению проблемы автоматизации фильтрации видеоматериалов с сервиса Youtube. Разработана рекомендательная информационная система в виде веб-приложения, основной задачей которой является подбор актуального видеоконтента с учетом потребностей пользователя. Для предоставления рекомендаций в разработанной системе использован метод коллаборативной фильтрации, основанный на коллективном сотрудничестве пользователей системы, которые просматривают видеоматериалы и оценивают их. В результате исследований было установлено, что использование разработанной рекомендательной системы обеспечивает точную фильтрацию материалов с YouTube. Данная работа является важной при использовании в образовательной деятельности, так как облегчает поиск актуальных учебных материалов по определенной тематике.

Ключевые слова: информационная система, коллаборативная фильтрация, поиск, видеоматериалы, YouTube.

АКТУАЛЬНІСТЬ РОБОТИ. На сьогодні, завдяки відкритості та розповсюдженості мережі Інтернет значні обсяги відеоматеріалів стали доступними широкому колу користувачів. Не виключенням є й галузь освіти, яка широко використовує відеоматеріали у навчальному процесі для аудиторної та самостійної роботи. Кількість навчальних відеоматеріалів в мережі Інтернет постійно зростає.

Не зважаючи на те, що пошук необхідного контенту користувачами здійснюється за допомогою пошукових систем, значні обсяги таких даних користувач не в змозі швидко переглянути та оцінити, настільки ці дані є для нього корисними. Тому з'являється потреба в якісній фільтрації контенту.

На даний час найбільшим сервісом, що зберігає відеоматеріали за різними тематиками, є Youtube. Для того, щоб з'ясувати, є корисним відеоматеріал, чи ні, його потрібно попередньо переглянути. Користувач не в змозі переглянути усі відеоматеріали за певною тематикою, тому знайти найбільш підходящий відеоматеріал наприклад, для вивчення конкретної мови програмування, є складною задачею.

Таким чином, задача автоматизації пошуку та фільтрації контенту, розміщеного на YouTube, для того, щоб підібрати контент, який є корисним конкретному користувачеві, є актуальною.

Метою даної роботи є розробка рекомендаційної інформаційної системи для пошуку необхідних відеоматеріалів на YouTube.

МАТЕРІАЛ І РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ. Рекомендаційною системою називають систему, яка використовує певний тип фільтрації і існуючі відомості про потреби користувачів, для подальшої рекомендації набору альтернатив, які система вважає актуальними для них [1, 2]. Рекомендаційні системи застосовуються у сферах розваг, електронної комерції, послуг, а також для надання персоналізованої рекомендації у системах, що зберігають різний контент – статті журналів, документи, відеоматеріали тощо.

Незалежно від призначення рекомендаційної системи маємо справу з користувачем, якому надається множина альтернатив, серед яких йому потрібно обрати найбільш прийнятний варіант. Користувачу може не вистачати власного досвіду і знань для того, обрати саме ті варіанти, які відповідають його потребам. Особливо це стосується навчальних матеріалів, так як зазвичай користувач не володіє достатньою кількістю інформації з тієї предметної області, яку збираються вивчати, щоб обрати змістовний та корисний матеріал.

Основними типами фільтрації, що застосовують при розробці рекомендаційних систем, є такі типи фільтрації: заснована на контенті; колаборативна; гібридна [3].

Контентна фільтрація надає рекомендації на основі поведінки користувачів в системі, де зберігається контент [4]. Наприклад, цей підхід може використовувати дані про перегляди статей та блогів (які теми читає користувач та характеристики таких блогів), а також відеоматеріалів.

Колаборативна фільтрація формує рекомендації, засновані на моделі поведінки попередніх користувачів. Використовуючи авторизований доступ до ресурсу, система може збирати інформацію про шаблони поведінки авторизованих користувачів протягом багатьох років. Наприклад, більшість сервісів пропонують оцінити переглянутий матеріал. Виставлені користувачами оцінки зберігаються в базі даних. Такий підхід дозволяє в подальшому рекомендувати користувачам найбільш підходящі їм матеріали, товари або інші об'єкти, в залежності від тематики веб-ресурсу. Загалом рекомендації базуються на колективній співпраці користувачів і на виділенні (методом фільтрації) тих користувачів, які демонструють схожі переваги або шаблони поведінки із користувачем, який здійснює пошук певного матеріалу. Колаборативна фільтрація широко використовується в комерційних сервісах і соціальних мережах. Підходи до надання рекомендацій щодо відеоматеріалів, розміщених на YouTube, на основі методу колаборативної фільтрації представлені в роботах [5, 6]. Для аналізу автори використовують готові набори даних, тоді як колекція матеріалів на Youtube постійно поповнюється, тому рекомендаційна інформаційна система повинна працювати в інтерактивному режимі та здійснювати пошук відеоматеріалів в актуальній на даний час колекції.

Гібридні методи фільтрації поєднують підходи колаборативної та контентної фільтрації. Порівняльна характеристика розглянутих методів фільтрації даних у рекомендаційних системах наведена у табл. 1.

У даній роботі для реалізації рекомендаційної інформаційної системи за основу був взятий метод колаборативної фільтрації.

У рекомендаційних системах як правило існує дуже багато матеріалів і користувачів, але кількість оцінених матеріалів може бути незначною, так як користувачі не завжди оцінюють всі об'єкти, які переглядають. Тобто, матриця рейтингів може мати пропуски. Також новому користувачу, який ще не виконував дії в системі, складно що-небудь рекомендувати. Ця проблема, що виникає в системах контентної фільтрації, називається проблемою «холодного старту», коли пасивна поведінка користувачів унеможливує використання рекомендаційної системи. Перевагою використання колаборативної фільтрації є те, що вона не вимагає конкретизованих запитів. Цей метод спрямований на визначення оцінок, які спрощують взаємодію користувачів з системою, за рахунок простих алгоритмів. Для знаходження невідомих переваг користувача використовуються відомі переваги, тобто оцінки інших користувачів в системі.

Таблиця 1 – Порівняльна характеристика методів колаборативної фільтрації

	Колаборативна фільтрація	Контентна фільтрація	Гібридна фільтрація
Переваги	Рекомендації незалежні від контенту сервісу Використовує оцінки користувачів	Немає проблеми «холодного старту» Немає проблеми розрідженості даних	Майже відсутні базові проблеми попередніх методів Висока швидкодія
Недоліки	Проблема «холодного старту» (першого користувача) Розрідженість матриці оцінок	Залежна від контенту Не використовує оцінки користувачів Вузькоспрямована	Висока складність розробки Важкість підтримки Вузькоспрямована
Застосування	Блоги Інформаційні портали	Інтернет-магазини Інтернет-аукціони Блоги	Великі інтернет-магазини Складні соціальні мережі
Приклади	Reddit Digg	Music Genome Project	Netflix

Основними задачами, що має вирішувати розроблена інформаційна система колаборативної фільтрації, є:

- пошук користувачем відеоматеріалів за ключовими словами, розміщених на YouTube;
- надання рекомендацій, тобто формування ранжованого списку N-відеоматеріалів для даного користувача.

Множину даних для рекомендаційної системи формують користувачі, які виставляють оцінки для різних відеоматеріалів. Оцінку можна представити у вигляді триплету – користувач, відеоматеріал, оцінка. Множина всіх триплетів оцінок формує розріджену матрицю, яка називається матрицею оцінок. Пари (користувач, відео), в яких користувачі не виставили оцінку відео, є невідомими значеннями цієї матриці. Для прогнозу оцінки, яку користувач міг виставити за відеоматеріал, використано метод сингулярного розкладу. Це дозволяє позбутися розрідженості матриці оцінок.

дженості матриці оцінок.

Сингулярний розклад матриці рейтингових оцінок користувачів X розміру $m \times n$, де m – користувачі, n – відеоматеріали, буде розкладанням на множини у вигляді USV^T [7,8]:

$$X_{[n \times m]} = U_{[n \times r]} S_{[r \times r]} (V_{[m \times r]})^T,$$

де U (матриця «властивостей» користувачів) – дійсна унітарна матриця розміру $n \times n$; S – прямокутна діагональна матриця розміру $r \times r$; V^T – дійсна унітарною матрицею розміру $m \times m$.

Було використано реалізацію методу сингулярного розкладу матриці за допомогою функції `svds()` бібліотеки SciPy мовою Python. Вхідними даними для функції `svds()` є розріджена матриця рейтингових оцінок відеоматеріалів. У результаті обробки вхідних даних функція `svds()` повертає матрицю з

прогнозованими апроксимованими оцінками кожного відеоматеріалу, із заповненими порожніми місцями в матриці. Після цього, використовуючи дані прогнозованих оцінок, залежно від користувача, авторизованого на сайті, який зробив запит на обчислення рекомендацій, з результуючого набору формується сортована вибірка ідентифікаторів відео (від найкращих до найгірших оцінок) для даного користувача, формується список з URL-адресами цих відеоматеріалів на YouTube, і вони відображаються на веб-інтерфейсі сторінки рекомендацій.

Для якісної роботи алгоритму система повинна мати базу даних, в якій повинні зберігатися дані у вигляді «користувач» – «оцінка» для більшості користувачів в системі. Цей набір даних формується динамічно у процесі перегляду користувачем певного відеоматеріалу та його оцінювання.

Після формування набору даних та виконання сингулярного розкладу матриці оцінок система має готові дані оцінок користувачів, на основі яких на-

даються рекомендації.

Надання рекомендацій для відеоматеріалів виконується за наступним алгоритмом.

1. Користувачі оцінюють переглянуті відеоматеріали за шкалою від 1 до 5.

2. Формується матриця рейтингових оцінок, що містить виставлені оцінки, який подається на вхід функції сингулярного розкладу матриці.

3. На виході з алгоритму повертаються матриці прогнозованих оцінок.

4. Виконується вибірка пар відео-оцінка для конкретного користувача, який зробив запит на знаходження рекомендованих відеоматеріалів.

5. Проводиться сортування вибірки з п.4 у порядку зменшення прогнозованої рейтингової оцінки, після чого виконується відбір 10 відеоматеріалів з найкращими рейтинговими оцінками.

Запропонований алгоритм реалізовано у веб-додатку, архітектура якого у вигляді UML діаграми компонентів представлена на рис. 1.

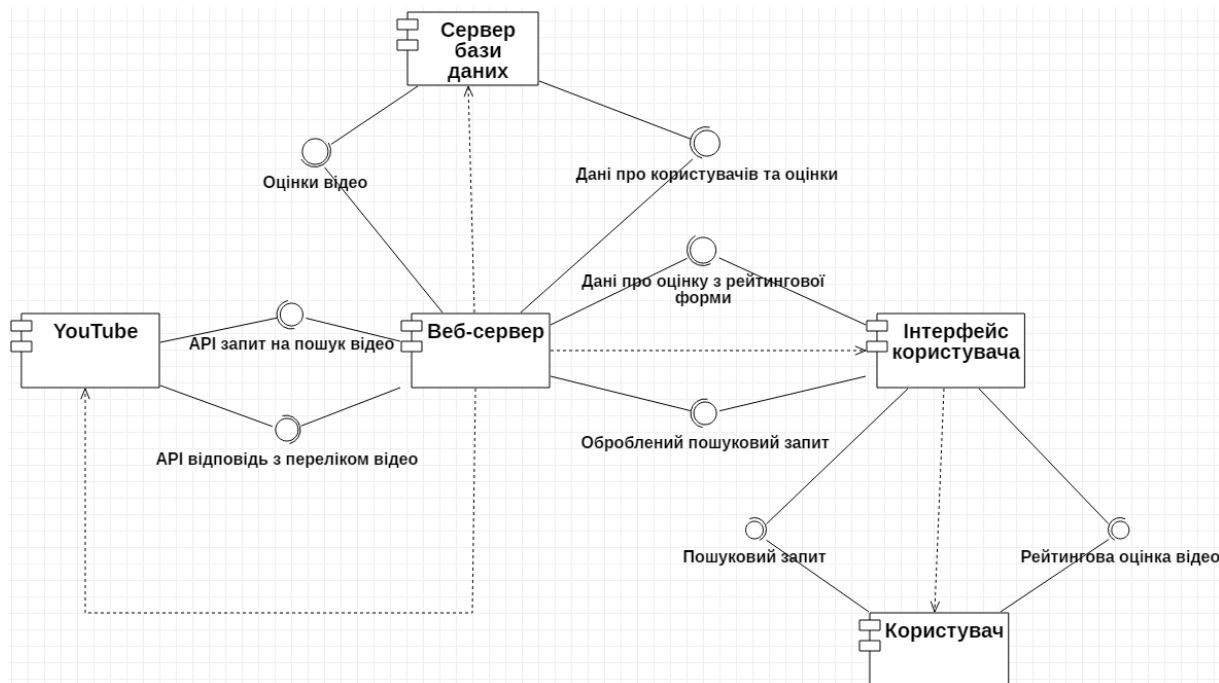


Рисунок 1 – Архітектура рекомендаційної інформаційної системи пошуку відеоматеріалів

Веб-додаток розроблено мовою Python на основі клієнт-серверної архітектури з використанням фреймворку Django [9]. Програмна реалізація рекомендаційної інформаційної системи відбувалась у середовищі PyCharm.

Веб-фреймворк Django використовує шаблон проектування MVC, але його реалізація відрізняється від класичної MVC архітектури. Головна відмінність полягає в тому, що Django має власну вбудовану частину Controller (програмний код, який контролює взаємодію між Model і View), розробнику необхідно створити власний Template. Template – це файл HTML, в якому також використовується Django Template Language. Таким чином, веб-фреймворк Django реалізує архітектурний патерн MVТ.

Головними компонентами розробленого веб-додатку є інтерфейс користувача, веб-сервер та сервер бази даних.

База даних з файлами конфігурації розташована на сервері баз даних MySQL. Було створено наступні таблиці бази даних рекомендаційної системи пошуку відеоматеріалів:

- Video – таблиця з даними про відеоматеріали;
- User – таблиця з даними про користувачів системи;
- Rating – таблиця, в якій зберігаються результати визначення рейтингу кожного відеоматеріалу;
- Rating_users – таблиця, що пов'язує відеоматеріал, користувача та виставлену ним оцінку за відеоматеріал;

Даталогічна модель бази даних рекомендаційної інформаційної системи пошуку відеоматеріалів представлена на рис. 2.

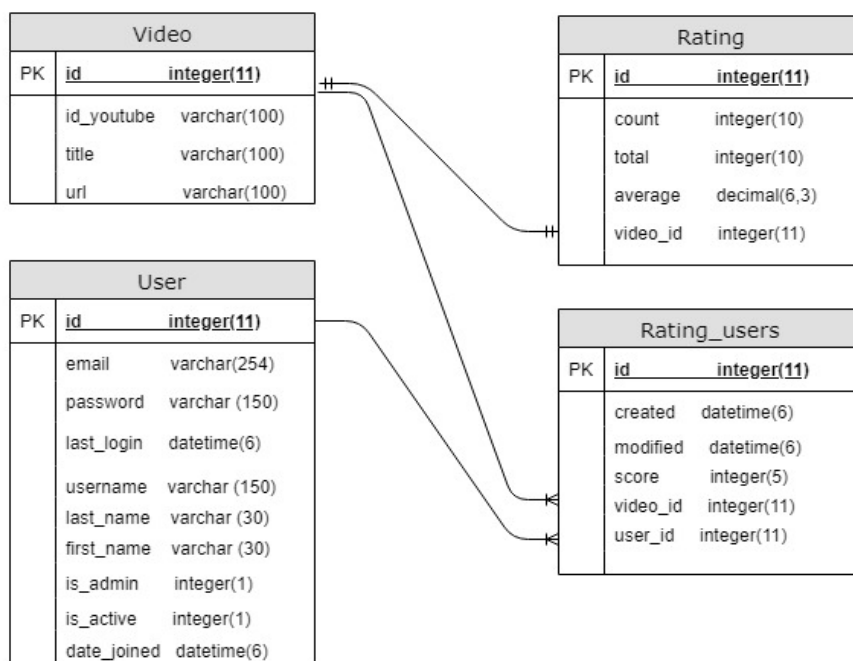


Рисунок 2 – Даталогічна модель бази даних рекомендаційної інформаційної системи пошуку відеоматеріалів

Робота з базою даних може проводитися через вбудований веб інтерфейс адміністратора Django та консольну утиліту manage.py.

Для зручної роботи з базою даних розроблено інтерфейс адміністратора. Особа, що має роль адміністратора в системі, може переглядати вміст усіх таблиць, а також виконувати будь які дії з додавання, видалення або редагування даних через веб-інтерфейс рекомендаційної системи.

Потоки даних розробленої рекомендаційної інформаційної системи в нотатції DFD показано на рис. 3 та рис. 4.



Рисунок 3 – Контекстна діаграма потоків даних

Діаграми потоків даних показують взаємозв'язки між процесами, які виконуються в системі та шлях обробки вхідних даних у вихідні, починаючи з пошукового запиту користувача і закінчуючи списком рекомендованих відеоматеріалів.

Відеоматеріали витягуються з бази даних сервісу YouTube за допомогою API запитів [10]. Цей процес відбувається у результаті обробки пошукового запиту користувача, у відповідь на який повертається список ідентифікаторів відеоматеріалів, які відповідають шуканій тематиці, та заносяться у базу даних, після чого відбувається рендеринг цих відеоматеріалів на сторінці додатку.

Користувач має змогу переглянути знайдені відеоматеріали та проставити рейтингові оцінки відео,

після чого вони заносяться до бази даних. Після кожного нового оцінювання відеоматеріалів система виконує оновлення рейтингу відеоматеріалів та відображає зміни в статистиці рейтингу на сторінці додатку.

Дані з посиланнями на відеоматеріали та їх рейтинги зберігаються у базі даних. Множина даних у вигляді списку відеоматеріалів та їх оцінок, виставлених кожним користувачем, формується на основі даних, взятих з бази даних.

Множина рекомендованих відеоматеріалів обчислюється за допомогою методу колаборативної фільтрації та відображається користувачеві на веб-інтерфейсі. Це дозволяє врахувати при наданні рекомендацій переваги користувачів системи, що переглядали та оцінювали відеоматеріали.

Розроблена рекомендаційна інформаційна система виконує такі функції:

- реєстрація та авторизація користувачів;
- взаємодія з YouTube API для витягу відеоматеріалів на обрану користувачем тематику;
- відображення відеоматеріалів на веб-сторінці додатку;
- оцінка відео-контенту користувачем;
- формування масивів даних для обробки рекомендаційним алгоритмом;
- надання рекомендацій користувачам щодо актуальних відеоматеріалів.

Практична реалізація створеної рекомендаційної інформаційної системи дає змогу отримати список рекомендованих відеоматеріалів для користувача.

Перевагою розробленої рекомендаційної системи є те, що вона реалізує конкретизовані параметри API запиту, який надсилається сервісу YouTube, для пошуку лише тих відеоматеріалів, які відповідають тематиці, що цікавить користувача, та мають високі рейтинги перегляду.

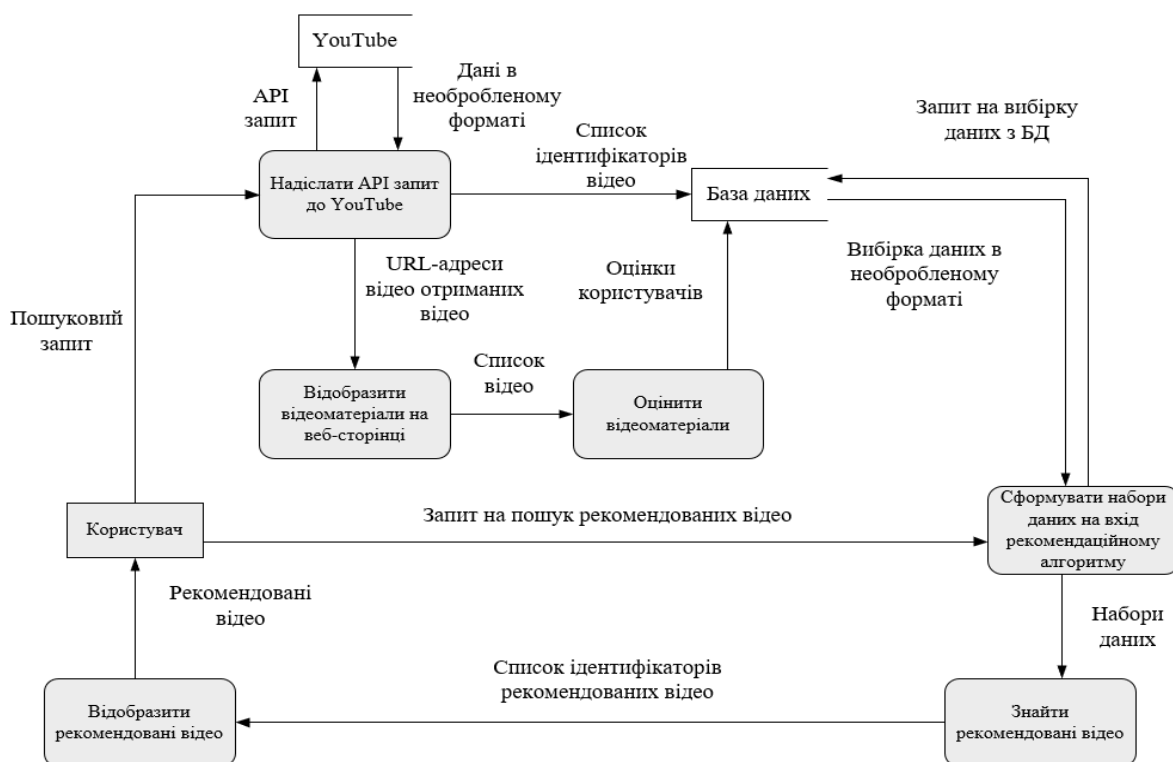


Рисунок 4 – Діаграма декомпозиції потоків даних рекомендаційної інформаційної системи

ВИСНОВКИ. Розроблено рекомендаційну інформаційну систему, яка призначена для знаходження актуальних відеоматеріалів для користувачів системи. Підбір відеоматеріалів базується на методі колаборативної фільтрації та алгоритмі матричної факторизації, який обробляє рейтингові оцінки відеоматеріалів користувачів системи.

Запропонована інформаційна система реалізована як веб-додаток на основі веб-фреймворку Django. Для роботи додатку необхідною умовою є наявність підключення до мережі Інтернет. Доступ до додатку здійснюється через будь-який сучасний веб-браузер.

Інтерфейс веб-додатку складається з декількох веб-сторінок, де користувач вводить дані для авторизації, виконує запит на пошук відеоматеріалів, оцінює переглянуті ним відеоматеріали, а також ознайомлюється з переліком рекомендованих йому відеоматеріалів.

Рекомендаційна інформаційна система може використовуватися як фізичними особами, так і підприємствами малого та середнього масштабу з метою пошуку та фільтрації якісного відео-контенту. Особливо корисною система може бути в освітній галузі та використовуватися для пошуку навчальних відеоматеріалів.

ЛІТЕРАТУРА

1. Sunidhi Sachdeva, Namita Tiwari. Analysis of Various Recommendation Systems. IOSR Journal of Computer Engineering. 2016. Vol. 18. Issue 4. P. 51–55.
2. D. Jannach, M. Zanker, Mouzhi Ge and Marian Gruning. Recommender Systems in Computer Science and Information Systems - a Landscape of Research. Electronic Commerce and Web Technologies: Proceedings of the 13th International Conference, Springer, Vienna, Austria, 2012. P. 76–87.

3. Francesco R., Lior R., Bracha S., Paul K. B. Recommender Systems Handbook. Dordrecht: Springer, 2015. 1009 p.

4. Joonseok Lee, Nisarg Kothari, Paul Natsev. Content-based Related Video Recommendations. Neural Information Processing Systems: Proceedings of the 30th Conference, Barcelona, Spain, 05-10 December, 2016.

5. James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, and Dasarathi Sampath. The Youtube Video Recommendation System. Conference on Recommender Systems: Proceedings of the Fourth ACM Conference, New York, USA, 2010. P. 293–296.

6. Shantanu Thakoor, Anchit Gupta, Parth Shah. Recommending Related YouTube Videos, 2017. URL: <http://snap.stanford.edu/class/cs224w-2017/projects/cs224w-66-final.pdf>.

7. Пятикоп Е. Е. Использование сингулярного разложения матриц в коллаборативной фильтрации. Проблемы информатизації та управління. 2013. № 4 (44). С. 76–81.

8. Brand, Matthew. Fast Online SVD Revisions for Lightweight Recommender Systems. Conference on Data Mining: Proceedings of the SIAM International Conference, 2003. URL: <http://www.merl.com/publications/docs/TR2003-14.pdf>.

9. Samuel Dazon, Aidas Bendoratis, Arun Ravindran. Django Web Development with Python, Birmingham, UK: Packt Publishing Ltd. 2016, 717 p.

10. Leon Yin, Megan Brown. Youtube-data-api Documentation, 2019. URL: <https://buildmedia.readthedocs.org/media/pdf/youtube-data-api/latest/youtube-data-api.pdf>.

RECOMMENDED INFORMATION SYSTEM FOR VIDEO SEARCH

Yu. Parfenenko, A. Kovtun, A. Verbytska

Sumy State University

vul. Rymkogo-Korsakova, 2, Sumy, 40007, Ukraine. E-mail: yuliya_p@cs.sumdu.edu.ua

Purpose. The aim of the study is to develop recommended information system to assist in efficient search for videos to the desired topic on YouTube. **Methodology.** The recommendation system is based on the collaborative filtering method which uses similar preferences of users who search for a concrete video. The input data for the recommendation algorithm: video id, user id, evaluation of video by the user. These data form a sparse matrix of rating, filling the gaps in which is implemented by the method of singular value decomposition. It allows predicting the rating for videos that are not rated by users. **Results.** The recommended information system is implemented as a web application using the Django framework and MVT architecture. The main components of the developed web application are the user interface, web server and database server. The database of the web-application is managed by MySQL database management system. The main functions of the system are user's registration and authorization; interaction with the YouTube API to access to video on user-selected topics; recommendation of the top 10 relevant video materials on the requested subject to user. **Originality.** The implementation of the proposed recommendation information system allows users to get a list of recommended video materials for the user excluding non-relevant video content. The specifying of API parameters when requesting video allows to find only those videos that are relevant to topics of interest to the user and have high view ratings. **Practical value.** The obtained results can be used in education especially self-study education, because of there are a lot of learning video on different YouTube channels. As a rule, users don't sufficiently aware in subject they are going to study, so they need advice on which video contains more understandable and relevant information. The developed recommendation information system may assist in searching learning videos. **Conclusions.** The developed information system allows users to make target search videos, evaluate viewed videos and recommend users videos on certain subject. The future research direction lies in improving the recommendations logic through the implementation of additional recommendation methods and interaction between them.

Key words: information system, collaborative filtering, search, video, YouTube.

REFERENCES

1. Sunidhi, Sachdeva, Namita, Tiwari (2016), "Analysis of Various Recommendation Systems", *IOSR Journal of Computer Engineering*, vol. 18, no. 4, pp. 51-55.
2. Jannach, D., Zanker, M., Mouzhi, Ge, Grüning, Marian (2012), "Recommender Systems in Computer Science and Information Systems - a Landscape of Research", *Proceedings of the 13th International Conference on Electronic Commerce and Web Technologies*, Springer, Vienna, Austria, pp. 76-87.
3. Francesco, R., Lior, R., Bracha, S., Paul, K. B. (2015), *Recommender Systems Handbook*, Dordrecht: Springer.
4. Joonseok, Lee, Nisarg, Kothari, Paul, Natsev (2016), "Content-based Related Video Recommendations", *Neural Information Processing Systems: Proceedings of the 30th Conference*, Barcelona, Spain.
5. Davidson, James, Liebal, Benjamin, Liu, Junning, Nandy, Palash, Van Vleet, Taylor, Ullas, Gargi, Sujoy, Gupta, Yu, He, Mike, Lambert, Livingston, Blake, Dasarathi, Sampath (2010), "The Youtube Video Recommendation System", *Proceedings of the Fourth ACM Conference on Recommender Systems*, New York, USA, pp. 293-296.
6. Shantanu, Thakoor, Anchit, Gupta, Parth, Shah (2017), *Recommending Related YouTube Videos*. Available from: <http://snap.stanford.edu/class/cs224w-2017/projects/cs224w-66-final.pdf>. [15 July 2019].
7. Piatykop, Ye. (2013) Ispolzovanie singulyarnogo razlozheniya matrix v kollaborativnoy filtratsii [Using singular decomposition of matrices in collaborative filtering], *Problemy informat`zatsiyi ta upravlinnya*, no. 4 (44), pp. 76-81.
8. Brand, Matthew (2003), "Fast Online SVD Revisions for Lightweight Recommender Systems", *Proceedings of the SIAM International Conference*, 2003. Available from: <http://www.merl.com/publications/docs/TR2003-14.pdf>. [15 July 2019].
9. Dazon, Samuel, Bendoratis, Aidas, Ravindran, Arun (2016), *Django Web Development with Python*, Birmingham, UK: Packt Publishing Ltd.
10. Yin, Leon, Brown, Megan (2019), *Youtube-data-api Documentation*. Available from: <https://buildmedia.readthedocs.org/media/pdf/youtube-data-api/latest/youtube-data-api.pdf>. [15 July 2019].

Стаття надійшла 05.09.2019.