

Інформаційна технологія відбору відеоматеріалів. Програмні засоби на основі асоціативних правил

Виконав:
студент групи 2КН-16м
Гайдейчук Юрій Анатолійович

Керівник:
д.т.н., проф. Яровий А.А.

Мета і завдання досліджень, об'єкт, предмет

Метою магістерської кваліфікаційної роботи є підвищення ефективності процесу надання рекомендацій користувачеві шляхом застосування інтелектуального аналізу даних.

Для досягнення мети розробки необхідно виконати такі задачі:

- ❖ здійснити аналіз предметної області інтелектуального аналізу даних;
- ❖ здійснити аналіз характеристик методів інтелектуального аналізу даних для задачі аналізу контенту;
- ❖ здійснити розробку інформаційної технології, зокрема програмного засобу на основі асоціативних правил та комбінованого алгоритму;
- ❖ здійснити програмну реалізацію інформаційної технології відбору відеоматеріалів.

Об'єкт дослідження – процес надання рекомендацій щодо відбору відеоматеріалів.

Предмет дослідження – програмні засоби відбору відеоматеріалів на основі асоціативних правил

Актуальність теми дослідження

Проблема вибору відеоматеріалу залишається досить актуальною тому, що безліч людей витрачають значну кількість часу на пошук необхідного відеоматеріалу. Із збільшенням кількості інформації в мережі інтернет цей процес буде займати все більше часу і його якість буде невпинно погіршуватися.

Велике число веб-ресурсів дозволяють фільтрувати вміст фільмів за деякими параметрами, але швидко і якісно отримати потрібний результат вони не дозволяють. Їхнім недоліком є те, що відсутнє використання методів штучного інтелекту при наданні користувачеві рекомендації.

Тому дана проблеми досить актуальна і потребує дослідження і вирішення або покращення її стану.

Характеристики та аналіз аналогів

Найбільш відомі рекомендаційні системи: **Amazon**, **Youtube**, **Reddit**, **Last.fm**, **livelib**, **Criticker**.

Дані системи є комерційними, тому детальний аналіз структури і методів реалізації аналізу контенту не розголошується.



Підходи до вирішення задачі відбору відеоматеріалу

Існує багато підходів вирішення задачі відбору відеоматеріалів, найбільш поширені з них:

- ▶ **Кластерний аналіз**
- ▶ **СТТР**
- ▶ **ІАД**

Кластерний аналіз - це задача розбиття заданої вибірки об'єктів на підмножини, що називаються кластерами, так, щоб кожен кластер складався з схожих об'єктів, а об'єкти різних кластерів істотно відрізнялися. Завдання кластеризації відноситься до статистичної обробки

Система підтримки прийняття рішень - це комп'ютеризована система, яка шляхом збору та аналізу великої кількості інформації може впливати на процес прийняття рішень.

Data Mining - це процес виявлення у необроблених даних раніше невідомих нетривіальних, практично корисних і доступних інтерпретацій знань, необхідних для прийняття рішень у різних сферах діяльності.

В даний час все більшого значення набуває інтелектуальний аналіз даних.

Методи ІАД

- ❖ Асоціативні правила
- ❖ Деревя рішень
- ❖ Кореляційний аналіз
- ❖ Нечітка логіка
- ❖ Регресивний аналіз

Завдання знаходження асоціативних правил розбивається на дві підзадачі:

- ❖ Знаходження всіх наборів елементів, які задовольняють порогу `minsupport`. Такі набори елементів називаються такими, що часто зустрічаються.
- ❖ Генерація правил з наборів елементів, що знайдені згідно з достовірністю, яка задовольняє порогу `minconfidence`.

Алгоритм Apriori має просту структуру, завдяки якій в неї можна легко вносити модифікації для оптимізації алгоритму під конкретні набори даних для вирішення задачі пошуку асоціативних правил.

Алгоритм Apriori

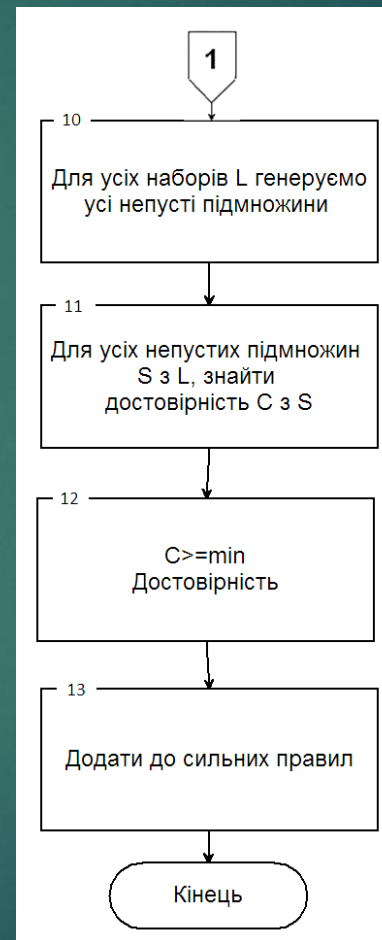
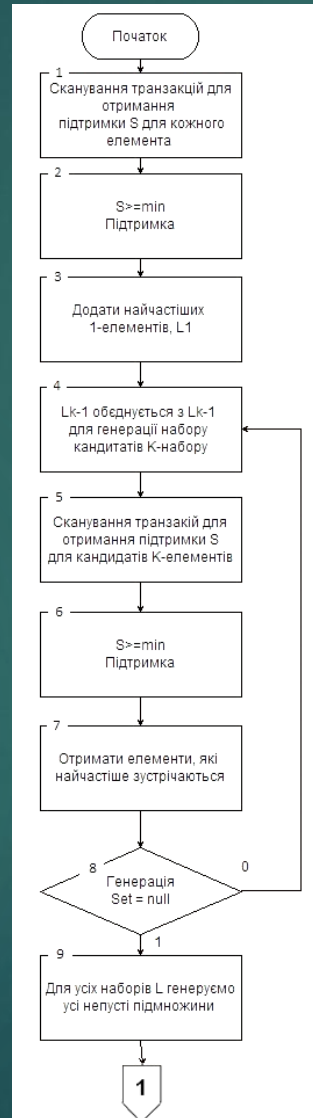
У роботі розглядається алгоритм Apriori. Він базується на методах інтелектуального аналізу (Data Mining), призначений для виявлення знань в базах даних.

Пошук асоціативних правил - ключова тема у інтелектуальному аналізі. В результаті можливий пошук приховані зв'язки в, на перший погляд, ніяк не пов'язаних даних. Ці зв'язки - асоціативні правила. Ті правила, кількість (підтримка) яких перевищує певний поріг, вважаються цікавими. Прикладом такого правила, є твердження, що в тому випадку, якщо відбулася подія A, то станеться і подія B з імовірністю X.

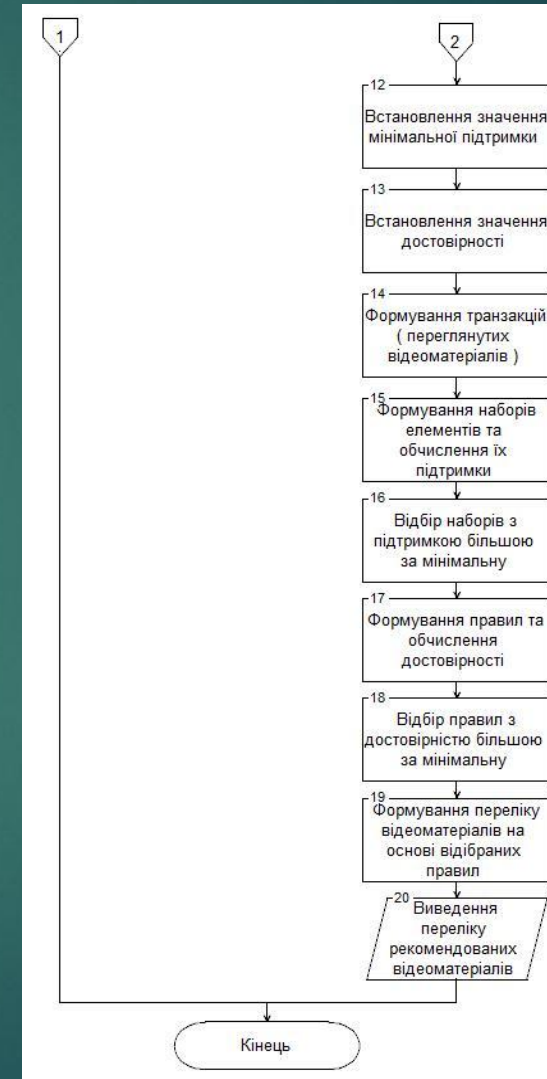
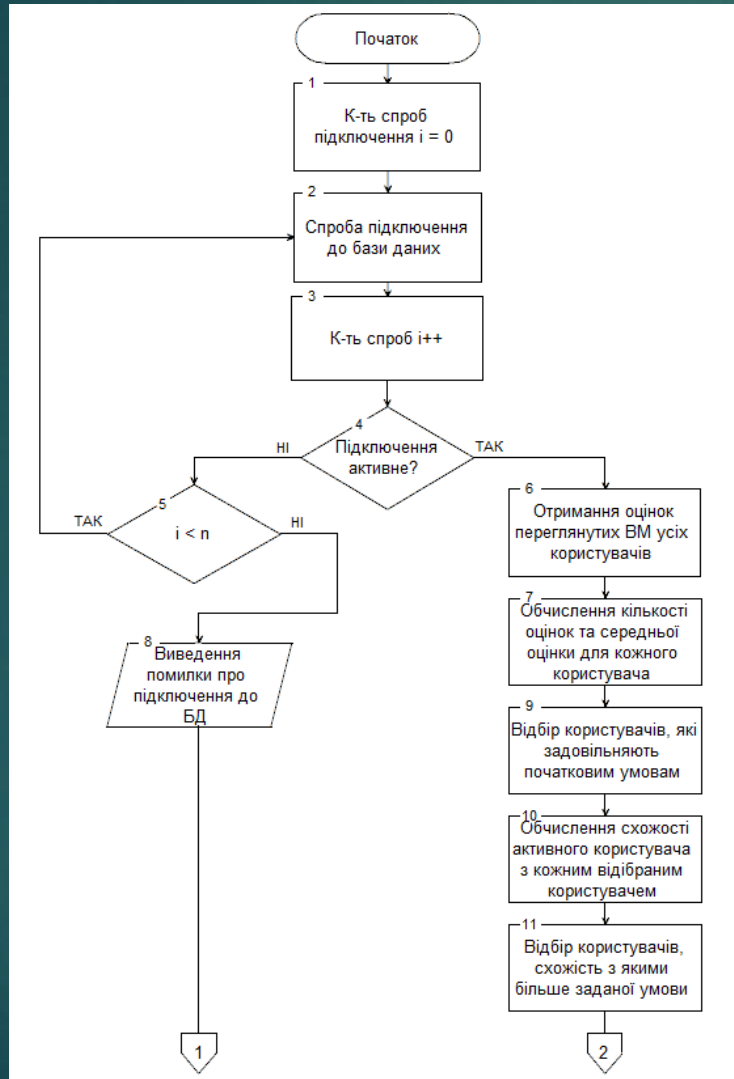
Одним з найбільш часто цитованих прикладів пошуку асоціативних правил служить проблема пошуку стійких зв'язків в кошику покупця (англ. Market Basket Problem). Проблема пошуку стійких зв'язків в кошику покупця полягає в тому, щоб визначити які товари купуються покупцями разом, так, щоб фахівці з маркетингу могли відповідним чином розмістити ці товари в магазині для підвищення обсягу продажів. Деякі виявляються правила можуть бути тривіальними, наприклад: «покупці, які купують пиво, також купують і горішки». Інші - цікаві і не тривіальні, наприклад: «покупці, які купують мило, так само купують і пиво». Саме здатність виявляти цікаві правила робить пошук асоціативних правил цінним і сприяє пошуку знань.

Також як приклад завдань пошуку асоціативних правил можна розглянути безлічі діагнозів, що зустрічаються одночасно у одного хворого, набори факторів ризику, що призводять до аварій на виробництві і т.д.

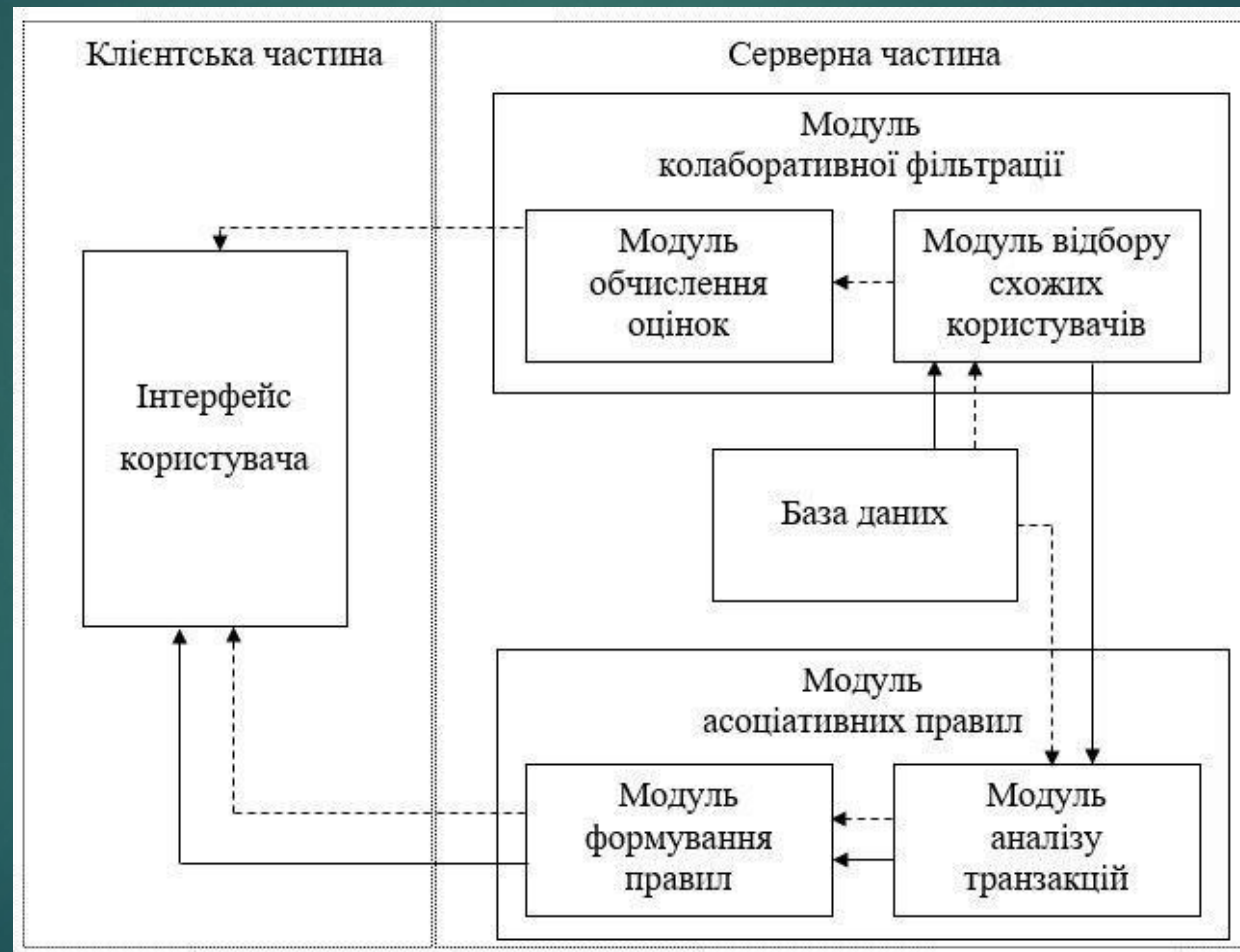
Низькорівнева схема алгоритму



Алгоритм роботи комбінованих алгоритмів



Загальна структурна схема інформаційної технології відбору відеоматеріалів



Обрані технології

- ❖ Для програмної реалізації системи обрано мови програмування JavaScript та TypeScript.
- ❖ Було здійснено порівняння, найбільш популярних REST API Node.js фреймворків, яке показало, що за переліком переваг найбільш доцільно використовувати Loopback.
- ❖ Обгрунтовано вибір технології реалізації програмних засобів клієнтської частини.
- ❖ Здійснено порівняльний аналіз трьох найбільш популярних клієнтських фреймворків.
- ❖ Для розробки клієнтської частини було обрано Angular 4.

Тестування

- ❖ Тестування проводилося на обліковому записі `jason_statham@gmail.com`, якому було присвоєно перегляд 100 фільмів. Вподобання та оцінки фільмів є суб'єктивними.
- ❖ Алгоритм розпочав своє функціонування після 10 заповнених облікових записів. До цього дані змінювалися, але це було не суттєво.
- ❖ Із збільшенням числа переглянутих фільмів рекомендації поступово змінювалися і починали відповідати вподобанням. Отже виявлено перший недолік системи, а саме високий поріг входу. Користуючись даним алгоритмом користувач не зможе зразу оцінити переваги системи.
- ❖ Після введення більше 100 фільмів дані алгоритми розпочали рекомендацію дійсно цікавих фільмів на погляд користувача.

Тестування швидкодії процесу надання рекомендацій

Здійснювалося на акаунті користувача `jason_statham@gmail.com`, який оцінив 100 фільмів. В таблиці 3.5 наведено схожість даного користувача до інших. Поріг схожості складає 0.3

На таблиці зображено схожість користувача `jason_statham` з іншими користувачами

№	UserID	Схожість
1	1	0.3316156231040968
2	2	0.36270665303380223
3	4	0.3541802926436001
4	5	0.35591542502953116
5	6	0.4003608339284706
6	7	0.3090738643394098
7	10	0.31469648083921886

Тестові умови:

- ❑ Колаборативна фільтрація рекомендує 8 відеоматеріалів. Асоціативні правила - 5. Комбінований алгоритм - 5. В даному випадку комбінований алгоритм та асоціативні правила генерують однакові фільми, що дозволяє достовірно оцінити час роботи кожного з алгоритмів та порівняти швидкодію.
- ❑ Процесор: Intel Core i3-6100 3.7GHz/8GT/s/3MB
- ❑ Материнська плата : MSI H110M PRO-VH
- ❑ ОЗУ : Kingston DDR4-2133 8192MB
- ❑ SSD: Kingston SSDNow V300 120GB
- ❑ Node.js : v8.8.1

№	Назва алгоритму	Час виконання (мс)
1	Apriori	6563
	Комбінований	400
2	Apriori	6707
	Комбінований	367
3	Apriori	6030
	Комбінований	346
4	Apriori	6384
	Комбінований	362
5	Apriori	6455
	Комбінований	350
6	Apriori	6432
	Комбінований	368
7	Apriori	7384
	Комбінований	365
8	Apriori	6114
	Комбінований	354
9	Apriori	6242
	Комбінований	432
10	Apriori	6570
	Комбінований	365

Середнє значення роботи алгоритму Apriori складає 6488.1 мс., а комбінованого – 370.9 мс., що в 17,493 рази швидше.

Результати роботи веб-ресурсу

MoviePark Movies Recommendations Hello : Jason Statham Log out

Collaborative filtering Apriori Combined algorithm

<p>MP: 8.9 IMDb: 8.7 Metacritic: 47</p> <p>The Matrix Revolutions</p>	<p>MP: 6.2 IMDb: 5.8 Metacritic: 14</p> <p>Night Club</p>	<p>MP: 7.9 IMDb: 6.8 Metacritic: 55</p> <p>Valerian and the City of a Thousand Planets</p>	<p>MP: 7 IMDb: 7.8 Metacritic: 79</p> <p>Spider-Man: Homecoming</p>	<p>MP: 8.8 IMDb: 7.8 Metacritic: 65</p> <p>Rogue One: A Star Wars Story</p>
<p>MP: 8.2 IMDb: 7.8 Metacritic: 83</p> <p>Avatar</p>	<p>MP: 7.2 IMDb: 7.8 Metacritic: 79</p> <p>The Incredibles 2</p>	<p>MP: 6.8 IMDb: 7.3 Metacritic: 59</p> <p>The Great Gatsby</p>	<p>MP: 7.3 IMDb: 6.7 Metacritic: 59</p> <p>Spectre</p>	<p>MP: 6.6 IMDb: 7.8 Metacritic: 83</p> <p>Mississauga</p>

MoviePark Movies Recommendations Hello : Jason Statham Log out

Rogue One

Premiere date: 15 December 2016
Slogan: Save the Rebellion. Save the dream.
Length: 133 min
Genres: Sci-Fi, Action, Adventure
Countries: USA
Budget: 200000000
Age limit: PG-13
Description: The daughter of an Imperial scientist joins the Rebel Alliance in a risky move to steal the Death Star plans.

7.8 IMDb 65 Metacritic 8.8 MP

Rogue One: A Star Wars Story Trailer (Official)

0:02 / 2:15 YouTube

MoviePark Movies Recommendations Hello : Jason Statham Log out

Collaborative filtering Apriori Combined algorithm

<p>MP: 6.8 IMDb: 8.3 Metacritic: 91</p> <p>Toy Story 3</p>	<p>MP: 7.8 IMDb: 8.2 Metacritic: 83</p> <p>Transporting</p>	<p>MP: 7 IMDb: 7.8 Metacritic: 79</p> <p>Spider-Man: Homecoming</p>	<p>MP: 8.8 IMDb: 7.8 Metacritic: 65</p> <p>Rogue One</p>	<p>MP: 7.3 IMDb: 6.7 Metacritic: 59</p> <p>Signs</p>
--	---	---	--	--

Економічна частина

В даному розділі було виконано оцінювання комерційного потенціалу розробки інформаційної технології відбору відеоматеріалів.

- ❖ Проведене порівня з сервісом Suggest Me Movie.
- ❖ Розраховано витрати на заробітну плату, амортизаційні відрахування, витрати на комплектуючі та на силову електроенергію, які становлять 90234,53 грн. Загальні витрати, які здійснюють усі виконавці даного етапу роботи складають 203741,5 грн.
- ❖ Виконано прогнозування комерційних ефектів від реалізації результатів розробки. Очікується, що основний прибуток буде отримуватись протягом трьох років. Чистий прибуток підприємства від впровадження розробки за три роки становить 533751,85 грн.
- ❖ Здійснено розрахунок ефективності вкладених інвестицій та періоду їх окупності. Життєвий цикл наукової розробки становить 3 роки. Абсолютна ефективність вкладених інвестицій дорівнює 330010,35 грн. Відносна щорічна ефективність вкладених коштів становить 38%, при мінімальному порозі в 19%. А термін окупності інвестицій складає 2,63 роки.

Тому фінансування розробки інформаційної технології відбору відеоматеріалів є доцільним для інвесторів і може принести їм потенційний прибуток.

Наукова новизна одержаних результатів.

- ❖ Підвищено швидкодію процесу формування асоціативних правил за рахунок впровадження попередньої фільтрації транзакцій.
- ❖ Запропоновано інформаційну технологію відбору відеоматеріалів на основі асоціативних правил та колаборативної фільтрації, що забезпечило підвищення швидкодії та якості надання рекомендації.

Практичне значення одержаних результатів

- ❖ На основі проведених досліджень розроблено програмне забезпечення відбору відеоматеріалів.
- ❖ Запропонована інформаційна технологія сприяє підвищенню ефективності процесу надання рекомендації, зокрема:
 - Розроблено алгоритм для підвищення ефективності надання рекомендацій на основі комбінованого застосування алгоритмів асоціативних правил та колаборативної фільтрації;
 - Реалізовано програмний засіб на сучасних технологіях розробки: Loorback 3.0, Fireloop, Angular4, що забезпечує підвищення надійності та швидкодії.
 - Розроблений програмний засіб для підвищення продуктивності надання рекомендацій, що дозволяє користувачам застосовувати комбіновані алгоритми асоціативних правил та колаборативної фільтрації, зберігати їхню історію переглядів і залишати оцінки відеоматеріалам.

Апробація, публікація, впровадження результатів

- ❖ Результати даних досліджень були апробовані на XLIV та XLV науково-технічній конференції Вінницького національного технічного університету (Вінниця, 2016 - 2017).
- ❖ Опубліковано 2 тез доповідей науково-технічних конференцій та подано заяву про реєстрацію авторського права на твір (комп'ютерну програму), реєстраційний номер АПС/2931-17 від 29.11.17р.
- ❖ Результати даного дослідження протестовано та плануються до подальшого використання у ФОП Носенко С.І., м. Київ (додаток Б).

Висновки:

- ❖ В ході виконання комплексної магістерської кваліфікаційної роботи розроблено інформаційну технологію відбору відеоматеріалів. При аналізі предметної області було з'ясовано, що найкращим підходом для розв'язання задачі відбору відеоматеріалів виявилися інтелектуальний аналіз даних та рекомендаційні системи.
- ❖ В другому розділі комплексної магістерської кваліфікаційної роботи було досліджено і здійснено класифікацію основних методів інтелектуального аналізу даних. Визначено сфери використання кожного з них. Найбільш актуальними методами для аналізу контенту є методи Data Mining.
- ❖ Було здійснено характеристику та аналіз методів інтелектуального аналізу даних для задачі аналізу контенту. Проаналізовано переваги та недоліки кожного з них. Обрано метод асоціативних правил для реалізації інтелектуальної системи відбору відеоматеріалів. Обґрунтовано доцільність комбінованого застосування асоціативних правил і методів рекомендаційних систем.
- ❖ Третій розділ присвячено проектуванню та розробці програмного забезпечення. Було розроблено: UML-діаграму класів, алгоритм роботи комбінованих алгоритмів на базі асоціативних правил та колаборативної фільтрації, схеми алгоритмів роботи внутрішніх частин програмного засобу.
- ❖ Здійснено програмну реалізацію інформаційної технології відбору відеоматеріалів.
- ❖ Проведено тестування програмного продукту, яке показало, що розроблена інформаційна технологія та програмний засіб відповідає поставленій меті. За допомогою комбінації асоціативних правил та колаборативної фільтрації вдалося підвищити швидкодію процесу надання рекомендацій у 17,493 раз.

ДЯКУЮ
ЗА
УВАГУ!