

Вінницький національний технічний університет
Факультет комп'ютерних систем і автоматики
Кафедра системного аналізу та інформаційних технологій

Інформаційна технологія аналізу та рекомендації кінофільмів для глядачів

Виконав: студент 2 курсу, групи 2ІСТ-19м
спеціальності 126 – «Інформаційні системи та технології»
Корпало А.В.

Керівник: доцент, к.т.н. Козачко О.М.

АКТУАЛЬНІСТЬ

Рекомендаційні системи використовуються для прогнозування рейтингу чи переваг, які користувач надав би елементу.

Практично кожна велика технологічна компанія у світі застосовує їх у тій чи іншій формі:

- Amazon використовує її для пропонування продуктів покупцям;
- YouTube використовує її, щоб вирішити, яке відео відтворювати далі під час автовідтворення;
- Facebook використовує його, щоб рекомендувати сторінки за якими людям захочеться слідкувати.

Більше того, успіх бізнесу таких компаній, як Netflix та Spotify, сильно залежить від ефективності їхніх рекомендаційних механізмів.

Існуючі рішення не завжди видають точний і персональний для користувача результат, саме тому необхідні нові рішення, що базуються на сучасних інтелектуальних технологіях.

ОБ'ЄКТ, ПРЕДМЕТ ТА МЕТА ДОСЛІДЖЕННЯ

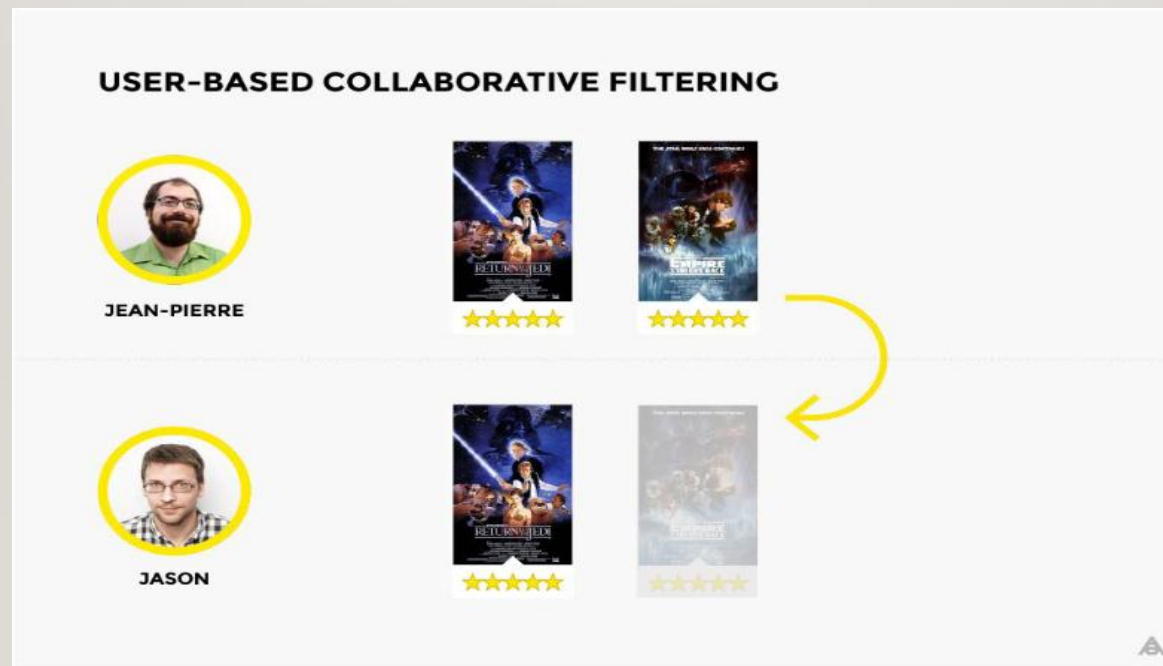
- **Метою** даної кваліфікаційної роботи є збільшення точності рекомендацій кінофільмів та прогнозованої оцінки для глядачів за рахунок використання інформаційних технологій.
- **Об'єктом дослідження** є процес створення інформаційної технології аналізу та рекомендації кінофільмів для глядача.
- **Предметом дослідження** є інформаційна технологія аналізу та рекомендації кінофільмів для глядача.

ПОСТАВЛЕНІ ЗАВДАННЯ ДОСЛІДЖЕННЯ

- ❑ оглянути існуючі стратегії створення рекомендаційних систем;
- ❑ проаналізувати способи застосування оглянутих стратегій для підвищення точності рекомендації кінофільмів;
- ❑ розробити інформаційну технологію аналізу та рекомендації кінофільмів для глядачів;
- ❑ програмно реалізувати інформаційну технологію аналізу та рекомендації кінофільмів для глядачів;
- ❑ порівняти точність передбачення розробленої технології з аналогами.

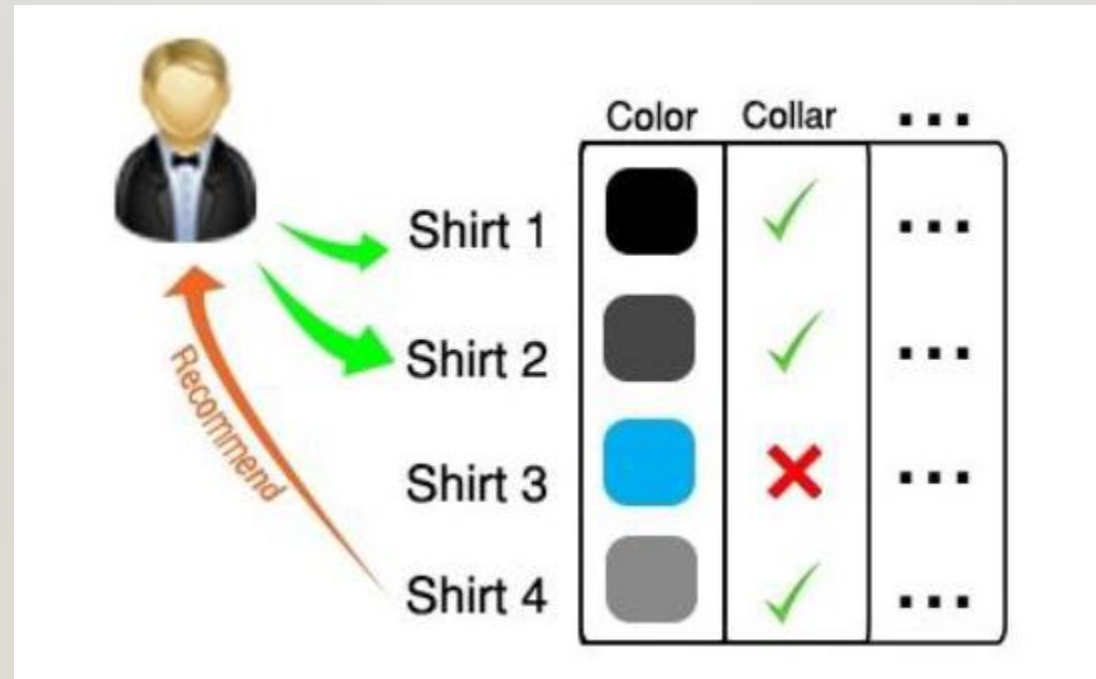
МЕТОД СПІЛЬНОЇ ФІЛЬТРАЦІЇ

- Головна ідея: «Хто однаково оцінював деякі об'єкти в минулому, дає таку ж оцінку на схожі об'єкти в майбутньому»



МЕТОД ФІЛЬТРАЦІЇ НА ОСНОВІ ВМІСТУ

- Методи фільтрації на основі вмісту базуються на описі елементів.



Алгоритм роботи модулю спільної фільтрації



$$S = 2 - \sqrt{\frac{\sum_i |y_i - x_i|}{m}}$$

де, S – схожість між двома користувачами,

i – номер фільму, $P = \sum_i \frac{S_i}{\sum S_i} * x_i$,

y – оцінка фільму користувача,
 де, P – рекомендована оцінка,
 i – номер фільму цього користувача,
 S – кількість фільмів
 m – кількість схожості користувача,
 x – оцінка фільму користувача

$$S = 2 - \sqrt{\frac{|5 - 4| + |2 - 4|}{2}} = 0,775$$

$$P = \left(\frac{1,455}{3,931} * 4\right) + \left(\frac{0,775}{3,931} * 2\right) + \left(\frac{1,701}{3,931} * 5\right) = 1,48 + 0,38 + 2,16 = 4,02$$

	1	2	3	Σ
S	1,455	0,775	1,701	3,931
x	4	2	5	

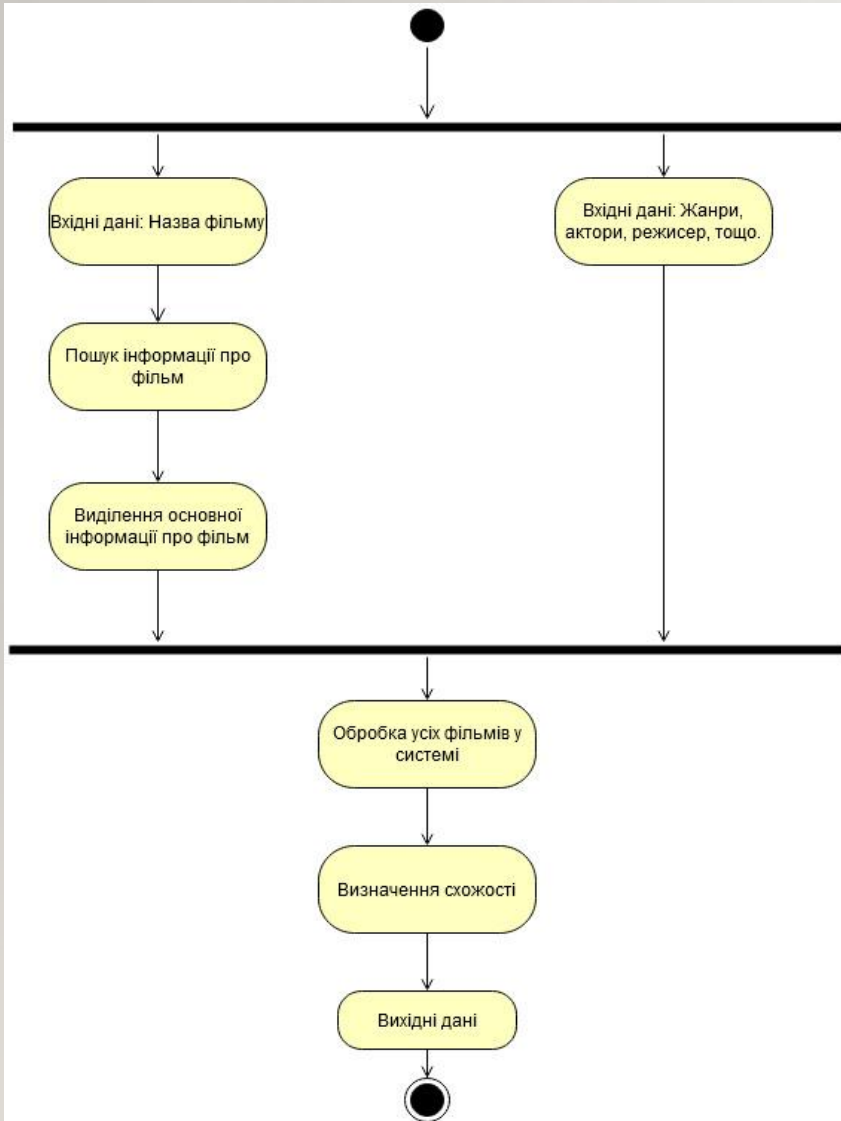
Алгоритм роботи модулю фільтрації на основі вмісту

Косинус подібності (англ. cosine similarity) — коефіцієнт подібності двох не нульових векторів у предгільбертовому просторі, який обчислюється як косинус кута між ними.

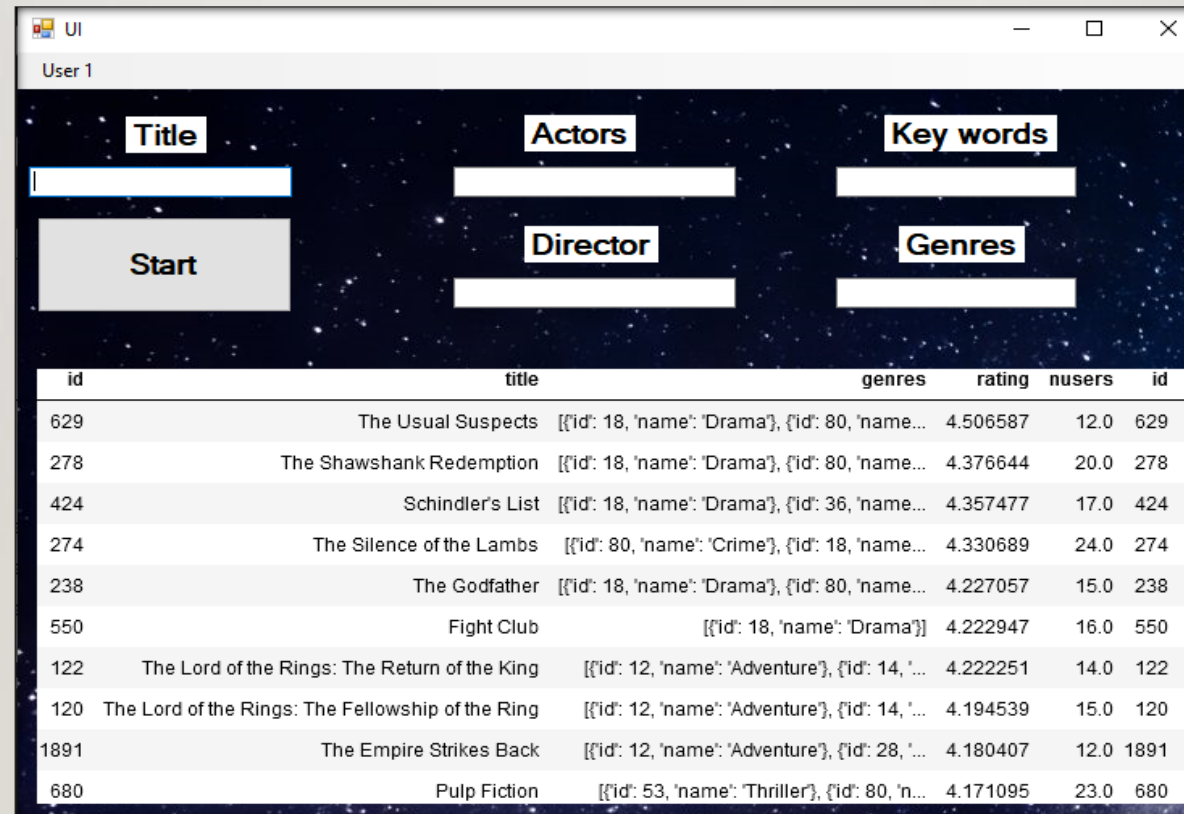
Для двох заданих векторів ознак, A та B , косинус подібності, $\cos(\theta)$, представляється за допомогою скалярного добутку та довжини, як:

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

де, A_i а B_i — координати вектору A та B відповідно



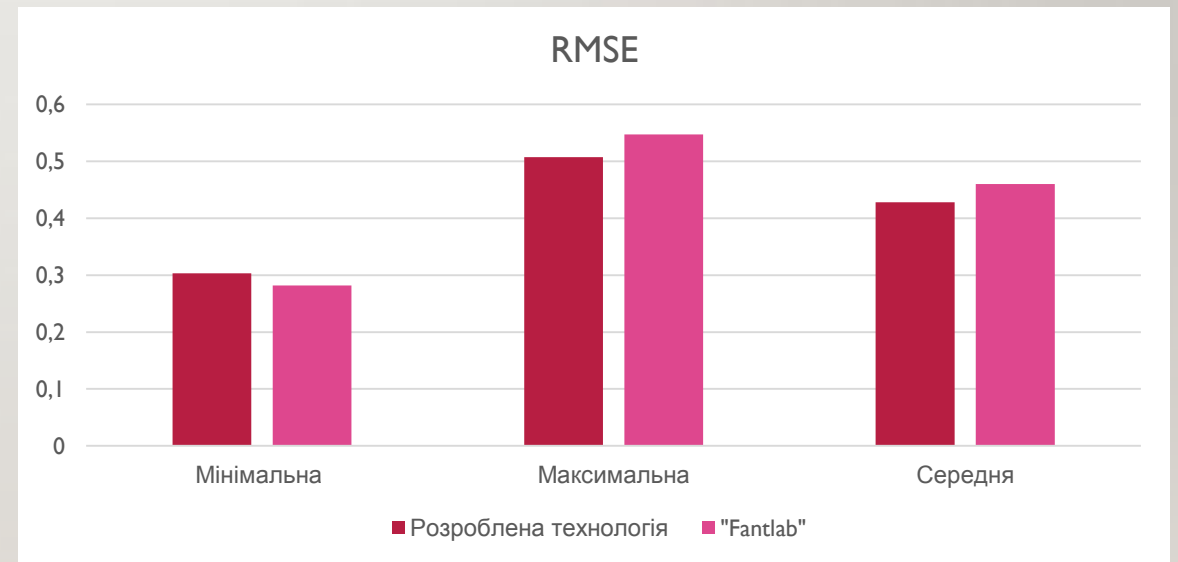
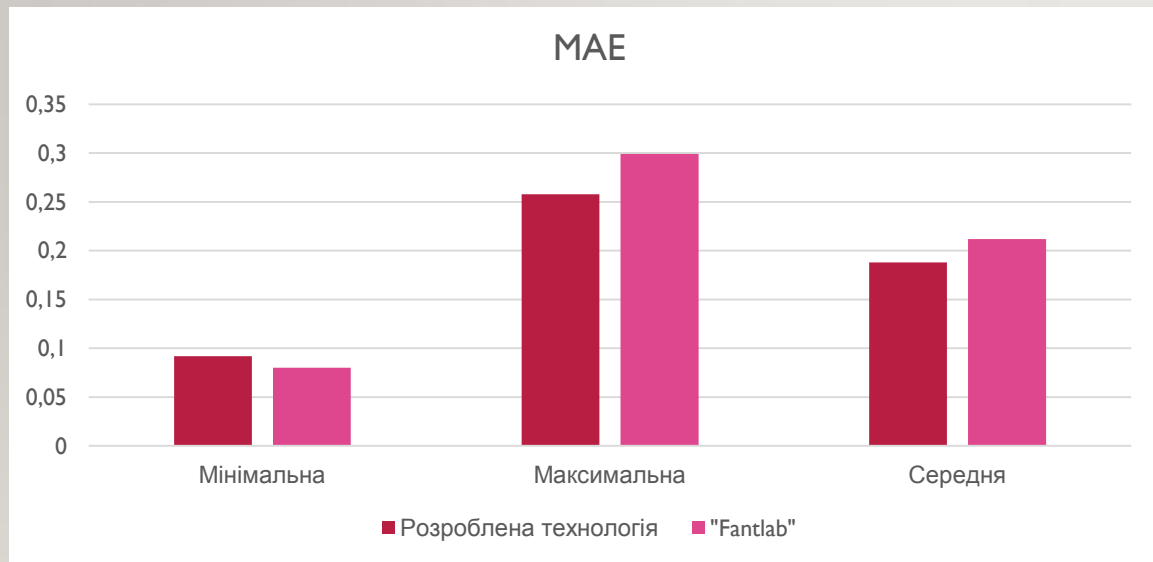
ПРИКЛАД РОЗРОБЛЕНОЇ ПРОГРАМИ



The screenshot shows a web application window titled "UI" with a "User 1" header. The main content area has a dark blue starry background. It features five input fields: "Title", "Actors", "Key words", "Director", and "Genres", each with a corresponding label above it. A "Start" button is positioned below the "Title" field. Below the input fields is a table displaying movie search results.

id	title	genres	rating	nusers	id
629	The Usual Suspects	[[{'id': 18, 'name': 'Drama'}, {'id': 80, 'name'...	4.506587	12.0	629
278	The Shawshank Redemption	[[{'id': 18, 'name': 'Drama'}, {'id': 80, 'name'...	4.376644	20.0	278
424	Schindler's List	[[{'id': 18, 'name': 'Drama'}, {'id': 36, 'name'...	4.357477	17.0	424
274	The Silence of the Lambs	[[{'id': 80, 'name': 'Crime'}, {'id': 18, 'name'...	4.330689	24.0	274
238	The Godfather	[[{'id': 18, 'name': 'Drama'}, {'id': 80, 'name'...	4.227057	15.0	238
550	Fight Club	[[{'id': 18, 'name': 'Drama'}]]	4.222947	16.0	550
122	The Lord of the Rings: The Return of the King	[[{'id': 12, 'name': 'Adventure'}, {'id': 14, 'n...	4.222251	14.0	122
120	The Lord of the Rings: The Fellowship of the Ring	[[{'id': 12, 'name': 'Adventure'}, {'id': 14, 'n...	4.194539	15.0	120
1891	The Empire Strikes Back	[[{'id': 12, 'name': 'Adventure'}, {'id': 28, 'n...	4.180407	12.0	1891
680	Pulp Fiction	[[{'id': 53, 'name': 'Thriller'}, {'id': 80, 'n...	4.171095	23.0	680

ТЕСТУВАННЯ ТОЧНОСТІ ПЕРЕДБАЧЕННЯ



Мінімальна – «Fantlab» на 15% точніше.

Максимальна – Розроблена технологія на 14% точніше.

Середня – Розроблена технологія на 12% точніше.

НАУКОВА НОВИЗНА

Подальшого розвитку набула інформаційна технологія аналізу та рекомендації кінофільмів для глядачів, заснована на удосконаленому методі гібридної фільтрації, що забезпечує підвищення точності рекомендацій та прогнозованої оцінки на 12%.

ПРАКТИЧНЕ ЗНАЧЕННЯ ТА ПУБЛІКАЦІЇ

Дослідження виконані в роботі дозволили розробити інформаційну технологію аналізу та рекомендації кінофільмів для глядачів, що надає рекомендації кінофільмів та прогнозує їх оцінку з точністю на 12% більше ніж в аналогів.

Окремі частини технології, з незначними змінами, можуть бути використані для рекомендації, будь яких, речей в рейтингових системах.

Тези за темою магістерської роботи були опубліковані у матеріалах XV міжнародної конференції "Контроль і управління в складних системах (КУСС-2020) ", м. Вінниця, 8-10 жовтня 2020 р.

ВИСНОВКИ

В процесі виконання магістерської кваліфікаційної роботи було розроблено та програмно реалізовано інтелектуальну технологію аналізу та рекомендації кінофільмів для глядачів.

Розглянуто основні методи побудови рекомендаційних систем. Обрано гібридну стратегію розробки, як таку яка найкраще підходить для досягнення поставлених цілей.

Розроблено загальну структуру інформаційної технології, яка складається з двох модулів, модулю спільної фільтрації та модулю фільтрації на основі вмісту, які працюють послідовно.

Програмно реалізовано розроблену технологію використовуючи мову програмування Python.

За результатами тестування було визначено, що розроблена технологія прогнозує оцінку з похибками $MAE = 0,188$, $RMSE = 0,428$, що приблизно на 12% точніше ніж в аналогів.

ДЯКУЮ ЗА УВАГУ!

