

УДК 004.9

DOI: 10.20998/2411-0558.2021.01.09

*С. А. ПОКРИШКА*, асп., СНУ ім. В. Даля, Сєверодонецьк,  
*Л. О. ШУМОВА*, канд. техн. наук, доц., СНУ ім. В. Даля,  
Сєверодонецьк

## **УДОСКОНАЛЕННЯ РЕКОМЕНДАЦІЙНОЇ ВЕБОРІЄНТОВАНОЇ СИСТЕМИ З ВИКОРИСТАННЯМ КОЛАБОРАТИВНОЇ ФІЛЬТРАЦІЇ**

Представлено вирішення задачі удосконалення рекомендаційної веборієнтованої системи за рахунок колаборативної фільтрації для надання рекомендацій користувачу вебсервісу зі статтями на тему ігор. В ході досліджень систематизовано сукупність етапів розробки та удосконалення рекомендаційної веборієнтованої системи. Представлена структура бази даних та метод колаборативної фільтрації з запропонованою мірою подібності для формування рекомендацій користувачам. Використано попереднє структурування статей за категоріями для виявлення вподобань користувачів та надання рекомендацій. Використання колаборативної фільтрації з запропонованою мірою подібності дозволило підвищити кількість переглянутих користувачами статей. Лл.: 2. Табл.: 1. Бібліогр.: 10 назв.

**Ключові слова:** рекомендаційна веборієнтована система; колаборативна фільтрація; бази даних; структурування статей за категоріями; міра подібності.

**Постановка проблеми.** Зростання кількості різноманітних вебсервісів обумовлює все більш жорсткі вимоги до їх конкурентоспроможності. Надання користувачу вебсервісу персоналізованого контенту є одним із методів, що робить взаємодію користувача з ресурсом більш комфортним. Застосування технології формування рекомендацій призводить до того, що споживачі все частіше віддають переваги ресурсам, що пропонують їм цікавий персоналізований контент. У сучасному інтернет-просторі використання рекомендацій контенту є важливою складовою популярності ресурсу, а їх відсутність може вважатись недоліком.

У зв'язку зі стрімко зростаючим числом вебсервісів, користувачу складно орієнтуватися у величезному обсязі інформації, тому актуальним завданням можна вважати удосконалення рекомендаційної веборієнтованої системи, орієнтованої на користувача.

**Аналіз останніх досліджень і публікацій.** Формування рекомендацій – це одне із важливих завдань інтелектуального вебсервісу будь-якого напрямку. Рекомендації дозволяють підбирати найкращі варіанти товару, музики чи фільму для користувача. Одна із важливих частин розробки – це вибір ефективного методу надання рекомендацій у

відповідності до індивідуальних завдань вебсервісу, тому варіантів реалізації безліч і кожен має як переваги та недоліки. Більшість досліджень у галузі рекомендаційних систем для надання рекомендацій використовують метод колаборативної фільтрації [1, 2] з опціональною мірою подібності, що використовується при фільтрації. Метод колаборативної фільтрації надає наступні переваги. Застосування цього методу не потребує детального опису досліджуваних елементів. Також, метод колаборативних фільтрації не залежить від відсутності інформації про елемент і дозволяє реєструвати зміну інтересів користувачів з плином часу і відстежувати приховані чинники, що забезпечують гнучкість щодо переваг користувача.

Розрізняють три типи методів формування рекомендацій [1]: фільтрація на основі вмісту, фільтрація на основі знань та колаборативна фільтрація, які можуть давати рекомендації від великих інформаційних груп відповідно до вподобань користувача.

Підходи до фільтрації на основі змісту [2] враховуються тільки минулі вподобання окремого користувача і намагаються вивчити модель вподобань, засновану на функціональному поданні вмісту рекомендованих елементів.

Метод фільтрації, заснований на знаннях, рекомендує елемент, обчислюючи вподобання користувача, базуючись на інформації про елементи рекомендацій, критерії вподобань користувача та критерії рекомендацій [3].

Колаборативна фільтрація – це метод автоматичного прогнозування (фільтрування) інтересів користувача шляхом збирання інформації про вподобання багатьох користувачів. Основна ідея колаборативної фільтрації заснована на ідентифікації шаблонів вподобань в співтоваристві користувачів та полягає в тому, що коли два користувачі з однаковими інтересами обрали один і той же елемент, ймовірність того, що в них буде однаковий інтерес в майбутньому, буде високою. Колаборативна фільтрація є широко використовуваним методом формування рекомендацій завдяки тому, що цей метод враховує тільки вподобання користувача та не враховує особливості або зміст рекомендованих об'єктів [4]. Також, метод колаборативної фільтрації надає наступні переваги. Застосування цього методу не потребує детального опису досліджуваних елементів. Також, метод колаборативних фільтрації не залежить від відсутності інформації про елемент і дозволяє реєструвати зміну інтересів користувачів з плином часу і відстежувати приховані чинники, що забезпечує гнучкість щодо переваг користувача.

У дослідженні [3] представлено розроблений алгоритм на основі  $k$ -means та обґрунтовано його переваги в порівнянні з алгоритмом  $k$ -means та сингулярним розкладом матриці. Новий алгоритм базується на зменшенні розмірності та кластерних технологіях. Для оцінки якості рекомендацій автор використовує середньоквадратичну помилку.

У роботі [4] автор досліджує застосування сингулярного розкладу матриці, кластерного аналізу в алгоритмах колаборативної фільтрації. Автором застосовано вдосконалені методи вимірювання схожості, які використовуються для пошуку найближчого сусіда користувача та набору рекомендованих кандидатів у кластері. Результати показали перевагу алгоритма формування рекомендацій з використанням методу колаборативної фільтрації на основі кластеризації.

В дослідженні [5] запропоновано алгоритми рекомендацій на основі інтегрованої колаборативної фільтрації та стохастичного градієнтного спуску на прикладі формування рекомендацій в соціальній кіберфізичній системі. Запропонований алгоритм використовує співвідношення між даними та враховує схожість між елементами, схожість між користувачами та два типи довіри серед користувачів для вибору найближчих сусідів обох користувачів та елементів, що надають найціннішу інформацію для рекомендацій. Ефективність використаних алгоритмів оцінено з використанням показників середньої абсолютної помилки та середньоквадратичної помилки.

В роботі [6] представлено розробку методів колаборативної фільтрації з використанням асоціативних правил на основі генетичного алгоритму. Порівнюють їх через характеристики точності, F1-міри та часу виконання.

Більшість проаналізованих досліджень використовують для визначення подібності користувачів алгоритми, основані на методах машинного навчання. Зазвичай застосування методів машинного навчання має певні обчислювальні обмеження, що можуть впливати на швидкість завантаження контенту. Тому використання методів машинного навчання для визначення подібності не завжди є доцільним, особливо для невеликих вебресурсів. Це обумовлює актуальність дослідження методу колаборативної фільтрації для надання рекомендацій та мір подібності без застосування методів машинного навчання.

**Метою дослідження** є удосконалення вказаної рекомендаційної веборієнтованої системи, контент якої складають статті на тему ігор, з використанням колаборативної фільтрації та запропонованої міри визначення подібності.

**Основна частина передбачає** удосконалення вказаної рекомендаційної веборієнтованої системи, контент якої складають статті на тему ігор.

*Етап 1.* Розробка бази даних. База даних є частиною вебсервісу. Розробка бази даних включає визначення сутностей та зв'язків між ними. Сутності визначено наступним чином: користувачі (users) і статті (article). Кожну сутність характеризує набір залежних підсутностей.

*Етап 2.* Налаштування категорій статей. В базу даних було додано та розподілено відповідно до категорій множина статей. Для більш швидкого обчислення в базу даних для кожного користувача додається лічильник перегляду кожної з категорій.

*Етап 3.* Визначення схожих користувачів та генерування рекомендацій. Цей етап складається з наступних кроків:

- визначення подібності користувачів;
- генерування рекомендацій.

В якості критерію схожості  $s$  запропоновано використання коефіцієнту, що розраховується як сума різниць переглядів між обраним користувачем  $u$  та іншими. Чим менша різниця, тим більша схожість користувачів.

Для генерування рекомендації статей використовується показник історії переглядів категорій, після чого виключаються з рекомендацій спільні статті та рекомендуються статті відсортовані по часу публікації.

Для персоналізації контенту на основі попередньої історії переглядів користувача удосконалено метод колаборативної фільтрації: колаборативна фільтрація застосовується до даних перегляду статей за категоріями (1):

$$r_{u,i} = \frac{\sum_{u' \in U} s(u, u') r_{u',i}}{\sum_{u' \in U} |s(u, u')|}, \quad (1)$$

де  $r_{u',i}$  – переглянута стаття користувачем  $u'$  в категорії  $i$ ;

$r_{u,i}$  – рекомендована стаття в категорії  $i$ ;

$U$  – множина подібних користувачів;

$u'$  – користувач, подібний до користувача  $u$  з множини  $U$ ;

$s(u, u')$  – критерій подібності користувача  $u$  та користувача  $u'$ .

В якості критерію схожості  $s$  запропоновано міру, що розраховується як сума різниць переглядів між обраним користувачем  $u$  та іншими наступним чином (2):

$$s = \sum_{i=1 \dots n, j=1 \dots m} (|u_{i,j} - u_{i+1,j}|, \dots, |u_{i,j} - u_{n,j}|), \quad (2)$$

де  $u_{j,i}$  позначає  $j$ -го користувача, що переглянув статтю категорії  $i$ .

### Результати експерименту.

Експеримент проведено з використанням розробленої рекомендаційної веборієнтованої системи. В статті наведено дані на момент проведення експерименту, але вказані параметри є масштабованими і можуть бути збільшені у майбутньому.

**Розробка бази даних.** Складовою частиною розробленої рекомендаційної веборієнтованої системи є база даних, яка містить дані про зареєстрованих користувачів (рис. 1) та про статті по ігровій тематиці (рис. 2).

u528782nwt_collaboration users	
id	int
login	varchar(40)
pass	varchar(100)
email	varchar(100)
dota2	int
minecraft	int
csgo	int
gta	int
fortnite	int
lol	int
wow	int

Рис. 1. Структура сутності users

u528782nwt_collaboration article	
id	int
name	varchar(100)
text	varchar(2000)
category	varchar(200)

Рис. 2. Структура сутності article

**Налаштування категорій статей.** На цьому етапі на розробленому вебсервісі визначено 7 категорій статей на тему комп'ютерних ігор. Визначені категорії статей представлені в табл. 1.

Таблиця 1.

Відповідність категорій до тематики статей

Тематика статті	Категорія статті
Dota 2	dota2
Counter Strike	csgo
Fortnite	fortnite
Minecraft	minecraft
League of Legend	lol
World of Warcraft	wow
Grand Theft Auto	gta

Всім існуючим статтям вручну було присвоєно категорію.

**Визначення схожих користувачів та генерування рекомендацій.**

Застосування методу колаборативної фільтрації для відслідковування активності зареєстрованих користувачів вебсервісу. Для персоналізації контенту на основі попередньої історії перегляду використано метод колаборативної фільтрації. Колаборативна фільтрація застосовується до даних переглянутих статей за категоріями. Коли користувач переходить до статті, в базі даних інкримінується показник відповідної категорії. Таким чином чим більше користувач переглядає статті в категорії, тим більше показник категорії в таблиці даного користувача. Після чого потрібно обрати найбільш схожого користувача. Для цього розраховується коефіцієнт  $s$  як сума різниць модулів даного користувача з кожним зареєстрованим на вебсервісі, наступним чином:

$$s = |dota2\_user-dota2\_similar| + |csgo\_user-csgo\_similar| + |fortnite\_user-fortnite\_similar| + |minecraft\_user-minecraft\_similar| + |lol\_user-lol\_similar| + |wow\_user-wow\_similar| + |gta\_user-gta\_similar|$$

На вебсервісі зареєстровано близько 1000 користувачів, таким чином процедура повторюється 1000 разів.

В результаті отримано найменшу різницю і найбільш схожого користувача. Знайдемо модуль різниці кожної з категорій користувача з найбільш схожим користувачем.

Дані поточного користувача:

Dota2 = 16; Csgo = 77; Fortnite = 66; Minecraft = 22; Lol = 1; Wow = 4;  
Gta = 1;

Дані схожого користувача:

Dota2 = 21; Csgo = 79; Fortnite = 62; Minecraft = 43; Lol = 38; Wow = 1;  
Gta = 14.

Отримуємо суму різниць модулів  
 $|16 - 21| + |77 - 79| + |66 - 62| + |22 - 43| + |1 - 38| + |4 - 1| + |1 - 14| = 85$ .

В результаті отримано суму різниць модулів категорій 85.

Ефективність запропонованого методу визначається за допомогою середнього показника прочитаних статей зареєстрованими користувачами на вебсервісі. Для визначення продуктивності запропонованого міри подібності для надання рекомендацій використано дані за 10 днів без застосування методу колаборативної фільтрації для рекомендування статей та з використанням методу колаборативної фільтрації.

Статистика вебсервісу:

Кількість зареєстрованих користувачів: ~ 1000,

Кількість статей на сайті: 105.

Середній показник читання статей без використання рекомендаційного модулю на 1 користувача за 1 сесію: 1, 3.

Середній показник читання після впровадження рекомендаційного модулю: 2, 1.

### **Висновки.**

В роботі представлено розробку та удосконалення рекомендаційної веборієнтованої системи з використанням колаборативної фільтрації. В роботі запропоновано міру визначення подібності користувачів для застосування в методі колаборативної фільтрації для надання рекомендацій. За результатами проведеного експерименту, запропонований коефіцієнт подібності дозволив удосконалити надання рекомендацій. Ефективність надання рекомендацій оцінена на підставі кількості прочитаних користувачами статей. Середній показник читання після впровадження системи збільшився з 1.3 до 2.1 статті на одного користувача за одну сесію перебування на сайті.

Таким чином, запропонований коефіцієнт подібності, що використовується в методі колаборативної фільтрації з метою надання рекомендацій, дозволив удосконалити рекомендаційну веборієнтовану систему та підвищити ефективність надання рекомендацій. Підвищення ефективності рекомендацій та збільшення залученості користувача надає перспективи збільшення часу перебування користувача на сайті, що є складовим критерієм пошукової оптимізації сайту для підняття позиції сайту в результатах пошукових систем.

Напрямок майбутніх досліджень спрямований на дослідження та аналіз критеріїв точності та ефективності наданих рекомендацій.

**References:**

1. Kato, Y., and Yamamoto, K. (2020), "A sightseeing spot recommendation system that takes into account the visiting frequency of users", *ISPRS International Journal of Geo-Information*, Vol. 9, No. 7, pp. 1-21.
2. Lops, P. et al. (2019), "Trends in content-based recommendation", *User Modeling and User-Adapted Interaction*, Vol. 29, No. 2, pp. 239-249.
3. Laseno, F.U.D., and Hendradjaya, B. (2019), "Knowledge-Based Filtering Recommender System to Propose Design Elements of Serious Game", *2019 International Conference on Electrical Engineering and Informatics (ICEEI)*, pp. 158-163.
4. Gorakala, S.K. (2015), *Usuelli M. Building a recommendation system with R*, Packt Publishing Ltd.
5. Collaborative filtering | Stanford University. URL: <https://www.youtube.com/watch?v=h9gpufJFF-0> (accessed 31.03.2021)
6. Collaborative filtering. URL: <https://www.coursera.org/lecture/machine-learning/collaborative-filtering-2WoBV> (accessed 05.03.2021)
7. Zarzour, H. et al. (2018), "A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques", *2018 9th International Conference on Information and Communication Systems (ICICS)*, pp. 102-106. doi:10.1109/iacs.2018.8355449
8. Xiaojun, L. (2017), "An improved clustering-based collaborative filtering recommendation algorithm", *Cluster Computing*, Vol. 20, No. 2, pp. 1281-1288, doi:10.1007/s10586-017-0807-6.
9. Xu, J. et al. (2017), "Integrated collaborative filtering recommendation in social cyber-physical systems", *International Journal of Distributed Sensor Networks*, Vol. 13, No. 12, 1550147717749745. doi:10.1177/1550147717749745.
10. Neysiani, B. S., Soltani, N., Mofidi, R., and Nadimi-Shahraki, M.H. (2019), "Improve performance of association rule-based collaborative filtering recommendation systems using genetic algorithm", *Int. J. Inf Technol. Comput. Sci*, Vol. 2, pp. 48-55. doi: 10.5815/ijitcs.2019.02.06.

*Статтю представив д-р техн. наук, проф. НТУ "ХПІ" Леонов С.Ю.*

*Надійшла (received) 14.04.2021*

Pokryshka Sergey, PhD student  
Volodymyr Dahl East Ukrainian National University  
59-a Central avenue, Severodonetsk, Luhansk region, Ukraine, 93400  
tel./phone: (064) 522-89-97, e-mail: xakermans@gmail.com

Shumova Larisa, Candidate of Engineering Sciences  
Volodymyr Dahl East Ukrainian National University  
59-a Central avenue, Severodonetsk, Luhansk region, Ukraine, 93400  
tel./phone: (064) 522-89-97, e-mail: shumova@ukr.net  
ORCID ID: 0000-0001-8113-5115



УДК 004.9

**Удосконалення рекомендаційної веборієнтованої системи за рахунок колаборативної фільтрації / Покришка С.А., Шумова Л.О. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2021. – № 1 (5). – С. 115 – 123.**

Представлено вирішення задачі удосконалення рекомендаційної веборієнтованої системи за рахунок колаборативної фільтрації для надання рекомендацій користувачу вебсервісу зі статтями на тему ігор. В ході досліджень систематизовано сукупність етапів розробки та удосконалення рекомендаційної веборієнтованої системи. Представлена структура бази даних та метод колаборативної фільтрації з запропонованою мірою подібності для формування рекомендацій користувачам. Використано попереднє структурування статей за категоріями для виявлення вподобань користувачів та надання рекомендацій. Використання колаборативної фільтрації з запропонованою мірою подібності дозволило підвищити кількість переглянутих користувачами статей. Іл.: 2. Табл.: 1. Бібліогр.: 10 назв.

**Ключові слова:** рекомендаційна веборієнтована система; колаборативна фільтрація; база даних; структурування статей за категоріями; міра подібності.

УДК 004.9

**Совершенствование рекомендательной вебориєнтованой системы за счет колаборативных фильтрации / Покришка С.А., Шумова Л.А. // Вестник НТУ "ХПІ". Серія: Інформатика и моделирование. – Харьков: НТУ "ХПІ". – 2021. – № 1 (5). – С. 115 – 123.**

Представлено решение задачи усовершенствования рекомендательной вебориєнтованной системы за счет использования колаборативной фильтрации для представления рекомендаций пользователям вебсервиса со статьями на игровую тематику. В ходе исследований систематизирована совокупность этапов разработки и усовершенствования рекомендательной вебориєнтованой системы. Представлена структура базы данных и метод колаборативной фильтрации использованием предложенной мерой сходства для формирования рекомендаций пользователям. Использовано предварительное структурирование статей по категориям для выявления предпочтений пользователей и предоставления рекомендаций. Использование колаборативных фильтрации с предложенной мерой сходства позволило повысить количество просмотренных пользователями статей. Ил.: 2. Табл.: 1. Библиогр.: 10 названий.

**Ключевые слова:** рекомендационная вебориєнтованная система, колаборативная фильтрация, база данных; структурирование статей по категориям; мера сходства.

UDC 004.9

**Improving the web-based recommender system through collaborative filtering / Pokrishka S.A., Shumova L.A. // Herald of the National Technical University "KhPI". Series of "Informatics and Modeling". – Kharkov: NTU "KhPI". – 2021. – № 1 (5). – P. 115 – 123.**

The solution to the problem of improving the recommendation web-oriented system through the use of collaborative filtering to provide recommendations to users of the web service with articles on gaming topics is presented. In the course of the research, a set of stages of development and improvement of the recommendatory web-oriented system was systematized. The structure of the database and the method of collaborative filtering using the proposed similarity measure to form recommendations for users are presented. The preliminary structuring of articles by categories was used to identify user preferences and provide recommendations. The use of collaborative filtering with the proposed similarity measure allowed us to increase the number of articles viewed by users. Ill.: 2. Bibliography: 10 titles.

**Keywords:** recommendation web-oriented system; collaborative filtering; database; structuring of articles by categories; measure of similarity.