

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧОРНОМОРСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ПЕТРА МОГИЛИ

Волков Андрій Михайлович

УДК 004.04

Інформаційна система перегляду відеоконтенту

Галузь знань 12 «Інформаційні технології» за спеціальністю

122 «Комп'ютерні науки»

122 - БКР.А – 401з. 21930112

Автореферат

бакалаврської кваліфікаційної роботи на здобуття освітньої кваліфікації

«бакалавр з комп'ютерних наук»

Миколаїв – 2021

Бакалаврська кваліфікаційна робота є рукопис.

Робота виконана в Чорноморському національному університеті імені Петра Могили Міністерства освіти і науки України на кафедрі інтелектуальних інформаційних систем

Науковий керівник:

канд. пед. наук, доцент
Болюбаш Надія Миколаївна

Рецензент:

доктор техн. наук, професор
Коваленко Ігор Іванович

Захист відбудеться «24» червня 2021 р. о 9⁰⁰ год. на засіданні екзаменаційної комісії (ауд. 2-403) у Чорноморському національному університеті імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68-ми Десантників, 10.

З бакалаврською кваліфікаційною роботою можна ознайомитися в бібліотеці Чорноморського національного університету імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68-ми Десантників, 10.

Автореферат представлений «___» червня 2021 р.

Секретар
екзаменаційної комісії,
викл. каф. ПС

Таранов М.О.

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Інформатизація суспільства на сучасному етапі характеризується великими обсягами накопиченого відеоконтенту в мережі Інтернет. Набуває поширення технологія OTT (анг. Over the Top), яка передбачає надання відеопослуг через Інтернет. У цих умовах перспективним напрямком підвищення ефективності інформаційних систем перегляду відеоконтенту є застосування автоматизованих систем побудови рекомендацій для ідентифікації вподобань користувача за їх попередньою поведінкою.

Сервіс YouTube у свій час зробив революцію в можливостях перегляду відеоконтенту й є найпопулярнішою платформою для перегляду відео в режимі онлайн. Існує також багато сервісів для перегляду ліцензійного відеоконтенту на ПК, планшетах, мобільних пристроях та SmartTV. Ці сервіси надають послуги як безкоштовно, так і на комерційній основі. Однак системи перегляду відеоконтенту із вбудованими рекомендаційними системами з високою швидкістю і точністю розпізнавання вподобань користувача є комерційними та дорогими, що стримує їх широке використання. Значний потенціал для вирішення цієї проблеми надає розробка веб-додатку, який дозволяє переглядати відеоконтент і будувати рекомендації в реальному часі на основі вподобань, які було виявлено у користувача.

Це обумовило **мету дослідження**, яка полягає у поліпшенні перегляду відеоконтенту шляхом розробки і впровадження веб-додатку із вбудованою системою рекомендацій по перегляду.

Відповідно до поставленої мети було сформульовано **завдання дослідження**:

1. Здійснити аналіз розвитку сфери сервісів перегляду відеоконтенту та дослідити теоретичні засади створення рекомендаційних систем для перегляду контенту на основі уподобань користувачів.
2. Обґрунтувати вибір технологій та інструментальних засобів розробки інформаційної системи перегляду відеоконтенту.

3. Розробити та здійснити програмну реалізацію інформаційної системи для перегляду відеоконтенту із вбудованою системою рекомендацій та дослідити якість програмної реалізації.

Об'єкт дослідження – перегляд відеоконтенту в мережі Інтернет.

Предмет дослідження – веб-орієнтовані програмні засоби для перегляду відеоконтенту із використанням алгоритмів побудови рекомендацій.

Методологічною основою дослідження є загальнонаукові та статистично-аналітичні методи, які дозволили комплексно вивчити предмет та об'єкт дослідження, дослідити напрями та шляхи оптимізації перегляду відеоконтенту на основі уподобань користувача.

Практичне значення отриманих результатів полягає в тому, що використання розробленої інформаційної системи дозволить підвищити якість перегляду відеоконтенту у мережі Інтернет за рахунок врахування уподобань користувачів.

Структура дипломної роботи. Відповідно до мети, завдань і предмета дослідження, дипломна робота містить основну та спеціальну частини. Основна частина роботи складається із вступу, чотирьох розділів, висновку, списку використаних джерел та 7 додатків. Загальний обсяг роботи – 88 сторінок, із них основного тексту основної частини – 48 сторінок, спеціальної – 18 сторінок. Кількість використаних джерел – 33.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі подано загальну характеристику досліджуваної теми, обґрунтовано актуальність дипломної роботи, сформульовано мету та завдання досліджень, вказано методи та засоби дослідження.

У першому розділі було розглянуто теоретичні засади перегляду відеоконтенту на основі рекомендаційних алгоритмів, розвиток мережевих сервісів перегляду відеоконтенту та основи створення рекомендаційних систем.

У сучасному світі в умовах глобалізації та поширенні Інтернет технологій, більшість бізнесів намагаються цифрувати свої послуги та надавати їх глобально за допомогою інтернету. Ця ситуація поширилась і на надання послуг з перегляду фільмів: від кабельного телебачення і DVD-дисків до цифрового телебачення. Якщо раніше потрібно було придбавати диск з фільмом, що сподобався, то тепер достатньо лише зробити декілька кліків і фільм можна буде переглядати у високій якості. Поступовий сплеск споживання OTT послуг багато в чому можна пояснити наступними факторами:

1. Необмежений вміст потокового аудіо та відео.
2. Доступна вартість, гнучкі моделі ціноутворення, засновані на оплаті, вільній оплаті (обмежений доступ до безкоштовного вмісту) або підписці
3. Поставляється за допомогою широкого спектру пристроїв, таких як смартфони, планшети, настільні комп'ютери, смарт-телевізори, переносні пристрої та приставки.
4. Послуги OTT можна легко придбати та скасувати.
5. Доступ до оригінального відео вмісту, такого як веб-серіали та найновіші фільми.
6. Вміст на замовлення доступний, а не чекає виходу шоу чи фільму на телевізор.

Рекомендаційна система – це система, що будує рейтинг товарів користувача серед колосального набору даних, на підґрунті вже існуючої інформації про користувача. Товари – елементи, що надаються сервісом, в яких користувач може

бути зацікавлений, такі як: фільми, музика, побутові товари, книги та ін. Є декілька способів, щоб описати зацікавленість користувачів: можна скористуватись оцінками та відгуками, що користувач залишає до послуг, або (та) використавши ключові слова, характеристики, що має товар.

Щоб зробити прогноз щодо вподобань користувачів стосовно фільмів, РС мають збудувати портрет користувача. Портрет користувача у більшості РС являє собою модель користувача, з наборів оцінок та ключових слів, характеристик, тощо. Оцінки, що користувачі залишають фільмам, можуть належати вільним інтервалам (0-1, 1-5, 1-10): чим рейтинг вище, тим більше можна сказати, що цей фільм вподобався користувачу [3]. Після створення кожної оцінки, всі рейтинги користувача агрегуються за допомогою специфічних алгоритмів, вимірюється схожість користувачів, а потім створюються рекомендації для цього користувача. Ключові слова автоматично обираються з опису фільму або серіалу, які користувачі проглядали або оцінювали в минулому. Після цього фільми зіставляються з портретом користувача та ті, що відповідають йому найбільше – рекомендуються.

Оцінки товарів можна поділити на явні та не явні. Явна оцінка – це та оцінка, якою користувач показав зацікавленість даним фільмом в межах своєї оціночної системи. Неявні оцінки будуються на основі історії перегляду або попередньої поведінки користувачів. Однією з переваг неявного оцінювання є зменшення потреби в оцінюванні фільмів користувачем. Неявні оцінки можуть походити з, наприклад часу, що був витрачений на перегляд фільму статті, посилання на фільм з інших сайтів (наприклад, так працює алгоритм ранжування сторінок Google). Інші дані про поведінку користувача такі як: швидкість перегляду сторінки, рух курсору або ввід клавіатури також були розглянуті в якості неявних показників зацікавленості та показали відмінні результати.

Базову архітектуру рекомендаційної системи можна зобразити наступним чином:

1) *Довідкова інформація* – існуюча інформація про послуги або товари, що надаються сервісом.

2) *Вхідна інформація* – інформація, що користувач явно або неявно надає системі для отримання рекомендацій. Може бути оцінками, поведінкою або прямими запитами, наприклад пошук за жанром або назвою.

3) *Рекомендаційний алгоритм* – алгоритм, що комбінує зібрані довідкові дані та вхідні дані, отримані від користувача, для побудови рекомендацій.

Типові системи в якості довідкових даних найчастіше використовують портрети користувачів, а в якості вхідних – дії користувача (оцінки товарів, час, проведений за переглядом товару, тощо).

Переважає більшість рекомендаційних використовує один з двох ключових методів: контентна фільтрація або колаборативна фільтрація. Окрім цього є ціла група методів, що базуються на поєднанні двох базових – гібридна фільтрація.

Здійснений аналіз показав, що сучасний стан інформатизації суспільства супроводжується накопиченням відеоконтенту у мережі Інтернет. У таких умовах існує потреба у автоматизованих системах перегляду відеоконтенту, у яких є вбудовані рекомендаційні системи на основі уподобань користувача. Наявність таких систем сприяє підвищенню ефективності систем перегляду відеоконтенту. Виявлено, що сьогодні існує багато кросплатформених сервісів для перегляду ліцензійного відеоконтенту на комерційній основі та безкоштовно. Однак системи з високою швидкістю і точністю розпізнавання уподобань є дорогими, що обумовлює необхідність створення доступних веб-застосунків для перегляду відеоконтенту з вбудованою рекомендаційною системою.

У другому розділі було розглянуто технології та засоби розробки веб-застосунків та обґрунтування їх використання.

Серед інструментальних засобів розробки інформаційної системи перегляду відеоконтенту було обрано середовище розробки програмного забезпечення IDE IntelliJ IDEA, яке має підтримку інструментів для проведення тестування TestNG і JUnit, системи контролю версій CVS, Mercurial і Git та має вбудований плагін DataGrid для зручної роботи з базами даних. Для розробки серверної частини веб-додатку було обрано мову програмування JavaScript та платформу NodeJS, яка надає можливість виконувати JavaScript-скрипти на сервері та відправляти результат їх

виконання користувачеві. Для серіалізації даних було використано формат JSON. Для розробки графічного інтерфейсу було обрано варіант з використанням односторінкових додатків та Javascript-фреймворк React.

Javascript є важливою мовою програмування, оскільки він є мовою веб-браузерів. Його пов'язаність з браузером робить його однією з найпопулярніших мов програмування у світі. Водночас це одне з найбільш зневажаваних програмних мов у світі.

NodeJS, як платформа має багато переваг у порівнянні з іншими інструментами. Javascript – спільна мова на фронтенді та бекенді. Розробники розуміють один одного, що покращує якість та швидкість розробки.

Швидкість – є однією з основних привабливих особливостей NodeJS. Причиною подібного є неблокуюча архітектура платформи, а конкретні результати залежать від використаних тестів швидкості, але, в цілому, NodeJS - є дуже швидкою платформою.

Легкість прототипізації - побудувати мінімальний та життєздатний додаток, та завантажити його на веб-сервер дуже легко через мінімалістичність та простоту платформи.

Асинхронність - JavaScript надзвичайно полегшує створення неблокуючого коду та асинхронного коду завдяки використанню єдиного потоку, функцій зворотного виклику (коллбеків) і підходу до розробки, заснованого на подіях. Кожного разу, коли програмі потрібно виконати витратні розрахунки, як аргумент передається коллбек, що буде виконаний відразу після завершення цього розрахунку. В результаті, для продовження роботи програми, чекати на завершення виконання подібних розрахунків не потрібно.

Кількість доступних бібліотек – менеджер пакетів під назвою npm дозволяє дуже легко інсталювати нові бібліотеки. У 2019 році число доступних для завантаження пакетів перевищило 1 мільйон. Це дозволяє використовувати готові бібліотеки для вирішення типових задач в програмуванні. Також менеджер пакетів дозволяє додавати скрипти як наприклад: для запуску веб-додатку, або перевірки додатку юніт-тестами.

JSON – спосіб серіалізації даних, який дуже нагадує звичайний Javascript об'єкт. Можна легко модифікувати, не звертаючись до стороннього програмного забезпечення.

У третьому розділі було описано розробку, програмну реалізацію, тестування та впровадження інформаційної системи перегляду відеоконтенту з вбудованою рекомендаційною системою, який дозволяє створювати користувачам особисті сторінки з контактною інформацією, здійснювати пошук та переглядати відеоконтент у зручному для користувача режимі, формувати персональний список відео. У додатку передбачена можливість оцінки переглянутого відео користувачем, збір та аналіз зібраної інформації, демонстрація проаналізованих даних.

Було побудовано рекомендаційну систему з використанням колаборативної фільтрації. Для вимірювання подібності користувачів скористаємося коефіцієнтом Жаккара. Є багато способів знаходити подібність у вподобаннях двох користувачів. Одним із способів є коефіцієнт Пірсона який добре підходить для вимірювання дискретних значень у невеликих проміжках (наприклад оцінка від 1 до 5 зірок). Але щоб покращити швидкість розрахунків, система буде використовувати коефіцієнт Жаккара. Він дуже корисний для вимірювання бінарних рейтингових даних (наприклад лайк/дизлайк). Багато великих компаній, як наприклад Youtube, пішли цим шляхом, оскільки користувачі найчастіше залишають оцінки 4-5 або 1. Коефіцієнт Жаккара також має відмінну синергію з системою Redis, що дозволяє виконувати операції порівняння множин даних у лінійний час.

В якості рекомендаційного алгоритму обрано класифікаційний метод пошуку k-найближчих сусідів. Щоб впоратися з великою кількістю користувачів, є важливим зробити оптимізації, які не передбачають порівняння кожного користувача з кожним іншим існуючим користувачем. Одним з шляхів впоратися з цією проблемою, є використання методу k-найближчих сусідів, що дозволяє порівнювати користувачів лише з його найближчими сусідами.

Після того, як подібність користувача була підрахована за допомогою коефіцієнта Жаккара, створюється відсортована множина, що відображає наскільки

користувач є подібним до інших користувачів. Користувачі з найбільшим значенням подібності вважаються його найближчими сусідами.

Рекомендаційні системи та колаборативне фільтрування запозичили багато ідей із статистики. Одним з прикладів є вимірювання кореляції між користувачами як кореляції між випадковими величинами через кореляцію Пірсона. Таким чином, це не дивно, що представлення оцінок довіри користувачам було зроблено навіть у попередніх системах, таких як MovieCritic3. Для покращення результатів видачі в рамках побудови найкращих рекомендацій топ-N фільмів використовується характеристика довіри системи до своїх прогнозів.

У контексті побудови списку найкращих рекомендацій з N записів є впевненість системи у прогнозуванні також можна інтерпретувати як силу рекомендації, тобто коли система впевнена, що певний предмет є адекватним для активного користувача. Коли система не впевнена, чи підходить елемент для даного користувача, ми можемо назвати це слабкою рекомендацією. В цій роботі ми використовуємо термін впевненість, але термін сила також має місце.

Іншим пов'язаним терміном у рекомендації найкращих топ-N фільмів є „актуальність”. Можна сказати, що система намагається рекомендувати лише релевантні фільми. Таким чином, низька впевненість може бути пов'язана з сумнівною релевантністю. Зрозуміло, коли система впевнена, що фільм не є актуальним, він не буде надавати така рекомендації. Коли система відображає нерелевантний фільм, це може бути тому, що вона не була впевнена у собі чи це було доречним чи ні.

Щоб прибрати з результатів нерелевантні рекомендації використаємо довірчий інтервал оцінок Вільсона. Він дозволить відфільтрувати ті фільми, що не отримали достатньо оцінок та мають низький рівень довіри

У розділі з охорони праці було проаналізовано умови праці на робочому місці програміста.

ЗАГАЛЬНІ ВИСНОВКИ

Бурхливий розвиток ІТ обумовлює необхідність розвитку інформаційних систем перегляду відеоконтенту з вбудованими рекомендаційними системами.

Проведене дослідження дозволяє зробити наступні висновки.

Здійснений аналіз показав, що сучасний стан інформатизації суспільства супроводжується накопиченням відеоконтенту у мережі Інтернет. У таких умовах існує потреба у автоматизованих системах перегляду відеоконтенту, у яких є вбудовані рекомендаційні системи на основі уподобань користувача. Наявність таких систем сприяє підвищенню ефективності систем перегляду відеоконтенту. Виявлено, що сьогодні існує багато кросплатформених сервісів для перегляду ліцензійного відеоконтенту на комерційній основі та безкоштовно. Однак системи з високою швидкістю і точністю розпізнавання уподобань є дорогими, що обумовлює необхідність створення доступних веб-застосунків для перегляду відеоконтенту з вбудованою рекомендаційною системою.

У результаті проведеного дослідження встановлено, що рекомендаційна система – це система, що будує рейтинг відеоконтенту серед великого набору даних, на підґрунті вже існуючої інформації про уподобання користувача при перегляді відео. Виявлено, що у більшості рекомендаційних систем використовуються рекомендаційні алгоритми колаборативної та контентної фільтрації та підхід, який базується на поєднанні цих алгоритмів – гібридна фільтрація. Колаборативна фільтрація передбачає здійснення кластеризації для виявлення користувачів з подібними характеристиками в оцінці відео й на основі цього прогнозує персональні рекомендації для відео перегляду. Контентна фільтрація передбачає надання рекомендацій для перегляду на основі ретроспективних даних про відеоконтент. Гібридна фільтрація поєднує у собі переваги алгоритмів колаборативної та контентної фільтрації й є найбільш популярною.

Серед інструментальних засобів розробки інформаційної системи перегляду відеоконтенту було обрано середовище розробки програмного забезпечення IDE

IntelliJ IDEA, яке підтримує інструменти для проведення тестування TestNG і JUnit, системи контролю версій CVS, Mercurial і Git та має вбудований плагін DataGrid для зручної роботи з базами даних. Для розробки серверної частини веб-додатку було обрано мову програмування JavaScript та платформу NodeJS, яка надає можливість виконувати JavaScript-скрипти на сервері та направляти їх результат користувачеві. Для серіалізації даних було використано JSON. Для розробки графічного інтерфейсу було обрано підхід одно сторінкових додатків, який передбачає оновлення лише частини вмісту при переході користувача на нову сторінку.

Здійснено розробку, програмну реалізацію, тестування та впровадження інформаційної системи перегляду відеоконтенту з вбудованою рекомендаційною системою, який дозволяє створювати користувачам особисті сторінки з інформацією про вподобання, здійснювати пошук та переглядати відеоконтент у зручному для користувача режимі, формувати персональний список відео. У додатку передбачена можливість оцінки переглянутого відео користувачем, збір та аналіз зібраної інформації, демонстрація проаналізованих даних.

У спеціальному розділі було проаналізовано умови праці на робочому місці програміста.

Поставлені завдання виконано повністю, однак є проблеми, які потребують подальшої розробки: налаштування більш глибокого динамічного аналізу зібраної інформації.

АНОТАЦІЯ

Волков Андрій Михайлович. Інформаційна система перегляду відеоконтенту. – На правах рукопису.

Бакалаврська кваліфікаційна робота на здобуття освітньої кваліфікації «бакалавр з комп'ютерних наук» в галузі знань 12 «Інформаційні технології» за спеціальністю 122 «Комп'ютерні науки».

Чорноморський національний університет імені Петра Могили, Миколаїв.

Дипломна робота присвячена проектуванню, розробці, програмній реалізації та впровадженню веб-додатку для перегляду відеоконтенту із вбудованою системою рекомендацій.

Об'єкт дослідження – перегляд відеоконтенту в мережі Інтернет.

Предмет дослідження – веб-орієнтовані програмні засоби для перегляду відеоконтенту із використанням алгоритмів побудови рекомендацій.

Метою дипломної роботи є поліпшення перегляду відеоконтенту шляхом розробки і впровадження веб-додатку із вбудованою системою рекомендацій по перегляду.

Дипломна робота складається з фахової частини і спеціальної частини з охорони праці.

Пояснювальна записка дипломної роботи складається зі вступу, чотирьох розділів, висновків та додатків.

У першому розділі проводиться дослідження теоретичних засад побудови рекомендаційних алгоритмів для сервісів перегляду відеоконтенту в мережі Інтернет. У другому розділі розглядаються технології і засоби розробки веб-додатку. У третьому розділі описано проектування та програмну реалізацію веб-додатку для перегляду відеоконтенту.

У четвертому розділі розкрито питання спеціальної частини з охорони праці.

Дипломна робота містить 79 сторінок, 22 рисунки, 33 джерела, 7 додатків.

Ключові слова: рекомендаційна система, колаборативне фільтрування, NodeJS.

ABSTRACT

Volkov Andrii. Video content viewing information system. – On the rights of the manuscript.

Bachelor's qualification work for the educational qualification "Bachelor of Computer Science" in the field of knowledge 12 "Information Technology" in the specialty 122 "Computer Science".

Petro Mohyla Black Sea National University, Mykolaiv.

Graduate work is devoted to the design, development, software implementation of a web application for viewing media-content with integrated recommender-system.

The object of research is viewing of media-content in the Internet.

The subject of the research is a web-based software for viewing media-content with integrated recommendation algorithms.

The purpose of the thesis is to improve viewing experience of media-content by developing and implementing a web application with integrated recommender system.

Explanatory note of the thesis consists of an introduction, four chapters, conclusions and appendix.

The first section examines the theoretical foundations of the building recommender algorithms for services serving media-content in the Internet. The second section discusses the technologies and tools for developing a web application. The third section describes the design and software implementation of a web application for viewing-media-content with integrated recommendation algorithms.

The fourth section reveals the issue of a special part of labor protection.

Thesis contains 79 pages (without appendix), 22 figures, 33 sources, 7 supplements.

Keywords: recommender system, collaborative filtering, NodeJS.