



**ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ОСВІТІ,
МИСТЕЦТВІ ТА КУЛЬТУРІ**
**INFORMATION TECHNOLOGIES IN EDUCATION,
ARTS AND CULTURE**

УДК 004.8:[005.953.2:002]:[159.942:028.2-043.4
DOI: 10.31866/2617-796X.8.1.2025.335530

Адріан Другак,

*магістрант, кафедра системотехніки,
Харківський національний університет радіоелектроніки,
Харків, Україна
adrian.druhak@nure.ua
<https://orcid.org/0009-0009-1303-0608>*

Андрій Коваленко,

*кандидат технічних наук, старший науковий співробітник,
доцент кафедри системотехніки,
Харківський національний університет радіоелектроніки,
Харків, Україна
andrey.kovalenko@nure.ua
<https://orcid.org/0000-0003-2882-5082>*

Олена Чайковська,

*кандидат педагогічних наук,
професор кафедри інформаційної діяльності
та зв'язків з громадськістю,
Київський національний університет
культури і мистецтв,
Київ, Україна
oachaikovska@knukim.edu.ua
<https://orcid.org/0000-0001-7769-1004>*

**РЕКОМЕНДАЦІЯ КНИГ З УРАХУВАННЯМ ЕМОЦІЙ,
ЩО МІСТЯТЬСЯ В ТЕКСТОВИХ ВІДГУКАХ ЧИТАЧІВ**

Метою статті є дослідження методів отримання даних емоційного аналізу текстових відгуків читачів для їх використання в рекомендаційних функціях систем електронної комерції, що надають послуги онлайн-читання книг або прослуховування аудіокниг.

© Другак А. О.,
© Коваленко А. І.,
© Чайковська О. А.

Методи дослідження – системний підхід, методи системного аналізу та синтезу, методи структурного моделювання реляційних баз даних.

Наукова новизна полягає у розробці модифікованого методу визначення рекомендацій сумісної фільтрації на основі користувачів (Collaborative Filtering User Based), що враховує дані емоційного аналізу текстових відгуків на прочитані книги.

Висновки. У цій статті розглянуто три підходи до оцінювання емоцій текстових відгуків читачів. За першим підходом використовуються лексикони-словники, слова в яких індексуються за наборами емоцій. Оцінки емоцій тексту розраховуються за статистикою входження слів до наборів емоцій у лексиконі. За другим підходом використовується мовна модель BERT, яка дає змогу отримати оцінки емоцій за запитом до неї. Мовна модель BERT – це нейронна мережа, яка забезпечує можливість урахування смислового значення слів у реченні зі словами, що розташовані ліворуч і праворуч від них. За третім підходом використовується трансформер-бот ChatGPT. Оцінка емоцій, що міститься в тексті відгуків, може здійснюватися в реальному часі за допомогою функції API ChatGPT. Вихідними параметрами функції API є текст установки до аналізатора настроїв ChatGPT і текст відгуку, а вхідним параметром – дані оцінки емоцій у форматі JSON.

З урахуванням розглянутих підходів розроблено модифікований рекомендаційний метод спільної фільтрації на основі користувачів, який враховує дані емоційного аналізу. Для цього запропоновано на етапі відбору даних урахувати усереднену косинусну подібність рейтингових оцінок та оцінок емоцій текстових відгуків користувачів. Запропонований метод визначення рекомендацій може використовуватися в бібліотечних системах з реалізованою функцією користувача для надання відгуків на прочитані книги.

Ключові слова: аналіз емоцій; рекомендаційні системи; спільна фільтрація; книги; методи оцінювання рейтингу; розширення для рекомендаційних систем.

Вступ. Одним зі шляхів поширення використання систем електронної комерції є використання рекомендаційних функцій. На сьогодні розроблено великий спектр рекомендаційних методів, що дають можливість визначити дані, які характеризують уподобання клієнтів. За допомогою цих методів, реалізованих як автоматизовані функції системи, формуються різні за призначенням рекомендації клієнтам із пропозицією придбання товарів, зокрема книг.

Книги є не лише одним з головних джерел знань та інформації для людства, але й джерелом емоцій. Для багатьох людей у світі читання книг є хобі, яке дає змогу отримати натхнення, пом'якшити стрес, полегшити зневіру чи самотність, стимулювати такі позитивні емоції, як сміх та радість. Для підбору книг використовують класичні методи визначення рекомендацій, які базуються переважно на аналізі історії покупок або онлайн-читання (прослуховування) та врахування рейтингових оцінок. На сьогодні набули широкого застосування методи машинного навчання, зокрема обробки природної мови (NLP), для визначення емоцій, що містяться в тексті. Це дозволяє змінити чи модифікувати методи визначення рекомендацій. У зв'язку з цим має практичний інтерес розгляд особливостей реалізації рекомендаційних функцій для бібліотечних систем електронної комерції, які враховують емоції, що містяться в текстових відгуках на прочитані або придбані книги.

Результати дослідження. Дослідження проводили за трьома етапами. На першому проведено аналіз та визначено методи для отримання даних емоцій-

ного аналізу тексту, які можна застосовувати для реалізації рекомендаційних функцій. На другому етапі розроблено реляційну базу даних рекомендаційної системи. На третьому етапі розроблено модифікований рекомендаційний метод сумісної (колоборативної) фільтрації на основі користувачів, який дає змогу враховувати емоції, що містяться в текстових відгуках на прочитані книги.

Аналіз наявних підходів для отримання даних емоційного аналізу тексту дав змогу відокремити такі методи:

- метод визначення емоцій з використанням лексиконів;
- метод визначення емоцій з використанням мовної моделі BERT;
- метод визначення емоцій з використанням трансформера-бота ChatGPT.

Метод визначення емоцій у тексті з використанням лексиконів (словників). Метод визначення емоцій за допомогою лексиконів (Mohsen, Hassan and Idrees, 2016) передбачає розрахунок коефіцієнтів TF-IDF (Term Frequency – Inverse Document Frequency). Для оцінки емоцій, що містяться в текстових документах, використовується спеціально розроблений лексикон NRC. Лексикон NRC (National Research Council Canada) є словником, що містить проіндексовані слова, які асоціюються з чотирма наборами емоцій:

- набір «Страх» (Fear), який містить слова, що пов'язані з емоціями тривоги, небезпеки, загрози; приклади слів: «страшний», «жахливий», «моторошний» тощо;
- набір «Сум» (Sadness), який містить слова, що пов'язані з емоціями втрати, розчарування або суму; приклади слів: «сумний», «трагічний», «розчарування» тощо;
- набір «Радість» (Joy), який містить слова, що пов'язані з емоціями щастя, задоволення і насолоди; приклади слів: «щасливий», «радісний», «чудовий» тощо;
- набір «Здивування» (Surprise), який містить слова, пов'язані з емоціями здивування, несподіванки або шоку; приклади слів: «несподіваний», «дивовижний», «приголомшливий» тощо.

Слід зазначити, що деякі слова можуть одночасно входити до складу кількох наборів емоцій. Фрагмент лексикону NRC, що містить індекс слова (index), саме слово (word) та назву набору емоцій (emotion), з яким асоціюється слово, подано на рис. 1.

index	word	emotion
0	відмовитися	fear
1	покинутий	fear
2	розпуста	joy
3	достаток	joy
4	відмовитися	sadness
5	покинутий	sadness
6	залишення	surprise
7	викрадення	surprise

Рис. 1. Фрагмент лексикону NRC

Для визначення емоцій у тексті відгуку на книгу розраховувати коефіцієнти TF-IDF необов'язково. Оцінку емоцій можна визначити за статистикою входжен-

ня слів до емоційних наборів лексикону NRC. Розглянемо процес визначення емоцій за прикладом текстового відгуку на книгу.

Текст відгуку (англ.): *This book pleasantly surprised me and made me smile sincerely. The story is heartwarming, cheerful, and truly delightful.*

Текст відгуку (укр.): Ця книга приємно здивувала мене та змусила щиро усміхнутися. Історія зворушлива, весела та справді захоплива.

Перший крок – приведення символів букв до нижнього регістру, токенизація слів у тексті та видалення стоп-слів:

– [«book», «pleasantly», «surprised», «made», «smile», «sincerely», «story», «heartwarming», «cheerful», «delightful»];

– [«книга», «приємно», «здивувала», «змусила», «щиро», «усміхнутися», «історія», «зворушлива», «весела», «захоплива»].

Другий крок – визначення слів, що пов'язані з емоціями (табл. 1). Після обробки тексту відгуку залишається 10 слів. З них пов'язані з емоціями 6 слів. З емоцією радості пов'язано 5 слів, а з емоцією здивування – 1 слово. Цей результат відразу надає можливість підрахувати відсоток присутності емоцій у тексті. Оцінки емоцій розраховуються за статистикою приналежності слів до наборів емоцій, як продемонстровано в табл. 1.

Таблиця 1

Приклад визначення оцінок, що пов'язані з емоціями

НАБІР ЕМОЦІЙ	ЕМОЦІЙНІ СЛОВА	КІЛЬКІСТЬ СЛІВ	ДАНІ ОЦІНКИ ЕМОЦІЙ
Радість (Joy)	усміхнутися (smile), зворушлива (heartwarming), весела (cheerful), приємно (pleasantly), захоплива (delightful)	5 слів з 6	$5/6 = 0,83$ (83 %)
Здивування (Surprise)	здивувала (surprised)	1 слово з 6	$1/6 = 0,17$ (17 %)
Сум (Sadness)	немає	0	0
Страх (Fear)	немає	0	0

Для реалізації цього методу можна використовувати лексикон NRC, який є у вільному доступі для некомерційного використання. Інший шлях – реалізувати власний лексикон або чотири вектори-словники, які відповідають емоціям радості, здивування, суму та страху. Оцінки емоцій розраховуються за статистикою приналежності слів до векторів за прикладом, наведеним у табл. 1. Під час розрахунку статистики потрібно враховувати всі слова, у тому числі й ті, що повторюються.

Метод визначення емоцій у тексті з використанням мовної моделі *BERT*. *BERT* (Bidirectional Encoder Representations from Transformers) – це модель представлення мови, яку розробила компанія Google у 2018 році (Devlin et al, 2019). Основною відмінністю нейронної мережі *BERT* є те, що вона використовує архітек-

туру перетворювачів і навчається у двох напрямках. Це забезпечує можливість урахування смислового значення слів у реченні зі словами, що розміщені ліворуч і праворуч від них.

Для дослідження можливості отримання емоційної оцінки тексту відгуків використовувалася функція `pipeline('text-classification', model='bert-base-uncased-emotion')`, що міститься в бібліотеці Hugging Face Transformers; її розробила мова Python. Ця функція дає змогу отримати доступ до вже навченої моделі BERT Base Uncased Emotion, яка призначена для розпізнавання емоцій у англомовному тексті. Модель здатна розпізнавати шість емоцій: радість (Joy), здивування (Surprise), сум (Sadness), страх (Fear), любов (Love) і злість (Anger). Приклад використання цієї моделі для аналізу тексту відгуку подано на рис. 2.

```
[7] prediction = classifier("This book pleasantly surprised me and made me smile sincerely. The story is heartwarming, cheerful, and truly delightful.")

for emotion in prediction[0]:
    label = emotion['label']
    score = emotion['score']
    print(f'Emotion: {label.capitalize()}, Score: {score:.4f}')

Emotion: Sadness, Score: 0.0002
Emotion: Joy, Score: 0.9965
Emotion: Love, Score: 0.0016
Emotion: Anger, Score: 0.0002
Emotion: Fear, Score: 0.0002
Emotion: Surprise, Score: 0.0013
```

Рис. 2. Приклад звертання до моделі BERT Base Uncased Emotion

На рис. 2 продемонстрований процес звертання до моделі та отримані числові дані аналізу за шістьма емоціями: радість (Joy) – 0,9965; здивування (Surprise) – 0,0013; сум (Sadness) – 0,0002; страх (Fear) – 0,0002; любов (Love) – 0,0016; злість (Anger) – 0,0002. Сума числових даних дорівнює одиниці ($0,9965 + 0,0013 + 0,0002 + 0,0002 + 0,0016 + 0,0002 = 1$).

Метод визначення емоцій у тексті з використанням трансформера-бота *ChatGPT*. Для оцінювання емоцій, що містяться в текстових відгуках на книгу, можна використовувати аналізатор настроїв ChatGPT (Ran, 2023). Трансформер ChatGPT використовує свої знання про мову та контекст, щоб розпізнавати емоційне забарвлення на основі лексики та синтаксису тексту, визначаючи такі емоції, як радість (Joy), здивування (Surprise), сум (Sadness), страх (Fear), любов (Love) і злість (Anger).

Реалізація діалогу з ChatGPT в бібліотечній системі з функцією автоматизованого аналізу емоцій у реальному часі здійснюється за допомогою API ChatGPT. Для цього потрібно імпортувати бібліотеку OpenAI до середовища розробки Python і отримати ключ реєстрації на сайті OpenAI (параметр `openai.api_key`). Вихідними параметрами функції API ChatGPT є текст установки (параметр `prompt`), яка використовується для побудови аналізатора настроїв, а також текст відгуку на книгу (параметр `input_text`). Вхідним параметром функції є набір даних емоційного аналізу, який можна отримати за допомогою функції `analyze_emotions(input_text)`.

Розглянемо процес визначення емоцій за прикладом текстового відгуку на книгу. Установка (`prompt`), яка використовується для побудови аналізатора настроїв, така:

– англійською мовою: «*Analyse the following text and return an emotion score as percentages in decimal format. Identify 6 emotions: joy, fear, surprise, sadness, anger, love. Return the response without explanations, in JSON format: {"joy": ..., "fear": ..., "surprise": ..., "sadness": ..., "anger": ..., "love": ... };*»;

– українською мовою: «Проаналізуй наступний текст і поверни оцінку емоцій у вигляді відсотків, у десяткових дробах. Визнач 6 емоцій: радість (joy), страх (fear), здивування (surprise), сум (sadness), злість (anger), любов (love). Поверни відповідь без пояснень, у форматі JSON: { "joy": ..., "fear": ..., "surprise": ..., "sadness": ..., "anger": ..., "love": ... }».

На рис. 3 продемонстрований діалог з трансформером ChatGPT та отримані числові дані аналізу за шістьма емоціями: радість (Joy) – 0,8; здивування (Surprise) – 0,1; сум (Sadness) – 0; страх (Fear) – 0; любов (Love) – 0,1; злість (Anger) – 0. Сума числових даних дорівнює одиниці ($0,8 + 0,1 + 0,1 = 1$).

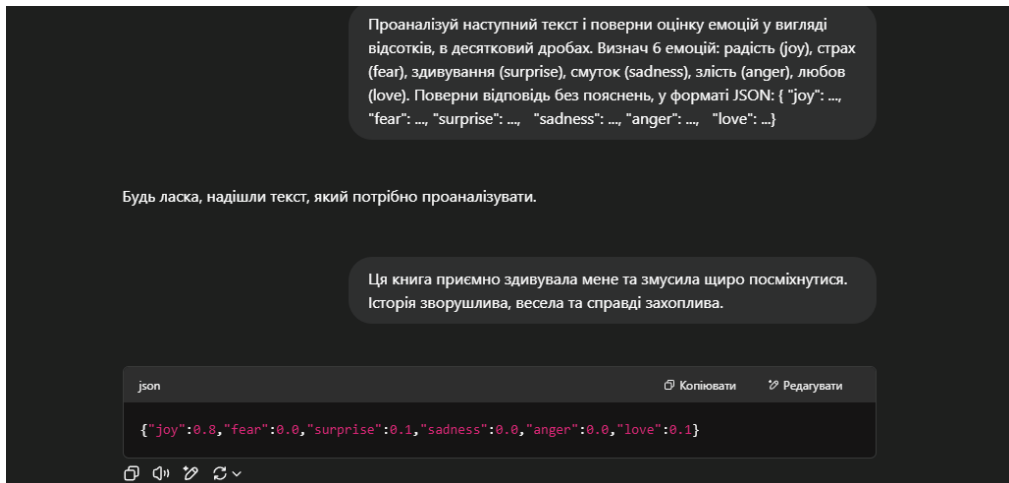


Рис. 3. Приклад діалогу з ChatGPT для визначення емоцій у тексті

Порівняння методів визначення емоцій у тексті. Дані порівняння розглянутих прикладів використання методів визначення емоцій у тексті подано в табл. 2.

Таблиця 2

Порівняння методів визначення емоцій у тексті

ВИДИ ЕМОЦІЙ	ЛЕКСИКОН NRC	МОДЕЛЬ BERT	ТРАНСФОРМЕР ChatGPT
Радість (Joy)	0,83	0,9965	0,8
Здивування (Surprise)	0,17	0,0013	0,1
Сум (Sadness)	0	0,0002	0
Страх (Fear)	0	0,0002	0
Любов (Love)	не підтримується	0,0016	0
Злість (Anger)	не підтримується	0,0002	0

За аналізом табл. 2 можна зробити висновки, що дані оцінок розглянутих методів приблизно збігаються. Найкращим є метод, що використовує модель BERT, але його недолік у тому, що він підтримує аналіз емоцій, які містяться тільки в англійських текстах. Якщо розглядати обов'язкову підтримку аналізу українськомовних текстів, то найкращим є метод, що використовує трансформер ChatGPT, бо він надає оцінку за шістьма видами емоцій.

База даних бібліотечної системи електронної комерції. Для дослідження методів визначення рекомендацій використовувалася база даних, яку наведено на рис. 4.

Призначення таблиць дослідницької бази даних:

- таблиця User (користувач) – облік персональних даних клієнтів;
- таблиця Book – облік інформації про книги;
- таблиця Review – облік відгуків користувачів – це текст відгуку (review_text); рейтингова оцінка (rating), яку виставляє книзі користувач за шкалою {1, 2, 3, 4, 5};
- таблиці Emotion і Emotions призначені для обліку оцінок емоцій, що містяться в текстовому відгуку на книгу. У таблиці Emotion (рис. 5) зберігається 4 записи з найменуваннями емоцій (поле name_em), якщо використовується метод визначення емоцій за лексиконом, або 6 записів, якщо використовується модель BERT і трансформер ChatGPT. Для зберігання числових оцінок для кожної емоції використовується поле valuation таблиці Emotions.

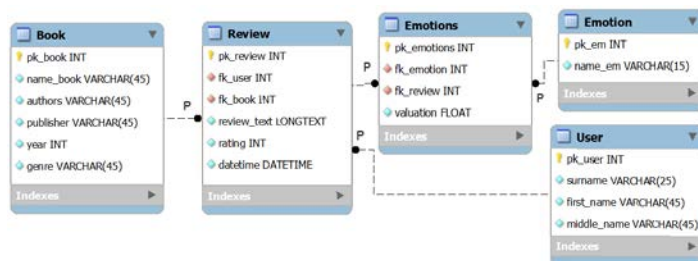


Рис. 4. ER-діаграма дослідницької бази даних

pk em	name em
1	joy
2	surprise
3	sadness
4	fear
5	love
6	anger

Рис. 5. Таблиця Emotion

Модифікація методу сумісної фільтрації для врахування емоцій читачів. Для реалізації спільної фільтрації на основі користувачів (Collaborative Filtering User Based) використовують рейтингову таблицю, що містить тільки оцінки книг (поле rating таблиці Review), які виставляють користувачі системи за власними уподобаннями після читання (Adomavicius and Tuzhilin, 2005). Для врахування емоцій потрібно модифікувати рейтингову таблицю (табл. 3), додаючи до неї не тільки рейтингові оцінки, а й оцінки емоцій. Інформація, що додається до табл. 3, отримується за запитом на вибірку з таблиць бази даних (рис. 4): User (поле pk_user, що позначене як індекс клієнта i); Book (поле pk_book, що позначене як індекс книги j); Review (поле rating – рейтингові оцінки, позначені як r_{ij}); Emotions (поле valuation – оцінки емоцій, що позначені як e_{ij}^w , де індекс w визначається ключем fk_emotion, $w = \{1, 2, 3, 4\} = \{\text{joy, surprise, sadness, fear}\}$, кількість оцінок емоцій $S = 4$).

Таблиця 3

Модифікована рейтингова таблиця оцінок книг від читачів

$j \rightarrow$	1					2					3					4					...	N				
$i \downarrow$	r_{i1}	e_{i1}^1	e_{i1}^2	e_{i1}^3	e_{i1}^4	r_{i2}	e_{i2}^1	e_{i2}^2	e_{i2}^3	e_{i2}^4	r_{i3}	e_{i3}^1	e_{i3}^2	e_{i3}^3	e_{i3}^4	r_{i4}	e_{i4}^1	e_{i4}^2	e_{i4}^3	e_{i4}^4	...	r_{iN}	e_{iN}^1	e_{iN}^2	e_{iN}^3	e_{iN}^4
1	5	0,2	0,7	0	0,1	1	0,2	0,5	0,1	0,2	4	0,5	0,2	0	0,3	4	0,1	0	0,3	0,6	...	3	0	0,3	0,4	0,3
2	4	0,4	0	0	0,6	?	?	?	?	?	2	0,1	0,1	0,8	0	5	0,5	0,2	0	0,3	...	3	0,1	0,3	0	0,6
3	4	0	0,1	0	0,9	2	0,1	0,3	0	0,6	3	0,7	0,1	0	0,2	2	0	0	0,4	0,6	5	5	0,2	0,3	0,5	0
4	3	0,9	0,1	0	0	5	0,2	0,7	0	0,1	1	0	0,4	0,3	0,3	1	0,9	0,1	0	0	4	4	0	0	0,7	0,3
5	1	0,4	0,2	0	0,4	5	0,5	0	0,5	0	5	0,2	0,4	0,3	0,1	2	0,5	0,1	0,4	0	...	1	0,4	0,2	0,2	0,2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
a	5	0,3	0,5	0,1	0,1	3	0,7	0,2	0,1	0	5	0,1	0,5	0	0,4	5	0,2	0	0,6	0,2	...	?	?	?	?	?
M	2	0,9	0,1	0	0	2	0,4	0,3	0	0,3	5	0	0	0,6	0,4	?	?	?	?	?	...	5	0,2	0,8	0	0

Головне завдання спільної фільтрації на основі користувачів – визначити прогнозу оцінку для книг, які користувач ще не оцінив. У табл. 3 – це користувач з індексом $i = a$ та неоцінена книга з індексом $j = N$, оцінка r_{aN} якої позначена символом «?» (клієнт та книга виділені зеленим фоном). Для прогнозу використовуються дані оцінок інших клієнтів, які за уподобаннями (за оцінками) найбільш близькі до нього. Метод спільної фільтрації реалізується за чотири кроки.

Крок 1. Підготовка таблиці даних для аналізу. До цієї таблиці відбираються дані клієнтів з рейтингової таблиці (табл. 3) за умови, що в них є оцінки книжок, які є i в клієнта з індексом $i = a$. Друга умова – у них є оцінки книг, які клієнт з індексом $i = a$ ще не оцінив. Оцінки емоцій клієнтів для книг, які не оцінив клієнт з індексом $i = a$, для розрахунку прогнозу не використовуються, а символ «?» замінюється нулем. Спрощений приклад даних для аналізу подано в табл. 4. Вона містить: N книг ($N = 5$) з рейтинговими оцінками r_{ij} , та Z книг ($Z = 4$) з оцінками емоцій e_{ij}^w . Усього враховується S оцінок емоцій ($S = 4$) для M відібраних клієнтів ($M = 4$) з індексами $P = \{1, 3, 4, 5\}$, індекс $i = a$ клієнта, для якого розраховується прогноз, розглядається окремо.

Таблиця 4

Дані для аналізу

$j \rightarrow$	1					2					3					4					5
$i \downarrow$	r_{i1}	e_{i1}^1	e_{i1}^2	e_{i1}^3	e_{i1}^4	r_{i2}	e_{i2}^1	e_{i2}^2	e_{i2}^3	e_{i2}^4	r_{i3}	e_{i3}^1	e_{i3}^2	e_{i3}^3	e_{i3}^4	r_{i4}	e_{i4}^1	e_{i4}^2	e_{i4}^3	e_{i4}^4	r_{i5}
1	5	0,2	0,7	0	0,1	1	0,2	0,5	0,1	0,2	4	0,5	0,2	0	0,3	4	0,1	0	0,3	0,6	3
3	4	0	0,1	0	0,9	2	0,1	0,3	0	0,6	3	0,7	0,1	0	0,2	2	0	0	0,4	0,6	5
4	3	0,9	0,1	0	0	5	0,2	0,7	0	0,1	1	0	0,4	0,3	0,3	1	0,9	0,1	0	0	4
5	1	0,4	0,2	0	0,4	5	0,5	0	0,5	0	5	0,2	0,4	0,3	0,1	2	0,5	0,1	0,4	0	1
a	5	0,3	0,5	0,1	0,1	3	0,7	0,2	0,1	0	5	0,1	0,5	0	0,4	5	0,2	0	0,6	0,2	0

Крок 2. Визначається міра збігу між рейтинговими оцінками й оцінками емоцій користувачів. Міра збігу визначається за розрахунками усередненої нормованої косинусної подібності рейтингових та емоційних оцінок:

$$\text{Sim}(a, i) = \left(R\text{Sim}(a, i) + \sum_{j=1}^Z E\text{Sim}(a, i, j) \right) / N, \quad (1)$$

де міра збігу N рейтингових оцінок книг, що виставили клієнти з індексами $i \in P$, визначається як:

$$R\text{Sim}(a, i) = \frac{\sum_{j=1}^N r_{aj} \times r_{ij}}{\sqrt{\sum_{j=1}^N r_{aj}^2} \times \sqrt{\sum_{j=1}^N r_{ij}^2}}, \quad i \in P, \quad (2)$$

а міра збігу S емоційних оцінок для Z книг – за формулою:

$$E\text{Sim}(a, i, j) = \frac{\sum_{w=1}^S e_{aj}^w \times e_{ij}^w}{\sqrt{\sum_{w=1}^S (e_{aj}^w)^2} \times \sqrt{\sum_{w=1}^S (e_{ij}^w)^2}}, \quad i \in P. \quad (3)$$

Крок 3. За значеннями $\text{Sim}(a, i)$ проводиться відбір клієнтів, які найбільш близькі за оцінками до клієнта з індексом $i = a$. Отриманий список клієнтів сортується за спаданням значень $\text{Sim}(a, i)$ і обмежується до K елементів. Фіксується множина U , що містить індекси K відібраних клієнтів.

Приклад розрахунку мір збігу згідно з даними табл. 4 подано в табл. 5. Якби оцінки емоцій не враховувалися, то відбір клієнтів здійснювався на основі значень $R\text{Sim}(a, i)$, для $K = 2$ – це клієнти з індексами $U = \{1, 3\}$. З урахуванням емоцій, відповідно до значень $\text{Sim}(a, i)$, це клієнти з індексами $U = \{1, 5\}$.

Таблиця 5

Приклад розрахунку мір збігу

$j \rightarrow$	1	2	3	4	5	$R\text{Sim}(a, i)$	1	2	3	4	Сума $\text{sim}()$	$\text{Sim}(a, i)$
$i \downarrow$	r_{i1}	r_{i2}	r_{i3}	r_{i4}	r_{i5}		$E\text{Sim}(a, i, 1)$	$E\text{Sim}(a, i, 2)$	$E\text{Sim}(a, i, 3)$	$E\text{Sim}(a, i, 4)$		
1	5	1	4	4	3	0,97	0,95	0,58	0,68	0,71	3,89	0,778
3	4	2	3	2	5	0,97	0,26	0,26	0,42	0,75	2,66	0,532
4	3	5	1	1	4	0,73	0,59	0,52	0,85	0,30	2,99	0,598
5	1	5	5	2	1	0,81	0,72	0,77	0,73	0,79	3,82	0,764
a	5	3	5	5	0							

Крок 4. Для обраного користувача ($i = a$) обчислюються прогнози – зважені середні рейтингові оцінки книг з індексами j на основі оцінок клієнтів, відібраних на кроці 3.

$$p(a, j) = \frac{\sum_{i \in U} r_{ij} \times R\text{Sim}(i, j)}{\sum_{i \in U} |R\text{Sim}(i, j)|}, \quad i \in U. \quad (4)$$

Використовуючи табл. 5, розглянемо приклад розрахунку $p(a, j)$ для $j = 5$:

$$p(a, j) = r_{a5} = 1 / (0,97 + 0,81) * (3 * 0,97 + 1 * 0,81) = 2,09.$$

Під час реальної роботи системи прогноз розраховується не для однієї, а для кількох десятків книг. Користувачеві подається рекомендований до читання або покупки список книг, який сортується за спаданням значень прогнозних оцінок $p(a, j)$. Кількість книг у списку також може бути обмежена заданим порогом. Недоліком запропонованого методу можна вважати обмеження, яке обумовлено тим, що не всі користувачі надають відгуки на прочитані книги.

Висновки. У цьому дослідженні розглянуто три підходи до емоційного аналізу текстових відгуків читачів – за допомогою лексиконів (структурованих словників), моделі глибокого навчання BERT, а також генеративного трансформера ChatGPT. З урахуванням розглянутих підходів розроблено модифікований рекомендаційний метод спільної фільтрації на основі користувачів (Collaborative Filtering User Based). Метод дає змогу під час спільної фільтрації даних визначати групи користувачів за найбільшим збігом рейтингових оцінок книг і оцінок емоцій текстових відгуків користувачів на книги. Розроблений метод може використовуватися в бібліотечних рекомендаційних системах, що надають послуги онлайн-читання книг або прослуховування аудіокниг.

REFERENCES

- Adomavicius, G. and Tuzhilin, A., 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, [e-journal] 17 (6), pp.734-749. <https://doi.org/10.1109/TKDE.2005.99>
- Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019*, Minneapolis, USA, June 2-7, 2019. Minneapolis: Association for Computational Linguistics, 2019, Vol. 1, pp.4171-4186.
- Mohsen, A.M., Hassan, H.A. and Idrees, A.M., 2016. Documents Emotions Classification Model Based on TF-IDF Weighting Measure. *International Journal of Computer and Information Engineering*, [online] 10 (1), pp.252-258. Available at: <<https://scispace.com/pdf/documents-emotions-classification-model-based-on-tf-idf-46yikynxgl.pdf>> [Accessed 6 March 2025].
- Ran, C., 2023. Emotion analysis of dialogue text based on ChatGPT: a research study. In: *International Conference on Algorithms, High Performance Computing, and Artificial Intelligence (AHPCAI 2023)*, Proceedings, 18-19 August 2023, [e-journal] Vol. 12941, pp.12941-37. <https://doi.org/10.1117/12.3011507>

UDC 004.8:[005.953.2:002]:[159.942:028.2-043.4

Adrian Druhak,
Master's Student,
Department of System Engineering,
Kharkiv National University of Radio Electronics,
Kharkiv, Ukraine
adrian.druhak@nure.ua
<https://orcid.org/0009-0009-1303-0608>

Andrii Kovalenko,
PhD in Technical Sciences,
Associate Professor at the System Engineering Department,
Kharkiv National University of Radio Electronics,
Kharkiv, Ukraine
andrey.kovalenko@nure.ua
<https://orcid.org/0000-0003-2882-5082>

Olena Chaikovska,
PhD in Pedagogy, Professor,
Kyiv National University of Culture and Arts,
Kyiv, Ukraine
oachaikovska@knukim.edu.ua
<https://orcid.org/0000-0001-7769-1004>

BOOK RECOMMENDATION WITH CONSIDERATION OF EMOTIONS EXTRACTED FROM READER TEXT REVIEWS

The purpose of the article is to study methods for obtaining emotional analysis data from readers' text reviews for use in the recommendation functions of e-commerce systems that provide online book reading or audiobook listening services.

The research methodology consists of systems approach, systems analysis and synthesis methods, and structural modelling of relational databases.

The scientific novelty lies in developing a modified method for determining user-based collaborative filtering recommendations, which considers data from the emotional analysis of text reviews of books read.

Conclusions. This article considers three approaches to evaluating the emotions of readers' text reviews. The first approach uses lexicon-based dictionaries, in which sets of emotions index words. Text emotion scores are calculated based on the statistics of the occurrence of words in sets of emotions in the lexicon. The second approach uses the BERT language model, allowing you to obtain emotional scores upon request. The BERT language model is a neural network that provides the ability to take into account the semantic meaning of words in a sentence, including the words to the left and right of them. The third approach uses the ChatGPT transformer bot. The evaluation of emotions in the reviews text can be done in real time using the ChatGPT API function. The output parameters of the API function are the installation text for the ChatGPT sentiment analyser and the review text, and the input parameter is emotion score data in JSON format.

Taking into account the considered approaches, a modified user-based collaborative filtering recommendation method has been developed, which takes emotional analysis data into account. For this purpose, it is proposed that the average cosine similarity of rating assessments and the emotional evaluations of users' textual reviews be considered during the data selection stage. The proposed method of determining recommendations can be used in library systems with an implemented user function of providing reviews of read books.

Keywords: emotion analysis; recommender systems; collaborative filtering; books; rating evaluation methods; extensions for recommender systems.

Надійшла 14.02.2025

Прийнята 16.05.2025

Стаття була вперше опублікована онлайн 18.07.2025



This is an open access journal, and all published articles are licensed under a Creative Commons Attribution 4.0.