

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук
(повна назва)
Кафедра _____ Системотехніки
(повна назва)

АТЕСТАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти _____ другий (магістерський)
(рівень вищої освіти)

Методи аналізу медіаконтенту в інформаційно-аналітичних системах
(тема)

Виконав: студент 2 курсу, групи _____
_____ СПРМ-18-2
_____ Лазарев Я.О.
(прізвище, ініціали)

Спеціальність 122 – Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Системне проектування
(повна назва освітньої програми)

Керівник _____ проф. Нечипоренко А.С.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри СПР

(підпис)

_____ Гребеннік І.В.
(прізвище, ініціали)

2020 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)
Кафедра Системотехніки
(повна назва)
Рівень вищої освіти другий (магістерський)
Спеціальність 122 – Комп'ютерні науки
(код і повна назва)
Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма Системне проектування
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
« ____ » _____ 2020 р.

ЗАВДАННЯ
НА АТЕСТАЦІЙНУ РОБОТУ

Студенту Лазареву Ярославу Олександровичу
(прізвище, ім'я, по батькові)

- 1 Тема роботи: Методи аналізу медіаконтенту в інформаційно-аналітичних системах
затверджена наказом університету від 30 03 2020 р. № 477 СТ
 - 2 Термін подання студентом роботи (проекту) 26 травня 2020р.
 - 3 Вихідні дані до роботи : Розробити модифікацію метода аналізу медіаконтенту та його програмну реалізацію
 4. Перелік питань, що потрібно опрацювати в роботі 4.1 Вступ; 4.2 Аналіз предметної області та постановка задачі дослідження, 4.3 Загальні положення інформаційно-аналітичних систем 4.4 Контент-аналіз як метод аналізу медіаконтенту, 4.5 Програмна реалізація, 4.6 Висновки
 5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій: Структура програми, Головна сторінка сайту, Сторінка «Топ 100», Сторінка «Анонси», Сторінка для входу та реєстрації, Функціонал для зв'язку з адміністратором сайту, Використання методу контент-аналізу
- _____
- _____
- _____
- _____

6. Консультанти розділів роботи (проекту)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про	
		підпис	дата
Спец. частина	проф. Нечипоренко А.С.		

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на виконання роботи	30.03.2020	
2	Огляд літератури та аналіз предметної області	31.03-10.04.2020	
3	Огляд та вибір методу розв'язання задачі	11-15.04.2020	
4	Вибір мови та середовища розробки ПЗ	16.04.2020	
5	Розробка алгоритму	17-20.04.2020	
6	Розробка програмного забезпечення	21-28.04.2020	
7	Проведення експерименту	29.04-3.05.2020	
8	Обробка результатів	4-9.05.2020	
9	Оформлення пояснювальної записки	10-14.05.2020	
10	Оформлення додатків	15-21.05.2020	
11	Представлення на рецензування	22.05.2020	
12	Представлення роботи	28.05.2020	

Дата видачі завдання 30 березня 2020 р.

Студент _____
(підпис)

Лазарєв Я.О.

Керівник роботи _____
(підпис)

проф. Нечипоренко А.С.
(посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка до магістерської атестаційної роботи: 61 сторінка, 19 рисунків, 2 додатка, 18 джерел інформації.

СИСТЕМА, ІНФОРМАЦІЙНО-АНАЛІТИЧНА СИСТЕМА, АНАЛІЗ, МЕДІАКОНТЕНТ, КОНТЕНТ АНАЛІЗ

Мета магістерської атестаційної роботи – вдосконалення методу аналізу медіаконтенту та його реалізація на основі веб-додатку.

Об'єкт розробки- процес аналізу медіаконтенту в інформаційно-аналітичних системах.

Предмет дослідження – методи аналіза медіаконтенту в інформаційно-аналітичних системах.

Результати магістерської атестаційної роботи – вдосконалений метод аналізу медіаконтенту та його реалізація на основі веб-додатку.

Область застосування – рекомендаційні системи.

ABSTRACT

Thesis contains: 61 pages, 19 images, 2 applications, 18 sources.

SYSTEM, INFORMATION-ANALYTICAL SYSTEM, ANALYSYS,
MEDIACONTENT, CONTENT-ANALYSIS

The purpose of the master's certification work is to improve the method of media content analysis and its implementation on the basis of a web application.

The object of development is the process of analysis of media content in information-analytical systems.

The subject of research is the methods of media content analysis in information-analytical systems.

The result of the master's thesis is an improved method of media content analysis and its implementation on the basis of a web application.

Field of application - recommendation systems.

ЗМІСТ

Вступ.....	8
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ І ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ.....	10
1.1 Медіаконтент та його реалізація в сучасних інформаційно-аналітичних системах	10
1.1.1 Основні види медіаконтенту.....	11
1.1.2 Методи аналізу медіаконтенту.....	12
1.2 Аналіз функціоналу інформаційно-аналітичної системи, де реалізовано аналіз медіаконтенту.....	14
1.3 Аналіз характеристик існуючих систем.....	21
1.4 Постановка завдання.....	22
2 ЗАГАЛЬНІ ПОЛОЖЕННЯ ІНФОРМАЦІЙНО-АНАЛІТИЧНИХ СИСТЕМ.....	24
2.1 Теоретичні основи інформаційно-аналітичних систем.....	24
2.1.1 Типи інформаційно-аналітичних систем,.....	25
2.2 Теоретичні основи рекомендаційних систем.....	26
2.3 Технології створення веб-додатків.....	27
2.3.1 Загальні положення.....	27
2.3.2 Технології веб-додатків.....	28
2.3.3 Види веб-систем і веб-додатків.....	30
2.4 Розробка функціоналу інформаційно-аналітичних систем	31
3 КОНТЕНТ-АНАЛІЗ ЯК МЕТОД АНАЛІЗУ МЕДІАКОНТЕНТУ.....	35
3.1. Контент-аналіз, переваги і недоліки методу.....	35
3.2 Методи контент-аналізу.....	37
3.3 Одиниці контент-аналізу	39
3.4 Удосконалений метод аналізу медіаконтенту.....	41
4 ПРОГРАМНА РЕАЛІЗАЦІЯ	44
4.1 Інструментарій для розробки.....	44

4.2 Структура реалізації веб- проекту.....	46
4.3 Технічні особливості побудови веб-додатків.....	49
4.3.1 Структура веб-додатків.....	49
4.4 Графічний інтерфейс.....	52
Висновки.....	57
Список літератури.....	58
Додаток А.....	60
Додаток Б.....	65

ВСТУП

У сучасному світі кожен користувач стикається з величезною кількістю інформації. І, оскільки в даний час людям доступна величезна кількість інформації, вони не здатні пропустити її всю через себе в пошуках того, що їм подобається. Це ускладнює вибір і робить цей процес більш тривалим.

Спростити цю задачу покликані інформаційно-аналітичні системи. Зараз більшість масштабних комерційних і соціальних сайтів використовують механізми, які сортують значні обсяги даних для визначення потенційних переваг користувачів.

Призначення інформаційно-аналітичних систем полягає в обробці інформації про користувача на основі наступного: дій користувача на веб сторінці, звичайного опитування, заповненого користувачем. За допомогою певних алгоритмів і отриманої інформації можна визначити, наприклад, які товари або послуги можуть бути корисні і придбані користувачем. Часто їх використовують для розміщення ефективної реклами.

Мета магістерської атестаційної роботи – вдосконалення методу аналізу медіаконтенту та його реалізація на основі веб-додатку.

Об'єкт розробки- процес аналізу медіаконтенту в інформаційно-аналітичних системах.

Предмет дослідження – метод аналізу медіаконтенту в інформаційно-аналітичних системах.

Для реалізації вказаної мети необхідно вирішити наступні основні завдання:

- провести аналіз предметної області;
- проаналізувати існуючі методи аналізу медіаконтенту та інформаційно-аналітичні системи, де вони реалізовані;
- удосконалити метод аналізу медіаконтенту та розробити його програмну реалізацію на створеному сайті;

Наукова новизна отриманих результатів полягає в тому, що вперше запропоновано вдосконалений метод аналізу медіаконтенту, який відрізняється від

існуючих областей та способом застосування.

Практичне значення одержаних результатів полягає у реалізації розробленого методу на основі веб-додатку у рекомендаційних системах пошуку фільмів.

За результатами виконання магістерської атестаційної роботи опубліковано тези доповіді у міжнародній студентській науковій конференції «Модернізація та сучасні українські та світові наукові дослідження».

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ І ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

1.1 Медіаконтент та його реалізація в сучасних інформаційно-аналітичних системах

Інтернет простір розвивається з кожним днем. Ми живемо в інформаційному суспільстві, де використання інформаційно-аналітичних систем є необхідним. Інформаційно-аналітичні системи (ІАС) призначені для накопичення знань і надання даних для прийняття рішень на основі комплексного аналізу інформації, у тому числі медіаконтенту.

Разом з розвитком технологій змінюється і інформація. А точніше спосіб її подачі. Якщо раніше інформація подавалася в текстовому і графічному вигляді, то тепер це неактуально. На перший план виходить оперативність, зручність використання інтернет ресурсів. Саме тому аналіз медіаконтенту в інформаційно-аналітичних системах сьогодні є наймовірніше актуальним, так само як і знаходження нових шляхів для аналізу для того, щоб економити час користувачів.

Сьогодні існує таке поняття, як медіа контент. Це спосіб подачі інформації за допомогою медіа технологій, за коштами аудіо і відео.

На сьогоднішній день медіа контент включає в себе різні складові. Це відеоролики, аудіо файли, зображення, ігри. Такий вид контенту представляє інтерес для інтернет-аудиторії. Тому його дуже вигідно використовувати для залучення інтернет-користувачів на різні ресурси. Дає можливість більш широко використовувати всі можливості інтернету.

Зараз якість інтернет-провайдерів набагато вище, ніж було раніше. Швидкість передачі і прийняття файлів займає лічені секунди. Тому для користування медіа контентом не виникає ніяких труднощів [1].

Також перевага такого виду контенту в тому, що більшість медіа файлів знаходяться у вільному доступі. Зараз не виникає труднощів з пошуком необхідної пісні, зображення плінтуса для стільниці з каменю, або того ж відеоролика про

улюблених кішок.

Так як існує чимала кількість соціальних мереж, будь-який користувач може формувати свій власний медіа контент. Це можна створити в особистому профілі. Ви можете додавати туди медіа файли, взяті з інтернету, а так само ті, які знаходяться у вас на комп'ютері. Різні відеоролики, ігри, аудіо файли або зображення будуть знаходитися в вашому профілі. До них будуть мати вільний доступ усі користувачі. Хоча зараз з'явилася і така можливість, як обмеження доступу до медіа файлів на ваше особисте бажання.

1.1.1 Основні види медіаконтенту

Типи медіаконтенту

1. Оригінальний професійний контент. Даний тип контенту створюється ЗМІ та іншими професійними виробниками контенту (продакшн-студії і т. П.). До такого контенту відносяться інформаційні повідомлення ЗМІ, професійні фото-, аудіо- і теле- (кіно-) матеріали, готові статті, програми, передачі, фільми, інформаційні та розважальні інтернет-ресурси.

2. Любительський контент, вироблений споживачами або User Generated Content. До даного типу контенту відносяться записи в блогах (різного масштабу і орієнтації) інтернет-користувачів, записи в форумах, коментарі до повідомлень ЗМІ на сайтах ЗМІ, записи на персональних сторінках в соціальних мережах, створені споживачем фото-, відео- і аудіоконтент, інтернет-ресурси, створені користувачами [1].

3. Третім типом контенту є продукт, який створюється штучним інтелектом, тобто роботами. Найбільш популярним типом такого контенту є новинні агре- Гатор і пошукові системи, які фактично структурують новинні потоки автоматично, тобто виконують фундаментальну функцію ЗМІ у встановленні пріоритетності новин. Також сюди відносяться автоматизовані системи створення корпоративної новин і листування.

1.1.2 Методи аналізу медіаконтенту

Насправді, у аналізі немає свого характерного, специфічного або особливого методу, який використовувався б саме для роботи з медіа. Прийнято використовувати міждисциплінарні методи, наприклад, такі як анкетування або контент-аналіз або методи «родинних» наук: соціології, психології, лінгвістики.

Перераховані далі методи можуть бути пов'язані між собою, що впливають одне з одного.

1) Спостереження (польове)

Використовується спостереження в якісних медіаисследования як цілеспрямоване пасивне вивчення предметів. «В процесі ми отримуємо знання не тільки про зовнішні сторони об'єкта пізнання, а й про його істотні властивості і відносини». Важливим моментом спостереження є інтерпретація його результатів - розшифровка показань приладів і т.п.

Спостереження при аналізі медіаконтенту частіше буває допоміжним методом, який використовується на попередньому етапі дослідження для постановки проблем, висунення гіпотез або перевірки теорій.

Можливості методу: дослідник знаходиться віч- на-віч безпосередньо до досліджуваного явища або ситуації, може розглянути відразу кілька параметрів, незалежно від готовності спостережуваних об'єктів. Обмеження: повторити цю ж ситуацію вже буде неможливо, якщо буде необхідність доповнити інформацію або перевірити її правдивість; дуже легко втратити важливі моменти або проінтерпретувати їх неправильно (через якихось суб'єктивних факторів з боку спостерігача) [2].

2) Лабораторний експеримент

Лабораторний експеримент це активне і цілеспрямоване втручання у перебіг досліджуваного процесу, відповідну зміну досліджуваного об'єкта або його відтворення в спеціально створених і строго контрольованих умовах, що визначаються цілями експерименту.

Лабораторний експеримент вважається одним з найпопулярніших. Часто

беруть дві групи: одна дивиться / слухає / читає певні типи медіа-продукції, які викликають певні емоції (наприклад, сцени насильства або кадри з місць катастроф), другої групи ж представляють медіа-продукцію більш нейтрального змісту. Потім ті зміни, які піддаються вимірюванню, порівнюються у представників обох груп і робляться відповідні висновки.

Головне достоїнство методу, на відміну від польового (природного) спостереження, це можливість його відтворення, повторення. Недолік же полягає в тому, що не завжди об'єкти поводяться в штучних умовах так само, як в реальності; виникають проблеми з репрезентативністю.

3) Польовий експеримент

Його відмінність від лабораторного полягає лише в тому, що він проводиться в реальних умовах.

Плюс такого методу в тому, що ставлення або поведінку учасників експерименту вивчається не в штучних умовах лабораторії, а в реальних ситуаціях, що підвищує ймовірність того, що і поведінку максимально наближене до природного і справжнього. Обмеження в тому, що він дає менше можливостей для контролю, ніж лабораторний, однак, використання методів статистичного аналізу дозволило вченим домогтися більшої контрольованої сторонніх змінних.

4) Контент-аналіз

Контент-аналіз це один із способів кабінетного дослідження, який дозволяє виявити наявність, відсутність або кількість певних медіа повідомлень, які, імовірно, сприяють певних дій.

Переваги методу: можлива точна фіксація показників в об'ємних масивах емпіричних даних і їх надійність, мінімізація впливу суб'єктивного фактора (дослідника) на об'єкт. Недоліки: труднощі визначення набору категорій аналізу, велика трудомісткість процедури [2].

5) Глибинне інтерв'ю

Глибинне інтерв'ю це слабоструктурована, вільна особиста бесіда інтерв'юєра з респондентом. Мета: вивести респондента до докладних відповідей на поставлені запитання. В ході розмови дослідник отримує від респондента дуже

докладну інформацію про причини його дій, про ставлення до різних питань.

Можливості методу: хороший шанс розібратися у внутрішніх переживаннях людей, вийти на довірчий рівень, що значно підвищує шанс чесних відповідей. Обмеження: при недостатній компетенції і професіоналізмі, дослідник може впливати на відповіді респондента, тим самим спотворюючи дійсність; великі витрати на проведення.

б) Фокус-група

Фокус-група відноситься до якісних методів аналізу. Його принцип полягає в наступному: збирається заздалегідь обрана з якихось критеріїв (які залежать від мети і характеру дослідження) група учасників (зазвичай до 12 осіб) і разом з модератором обговорюють певну тему. В ході обговорення виявляються важливі пункти: чому люди діють певним чином і що думають з приводу якихось питань.

Переваги методу: дозволяє простежити взаємодію учасників з різними думками, побачити логіку їхніх суперечок, почути аргументацію. Недоліки: при недосвідченому або непрофесійному модератором деякі особистості можуть залишатися "нерозкритими" і досить важко дізнатися переживання за темою кожного в однаковій мірі.

1.2 Аналіз функціоналу інформаційно-аналітичної системи, де реалізовано аналіз медіаконтенту

Рекомендаційні системи - це комплекси алгоритмів, які намагаються передбачити які об'єкти (товари) будуть цікаві користувачеві, маючи певні дані про його профіль (інтересах, історії переглядів, оцінок).

Традиційно рекомендаційні системи поділяють на чотири типи :

1) колаборативна фільтрація (collaborative filtering).

Рекомендації на основі оцінок інших користувачів.

Приклади умов:

Користувачеві N, зі схожими з вашими оцінками, сподобався цей фільм, ймовірно вам він теж сподобається. Фільм M схожий на фільми, які ви вже високо

оцінили, можливо він вам теж сподобається.

Плюси:

Теоретично висока точність.

Мінуси:

Високий поріг входу: не знаючи нічого про інтереси користувача, рекомендації практично марні, багато користувачів будуть просто відразу йти.

Коллаборативна фільтрація завжди передбачає те, що у нас є дані про користувачів і об'єктах, які вони оцінили (в тому числі про поточного користувача). Тоді фільтрація розділяється на два типи: user-based і item-based (Рисунок 1.1).

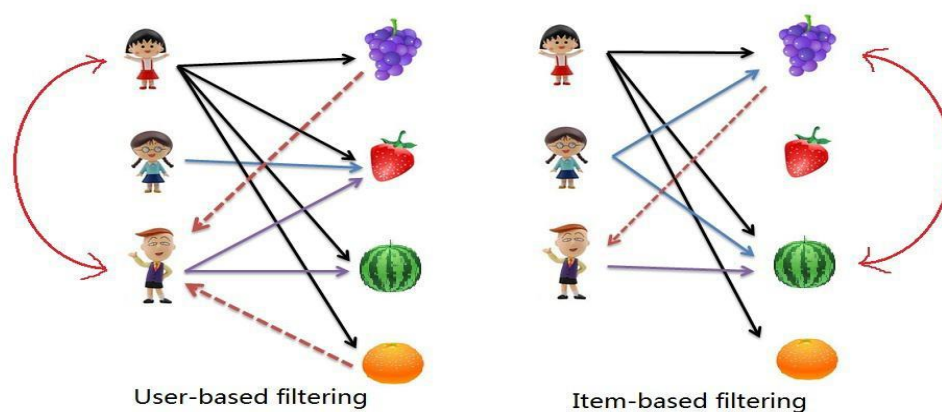


Рисунок 1.1 - Типи коллаборативної фільтрації

У разі user-based рекомендації будуються за принципом: «знайти користувачів, схожих на мене, і подивитися чи подобається їм даний об'єкт». Наприклад: «Користувачі Альоша і Борис люблять гамбургери, латте і стартапи, а користувач Валера любить гамбургери і латте, значить швидше за все йому потрібно починати стартап» [3].

У разі item-based логіка рекомендацій повертаються іншою стороною: «знайти об'єкти, схожі на заданий, і подивитися як я їх оцінював раніше». Наприклад: «Чи сподобається Валері робити стартапи? Валера любить латте і м'ячболі, значить швидше за все йому сподобаються стартапи».

Коллаборативна фільтрація в теорії - штука досить проста: знайти користувачів, які оцінили даний об'єкт, а потім порахувати коефіцієнт схожості векторів їх оцінок всіх об'єктах в базі. Взяти, наприклад, k користувачів з

найвищими коефіцієнтами схожості і подивитися як вони найчастіше оцінювали даний об'єкт (типовий алгоритм k-найближчих сусідів). Для збільшення точності не забути розділити кожну оцінку на середню оцінку юзера (щоб згладити вплив «Хейтер» і «мені подобається все» користувачів) [4].

Тут можна згадати ще метод SVD (Сингулярне Розкладання Матриць), який на даний момент є передовим в області колаборативної фільтрації.

2) системи, засновані на контенті (content-based).

Рекомендації на основі даних, зібраних про кожен конкретний товар.

Приклади умов:

Починаючи від простих: «Книги того ж жанру чи автора», «Речі того ж виробника», до «Вам подобаються комедії з Джимом Керрі, ось вам добірка комедій з НИМ».

Приклади сервісів:

Prismatic (Рисунок 1.2), Surfingbird.

Плюси:

Можна робити рекомендації навіть незнайомим користувачам, тим самим залучаючи їх в сервіс. Можливість рекомендувати ті об'єкти, які ще не були ніким оцінені.

Мінуси:

Точність сильно падає, час розробки трохи зростає.

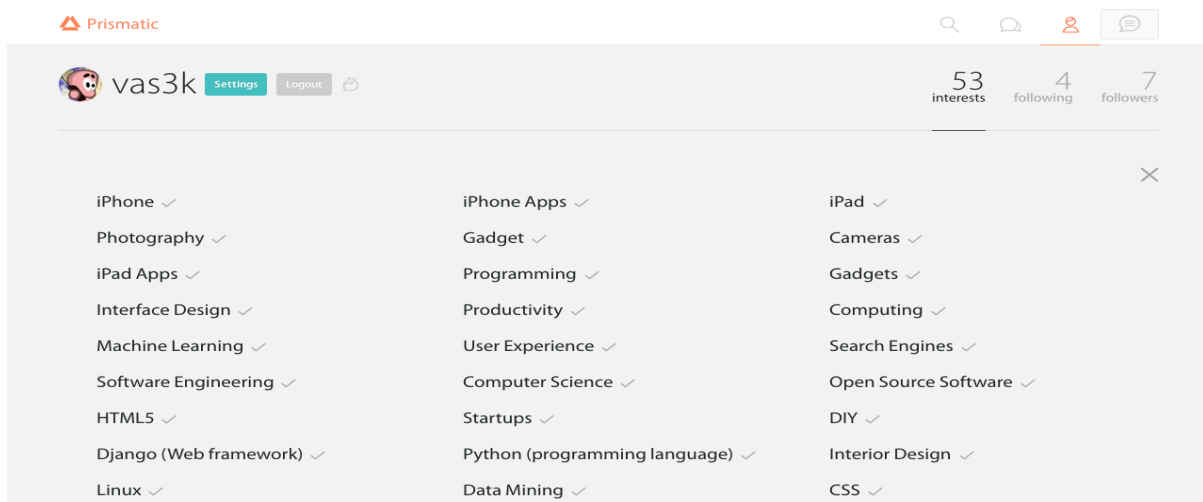


Рисунок 1.2 - Prismatic - типова content-based система

Content-based системи спираються на знання про об'єкти, абсолютно ігноруючи як там ці об'єкти оцінювали інші користувачі [3]. Як знань про об'єкт виступають будь-які його властивості, які ми можемо добути: жанр, виробник, автор, країна, в загальному все, що ви зможете добути в вашій області. Тільки захоплюватися теж не варто, наприклад, виробник процесора не часто є релевантним властивістю при виборі смартфона.

Останнім часом саме content-based рекомендації стають найпопулярнішими, тому що ніхто з користувачів не хоче витратити час на «навчання» системи своїм уподобаннями, а ніхто з создателів систем не хоче втрачати купу користувачів відразу після реєстрації. Звідси і маємо купу скарг і жартів про те, що «я купив джинси на ібее і він тепер мені постійно джинси пропонує» або «зайшов на сайт з обручками, тепер Яндекс. Директ всюди пропонує купити їх ».

Content-based системи стали настільки популярні, що всіх починає нудити від їх дурних і недоречних рекомендацій. Так що думайте, перед тим як реалізовувати чергову контент-орієнтовану систему, чи не буде вона дратувати ваших користувачів як 90% існуючих. Або може у вас є час і мізки, щоб спробувати розширити її додатковими умовами і рушити в бік knowledge-based.

Зате ці системи спираються на відомі нам знання про TF-IDF і схожість документів і практично нічого нового при їх написанні вивчати не доведеться. Та ж обробка мови, та ж класифікація, все це зрозуміло, відомо і досить просто: знайшов схожі документи і видав на сторінці [3].

3) Системи, засновані на знаннях (knowledge-based).

Рекомендації, засновані на знаннях про предметну область (а не про кожен товар). Часто попередній тип (content-based) визначають як окремий випадок knowledge-based, де знаннями є відомості про товар, але content-based має таку широку поширеність, що має сенс виносити його в окремий тип (і в більшості літератури так і роблять). Ці самі додаткові знання дозволяють робити рекомендації не ґрунтуючись на «схожості» чого-небудь, а з більш складними умовами.

Приклади умов:

До цього плеєра вам можливо знадобляться навушники. Ви купили відмінну

камеру, ось вам 10% знижки на сумку для неї.

Приклади сервісів:

Великі інтернет-магазини, наприклад Rozetka, Allo і т. П. (Рисунок 1.3).

Плюси:

Можливість виключити улюблену ситуацію всіх систем рекомендації: «Ви тільки що купили квартиру в Харкові? Ймовірно ось ці 5 квартир в Харкові вам теж знадобляться!». Можливість регулювати ситуації типу «макбук 2007 року оцінили на 5.0, а макбук 2014 на 4.9, значить макбук 2007 року краще».

Мінуси:

Висока складність розробки та збору даних [4].

Рекомендуемый комплект

	Apple iPhone 5s 32Gb (золотистый) 32 990.-	+		Флип-кейс Puro Ultra Slim для iPhone 5 (белый) 1 290.- <input checked="" type="checkbox"/> В комплект	+		Red Line Защитная пленка для iPhone 5 (глянцевая) 290.- <input checked="" type="checkbox"/> В комплект	=	34 570.-
---	--	---	---	--	---	---	---	---	-----------------

Составь свой комплект
Купить комплект

Рисунок 1.3 - Приклад заснованих на знаннях рекомендацій

Рекомендаційні системи, засновані на знаннях, так само не спираються на оцінки інших користувачів, а дивляться тільки на профіль користувача і товару. Подальші шляхи завжди різні. Об'єднує їх те, що кожен шлях має на увазі наявність знань про якусь предметної області. Або про користувачів, або про товари, або ще про що-небудь, що може допомогти ранжирування.

Case-based підхід має на увазі наявність додаткової сутності: вимог користувача (user requirements). Вимоги можуть виглядати як: «сучасна дзеркальна камера з мінімум 32 точками автофокусування і ціною до \$ 2000» і можуть задаватися шляхом вибору потрібних чекбоксів. Завдання підходу полягає просто в тому, щоб знайти рекомендовані товари, згідно з вимогами.

Demographic-based враховує властивості безпосередньо користувача (Рисунок

1.4). Наскільки він забезпечений, де живе, скільки заробляє, і. т. д.

Utility-based розраховує відносну корисність кожного товару для користувача. Якщо користувач гик, швидше за все і фотоапарат він вибере з найбільшою кількістю функцій. Якщо користувач купив фотоапарат, швидше за все він за об'єктивами для нього прийде.

Critique-based мають додатковий інтерфейс для «критики». Наприклад дві осі, за якими користувач буде ставити «хочу дорожче, але з великою кількістю мегапікселів», а система намагатися підлаштуватися під запит.

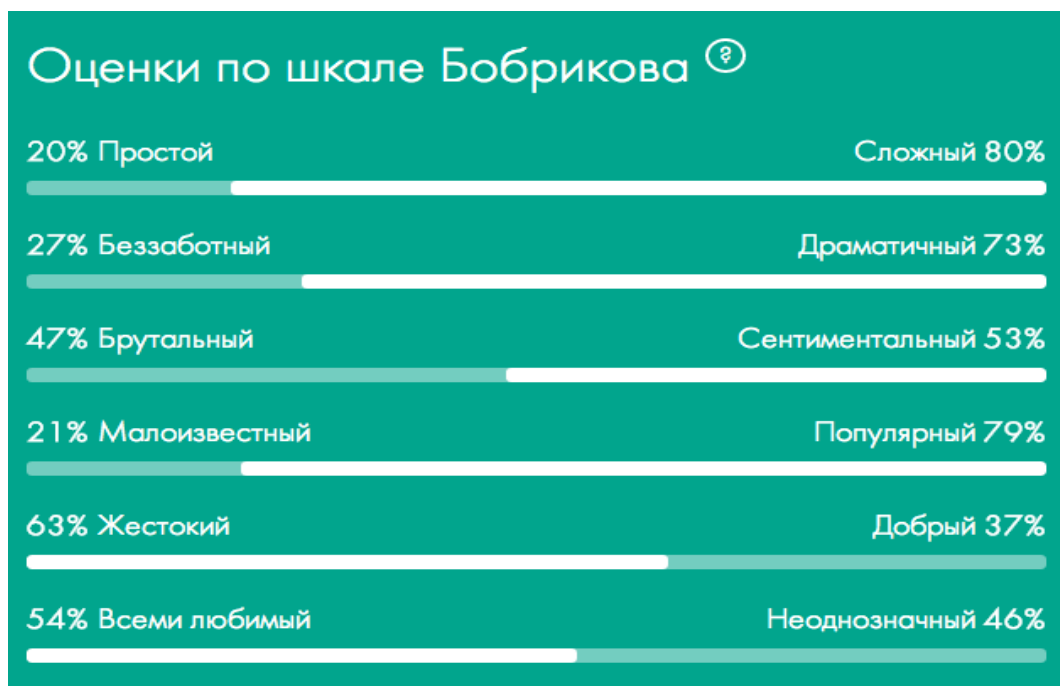


Рисунок 1.4 - Шкала на Імхонета

Whatever-you-want-based як можна помітити, практично всі підвиди knowledge-based систем залежать виключно від уяви їх авторів. У кожному конкретному випадку можна вигадувати свої галузі знань і ранжувати за ним, досягаючи максимальних результатів. Все залежить тільки від вас [4].

4) Гібридні (hybrid) системи.

Як завжди ідея «взяти найкраще від кожного типу» не змусила себе довго чекати. І при певному завзятості, виходить. Наприклад найкрутішою на даний

момент рекомендаційної системою вважається система Netflix (BellKor), яка виграла мільйон доларів на Netflix Prize в 2009-му, вона є комбінацією 27 рекомендаційних алгоритмів.

Приклади умов:

Обмежені тільки уявою творців.

Приклади сервісів:

Netflix (Рисунок 1.5).

Плюси:

Найвища точність, у Netflix похибка приблизно 0.86 бала для 5-бальної шкали оцінок. Хоча, звичайно, все залежить від реалізації.

Мінуси:

Найвища складність розробки.



Рисунок 1.5 - Система рекомендацій Netflix

Всі «справжні» рекомендаційні системи, як і пошукові системи, майже завжди - гібриди. Їх розробники застосовують машинне навчання щоб виділяти самі релевантні для користувача алгоритми, застосовують методи оптимізації для пошуку найкращих коефіцієнтів, в загальному розважаються як хочуть. Єдиного алгоритму побудови гібридних систем немає, до них просто приходять всі, хто спробує вищепереліченими [3].

1.3 Аналіз характеристик існуючих систем

Рекомендаційні системи мають велику популярність [4]. На основі одних тільки рекомендаційних алгоритмів будують цілі проекти, той же LastFM (Рисунок 1.6) або Імхонет, Prismatic або Surfingbird, А Яндекс недавно вирішив знову всіх побити і випустив «рекомендації для всіх і про все» під назвою Яндекс.Атом.

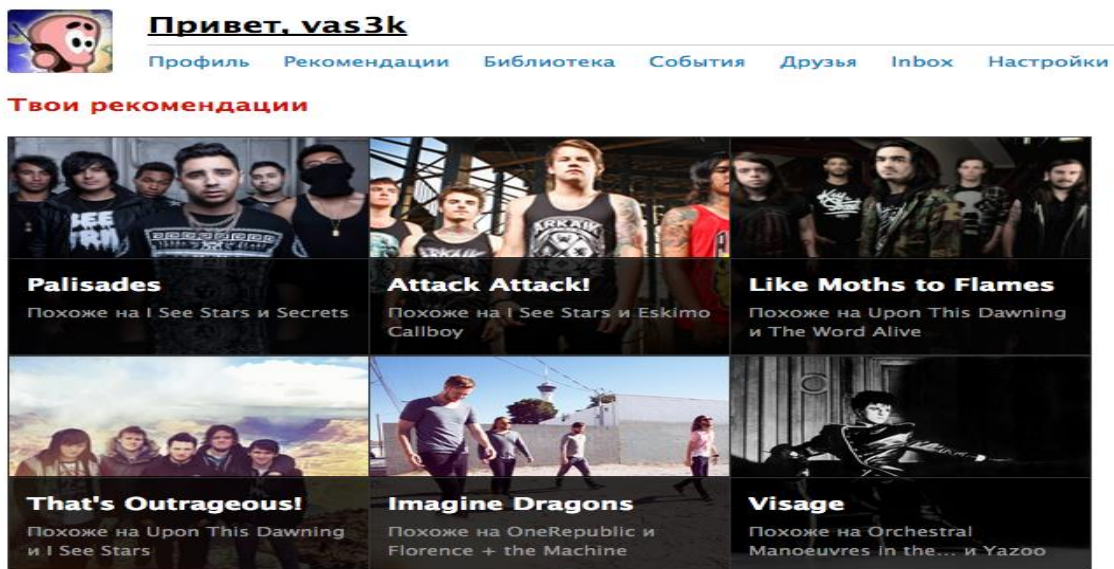


Рисунок 1.6 - Головна сторінка сайту проект last.fm

Ось кілька інших прикладів широко відомих рекомендаційних систем:

Amazon- один з лідерів галузі; Amazon рекомендує вам книги і інші товари, ґрунтуючись на тому, що ви купували, що переглядали, які рейтинги ставили, які залишали відгуки. Так, як це зазвичай і буває, Великий Брат збирає все, навіть якщо щось поки не вміє використовувати.

Netflix- компанія, основою бізнесу якої є оренда фільмів; Зараз компанія перейшла на потокове відео, однак перші десять років свого життя вони розсилали поштою фізичні DVD, які потім потрібно було посилати назад, щоб отримати наступний (гроші при цьому бралися за підписку). Така модель бізнесу виявилася дуже успішною, а опублікований для Netflix Prize набір даних на кілька років став головним тестовим прикладом для систем колаборативної фільтрації (зараз Netflix прибрала його з відкритого доступу через можливу деанонімізація, і на зміну йому

прийшов Yahoo! KDD Cup Dataset).

Last.fm Pandora рекомендують музику. Вони дотримуються різних стратегій рекомендації: Last.fm використовує, крім власне рейтингів інших користувачів, виключно "зовнішні" дані про музику - автор, стиль, дата, теги і т.п., а Pandora ґрунтується на "утриманні" музичної композиції, використовуючи дуже цікаву ідею - Music Genome Project, в якому професійні музиканти аналізують композицію декількома сотнями атрибутів. Аналізувати композиції автоматично поки, щоправда, ніхто добре не вміє - і це ще одне цікаве додаток зусиль для машинного навчання.

Google, *Yahoo* !, *Яндекс* - можна сказати, що вони теж рекомендують користувачам сайти? Формально - так, але в реальності це інші системи: пошуковики намагаються передбачити, наскільки даний документ релевантний даному запиту, а рекомендувача - намагаються передбачити, який рейтинг даний користувач поставить даному продукту. Звичайно, в успіху пошукових систем є чимала заслуга моделей, заснованих на даних від користувачів (click logs), і, звичайно, пошукова видача часто буває Персоналізація, але все-таки завдання трошки інша. Кілька ближче до нашого завдання проблема того, яку рекламу показувати користувачеві (*AdSense*, *Яндекс.Директ* і т.д.) - тут користувачі дійсно «голосують ногами» за рекламні оголошення, і потрібно «порекомендувати» ті з них, які швидше за все викличуть позитивну реакцію [3]. Але справа ускладнюється економічною стороною питання (рекламодавці платять гроші, між ними потрібно влаштувати аукціон за право розмістити рекламу), так що ці завдання ми теж розглядати зараз не будемо. Однак у провідних пошукових систем є маса побічних проєктів, заснованих на рекомендаційних системах - наприклад, ми вже згадували *Yahoo! Music*.

1.4 Постановка завдання

У першому розділі атестаційної роботи було розглянуто поняття медіаконтенту в інформаційно-аналітичній системі та приклади методів аналізу медіаконтенту.

З усіх методів найбільш детальним є метод контент-аналізу.

Контент-аналіз - техніка укладання, виробленого завдяки об'єктивному і

систематичному виявленню відповідних завданням дослідження характеристик тексту.

Головний недолік цього методу - це досить трудомістка процедура, він досліджує явно виражені параметри, а це істотно звужує пізнавальні можливості методу і фактично зводить його до опису.

Для вирішення цієї проблеми необхідно проводити більш детальний аналіз об'єкта дослідження та враховувати неповерхневі фактори.

Метою роботи являється вдосконалення методу аналізу медіаконтенту та його реалізація на основі веб-додатку.

Для реалізації вказаної мети необхідно вирішити наступні основні завдання:

- провести аналіз предметної області;
- проаналізувати існуючі методи аналізу медіаконтенту та інформаційно-аналітичні системи, де вони реалізовані;
- удосконалити метод аналізу медіаконтенту та розробити його програмну реалізацію на створеному сайті;

2 ЗАГАЛЬНІ ПОЛОЖЕННЯ ІНФОРМАЦІЙНО-АНАЛІТИЧНИХ СИСТЕМ

2.1 Теоретичні основи інформаційно-аналітичних систем

На сьогоднішній день в області інфоомаційних технологій особливої актуальності набувають питання побудови інформаційно-аналітичних систем.

Спочатку розглянемо, що таке інформаційно-аналітична система. Інформаційно-аналітичні системи (ІАС) покликані на основі даних, одержуваних в режимі реального часу, допомагати в прийнятті рішень. ІАС - це сучасний високоефективний інструмент підтримки прийняття тактичних, стратегічних і оперативних рішень на основі оперативного і наочного надання всієї необхідної сукупності даних про користувачів.

Інформаційно-аналітична система (ІАС) — це комп'ютерна система, яка дозволяє отримувати інформацію, створювати її та здійснювати її обробку та аналіз [5].

Завданнями ІАС є ефективно зберігання, обробка та аналіз даних.

Побудова ІАС по суті своїй дуже дорогий і трудомісткий процес. Для того щоб уникнути таких проблем і забезпечити можливість проектування досить уніфікованих ІАС, легко адаптуються до зміни зовнішніх умов, необхідна теоретична і методологічна база і відповідні інструментальні засоби її реалізації.

Проведення аналізу вимагає доступу до великого обсягу даних. Час - найбільш істотний вимір, і зазвичай аналіз проводиться з метою виявлення певних тенденцій: спочатку аналізуються дані про минуле та сьогодні, а потім виявлена тенденція екстраполюється на майбутній період часу.

Визначимо основні вимоги до ІАС.

ІАС повинна:

- 1) зберігати в собі великі обсяги даних;
- 2) підтримувати внутрішню несуперечливість даних;
- 3) підтримувати якісний процес поповнення даних;
- 4) підтримувати високу швидкість отримання даних;

- 5) забезпечувати наявність зручних утиліт перегляду даних;
- 6) забезпечувати повноту і достовірність даних, що зберігаються.

На рисунку 2.1 наведено структурну схему інформаційно-аналітичної системи.

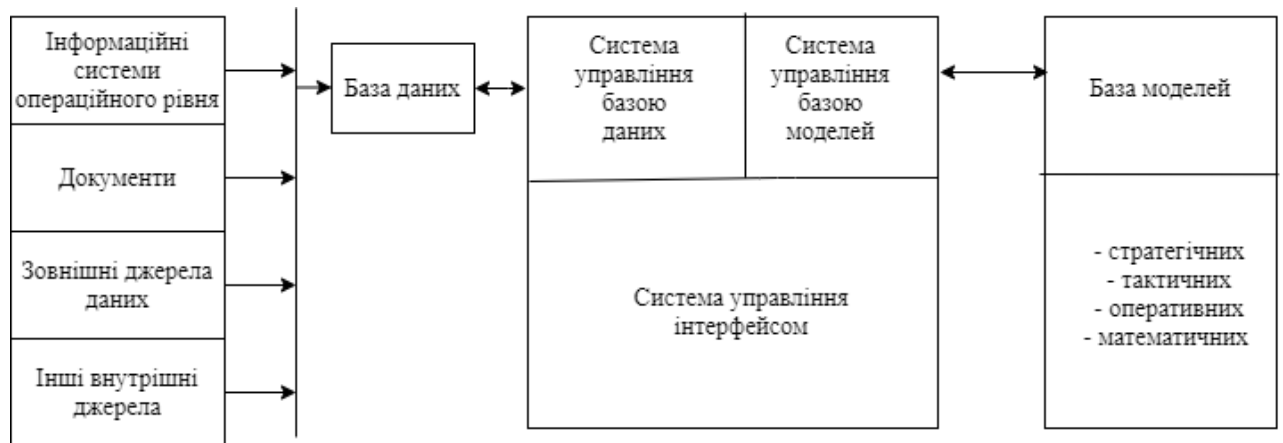


Рисунок 2.1 – Структурна схема інформаційно-аналітичної системи

2.1.1 Типи інформаційно-аналітичних систем

Інформаційно-аналітичні системи (ІАС) покликані на основі даних, одержуваних в режимі реального часу, допомагати в прийнятті рішень.

Основне призначення ІАС - динамічне представлення і багатовимірний аналіз історичних і поточних даних, аналіз тенденцій, моделювання і прогнозування результатів різних управлінських рішень.

Результатом застосування засобів ІАС є з одного боку - регламентні аналітичні звіти, орієнтовані на потреби користувачів різних категорій, з іншого - кошти інтерактивного аналізу інформації і швидкої побудови звітів користувачами-непрограмістів з використанням звичних понять предметної області.

Функцію збору і зберігання інформації з супутньою доопрацюванням в інформаційно-аналітичних системах, виконують інформаційні сховища (Data Warehouse).

У зв'язку з великим обсягом і складністю аналіз даних має два напрямки - оперативний аналіз даних (інформації), широко поширена англійська абревіатура

назви - OLAP. Основним завданням OLAP-аналізу є швидке вилучення необхідної для обґрунтування або прийняття рішення інформації [6].

Інтелектуальний аналіз інформації - має також широко поширене в спеціальній літературі англomовну назву Data Mining. Він призначений для фундаментального дослідження проблем в тій чи іншій предметній області. Вимоги по часу менше жорсткі, але використовуються більш складні методики. Ставляться, як правило, завдання і отримують результати стратегічного значення. Цю функцію виконують різноманітні засоби Data Mining.

Інформаційно-аналітичні системи є надбудовою над уже функціонуючими над інформаційними додатками і не вимагають їх заміни.

Одним з головних видів інформаційно-аналітичної системи є рекомендаційні системи

2.2 Теоретичні основи рекомендаційних систем

Рекомендаційні системи - це моделі, які краще за вас знають, чого вам хочеться [3].

Ці системи з'явилися в інтернеті досить давно, близько 20 років тому. Однак справжній підйом в цій області трапився приблизно 5-10 років тому, коли відбулося змагання Netflix Prize. Компанія Netflix тоді давала в прокат не цифрові копії, а розсилала VHS-касети і DVD. Для них було дуже важливо підвищити якість рекомендацій. Чим краще Netflix рекомендує своїм користувачам фільми, тим більше фільмів вони беруть в прокат. Відповідно, зростає і прибуток компанії. У 2006 році вони запустили змагання Netflix Prize. Вони виклали у відкритий доступ зібрані дані: близько 100 мільйонів оцінок за п'ятибальною шкалою з зазначенням ID проставити їх користувачів. Учасники змагання повинні були якомога краще передбачати, яку оцінку поставитиме певного фільму той чи інший користувач. Якість передбачення вимірювалося за допомогою метрики RMSE (середньо - квадратичне відхилення). У Netflix вже був алгоритм, який передбачав оцінок з якістю 0.9514 за метрикою RMSE. Завдання було поліпшити прогноз хоча б на 10% - до 0.8563. Переможцю був

обіцяний приз в \$ 1000000. Змагання тривало приблизно три роки. За перший рік якість поліпшили на 7%, далі все трохи сповільнилося. Але в кінці дві команди з різницею в 20 хвилин надіслали свої рішення, кожне з яких проходило поріг в 10%, якість у них була однакова з точністю до четвертого знака. У задачі, над якою безліч команд билосся три роки, все вирішили якихось двадцять хвилин. Італійка, яка запізнилася команда (як і багато інших, які брали участь в конкурсі) залишилися ні з чим, однак сам конкурс дуже сильно прискорив розвиток в цій області .

З рекомендаційними системами нам доводиться зустрічатися все частіше і частіше з банальної причини: маркетологи щороку проводять дослідження, в яких величезними відсотками натякають як збільшується кількість покупок, наприклад, в інтернет-магазині, якщо під кожним товаром висить блок «схожі товари», або « можливо ви захочете придбати », або « з даним товаром купують ще ». І це дійсно так.

2.3 Технології створення веб-додатків

2.3.1 Загальні положення.

Веб-додаток - клієнт-серверний додаток, в якому клієнтом виступає браузер, а сервером - веб-сервер. Логіка веб-додатки розподілена між сервером і клієнтом, зберігання даних здійснюється, переважно, на сервері, обмін інформацією відбувається по мережі. Одним з переваг такого підходу є той факт, що клієнти не залежать від конкретної операційної системи користувача, тому веб-додатки є кроссплатформенною сервісами. Веб-додатки стали широко популярними в кінці 1990-х - початку 2000-х років.

За роки існування WWW складу веб-додатків, виконувані ними функції, принципи та архітектура їх побудови зазнали значних змін - від найпростіших засобів зберігання HTML-сторінок до рішень, орієнтованих на підтримку роботи корпоративними інформаційних систем [7].

2.3.2 Технології веб-додатків

1) AJAX (Asynchronous JavaScript and XML) - підхід до побудови призначених для користувача інтерфейсів веб-додатків, при якому у відповідь на кожну дію користувача веб-сторінка на його браузері, що не перезавантажується повністю - з веб-сервера тільки догружаються потрібні йому дані. Цим забезпечується оперативна робота як одного, так і груп користувачів з додатками.

AJAX являє собою не одну, а групу технологій і базується на принципах використання DHTML для динамічної зміни змісту сторінки і використання XMLHttpRequest для звернення до сервера (альтернативний варіант - динамічне підвантаження JavaScript з використанням об'єктної моделі документа - DOM). З урахуванням цих принципів можна створювати зручні веб-інтерфейси на тих сторінках сайтів, де необхідна активна взаємодія з користувачами.

2) ASP (Active Server Pages) - технологія створення веб-додатків, що використовує об'єктну модель інтерфейсу, створеного на основі ISAPI-фільтра. ASP спростила завдання генерації HTML-сторінок і дозволила виробляти звернення до компонентів баз даних [8].

Принцип, закладений в основу інтерфейсу додатку, полягає в тому, що на веб-сторінці присутні фрагменти коду, який інтерпретується веб-сервером і надає користувачеві готовий результат виконання обраних фрагментів коду. веб-сторінка, створена з використанням технології ASP, має розширення «asp».

3) Common Gateway Interface (CGI) - технологія створення веб-додатків, яка є найбільш низкоуровневою і є стандартом інтерфейсу, який служить для зв'язку зовнішньої програми з веб-сервером.

Сам протокол розроблений таким чином, щоб можна було використовувати будь-яку мову програмування, який може працювати зі стандартними пристроями введення / виводу.

Оскільки така можливість є на рівні операційної системи, то, якщо не потрібно складний скрипт, його можна оформити у вигляді командного файлу.

4) Personal Home Page (PHP) - технологія, яка отримала дуже широке

поширення завдяки своїй безкоштовності і підтримки найпопулярніших платформ. Вона базується на принципі побудови сторінок з шаблонів, вперше з'явилося в Active Server Pages, але розвиває і доповнює його [7].

Сторінки PHP мають вигляд звичайних HTML сторінок, в яких можуть використовуватися спеціальні теги виду `<? Php i?>`. Між тегами вставляються рядки програмного коду на спеціальній мові сценаріїв PHP.

Принцип шаблонів дозволив розробникам писати програми набагато швидше і без помилок, властивих традиційним CGI - програм, що видає HTML вміст в потік виводу. На сьогоднішній день діапазон систем, побудованих на шаблонах простирається від простих сторінок з вибірками з бази даних до великих додатків електронної комерції, заснованих на XML. Шаблонні системи користуються великою популярністю серед розробників, оскільки найбільш підходять для типових сайтів. Такі рішення включають ColdFusion, PHP, JSP і ASP, з яких PHP є найбільш поширеною.

5) JSP (Java Server Pages) - технологія створення веб-додатків, заснована на одноразовій компіляції Java-коду (сервлета) при першому зверненні до нього з подальшим виконанням методів цього сервлета і приміщенням отриманих результатів в набір даних, які відправляються в браузер [9].

6) NET - технологія створення веб-додатків, яка є новітньою розробкою компанії Microsoft і заявлена як новий етап в розвитку засобів взаємодії між додатками. На даний момент вона доступна в якості доповнення .NET Framework до сімейства операційних систем Microsoft Windows, а також в новому продукті Windows Server 2003.

Також ведуться роботи по створенню .NET Framework на інших операційних системах. Платформа .NET спрощує розробку додатків і підвищує надійність коду. Зокрема, вона забезпечує автоматичне керування часом життя об'єктів, нейтральні до мов бібліотеки класів і перетинають кордони мов спадкування, обробку винятків і налагодження.

2.3.3 Види веб-систем і веб-додатків

В даний час існує безліч різних веб - систем і веб- додатків. Розглянемо деякі з них:

1) статичні сайти - це сайти, що складаються з веб-сторінок, написаних цілком на html і зберігаються на сервері в тому вигляді, в якому їх створив веб-майстер. Такі сторінки можуть містити клієнтські скрипти, графіку і вставляєміє інтерактивні елементи, але вони не впливають на вміст сторінки, поки не будуть завантажені в браузер. При цьому на сервері сторінка ніяк не зміниться і буде відображатися однаковий для всіх користувачів.

2) динамічні сайти - це сайти, вміст яких генерується «на льоту»: запитувана веб-сторінка формується програмно, за допомогою серверного додатка (з бази даних, що включаються файлів і т.п.). Таким чином, одна і та ж сторінка може виглядати по різному для різних користувачів [9].

3) портал - це багатофункціональний сайт, який є «шлюзом» до різної інформації і онлайн-сервісів. Сайти портального типу складні в розробці і супроводі, але вони охоплюють максимальну кількість користувачів. Всі популярні пошукові системи є сайтами цього типу.

4) спеціалізовані сайти - це категорія, до якої відносяться веб-сайти, орієнтовані на певну тематику, функціональність і / або цільову аудиторію. Розглянемо деякі типи сайтів, які потрапляють в цю категорію:

а) персональний сайт - ресурс, створений і підтримуваний автором і представляє, в першу чергу, інформацію про нього самого і його інтересах. Персональні сайти часто є статичними і розміщуються на безкоштовних хостингових майданчиках.

б) блог (онлайн-щоденник) - просунута різновид персонального сайту, де інформація може надаватися в хронологічному порядку (як в щоденнику). Іронія в тому, що свої рукописні щоденники люди вважають за краще нікому не показувати, а до онлайн-версії намагаються залучити найбільшу кількість користувачів.

в) соціальна мережа - онлайн-сервіс, орієнтований на пошук, встановлення

контактів і спілкування між користувачами системи. Раніше популярні сервіси онлайн-знайомств тепер перейняли багато рис соціальних мереж, тому віднесені до цієї ж групи [7].

г) форум - майданчик для спілкування користувачів, що працює на дискусійному принципі: користувач створює тему, інші учасники можуть приєднатися до її обговорення.

д) корпоративний сайт - офіційне представництво підприємства або організації в глобальній мережі. Розміри і функціональність таких сайтів варіюються від дуже маленьких «візиток» в кілька статичних сторінок з загальною інформацією до корпоративних порталів. Реальність така, що підприємці розглядають Веб щонайменше як рекламний майданчик.

е) інтернет-магазин - найпопулярніша група сайтів для електронної комерції (e-commerce). Інтернет-магазин може виступати як «вітрина» і представляти інформацію про товари і / або послуги, але основна функціональність все-таки спрямована на здійснення покупок в Інтернет. Для цього сайт інтегрується з платіжним шлюзом, обслуговуючим банківські перекази або електронні гроші. Це дає можливість користувачеві оплачувати покупки онлайн, а власнику інтернет-магазину обробляти замовлення і отримувати оплату за них через сайт.

е) wiki-сайт - специфічна різновид сайтів, характерною рисою яких є колаборативного управління інформацією. Користувачі можуть не тільки генерувати контент, але і редагувати інформацію, додану іншими користувачами, обговорювати, приймати або відхиляти правки інших користувачів так само, як це зроблено в Wikipedia.org.

2.4 Розробка функціоналу інформаційно-аналітичної системи

Представлена функціональна схема, у вигляді діаграми прецедентів, на якій можна побачити основні функції для трьох типів користувачів системи. Цими користувачами є [7]:

1) замовник;

- 2) виконавець;
- 3) адміністратор.

Функції замовника:

- 1) реєстрація;
- 2) авторизація;
- 3) відновлення пароля;
- 4) відправка і отримання повідомлень;
- 5) управління профілем (перегляд профілю, зміна профілю);
- 6) управління замовленнями (перегляд, створення, зміна, видалення);

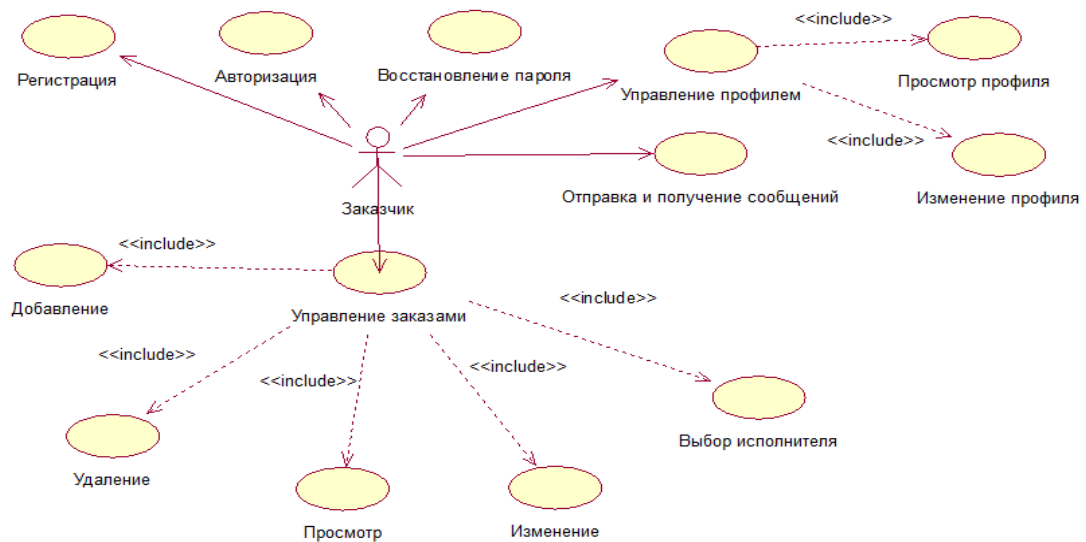


Рисунок 2.2 – Функції замовника

Функції виконавця:

- 1) реєстрація;
- 2) авторизація;
- 3) відновлення пароля;
- 4) відправка й отримання повідомлень;
- 5) управління профілем (просмотр профілю, редагування профілю, редагування портфолію);
- 6) управління замовленнями (подати заявку на виконання замовлення,

виконати замовлення).

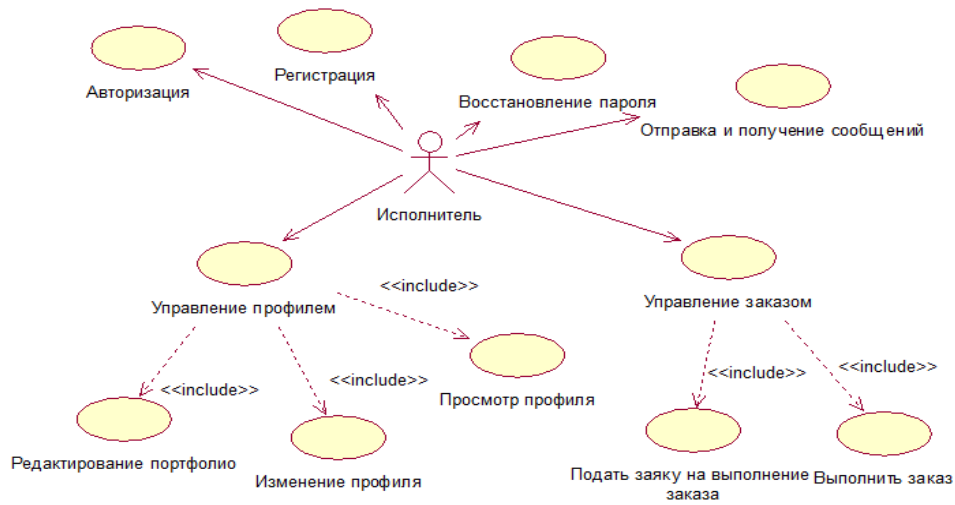


Рисунок 2.3 – Функції виконавця

Функції адміністратора:

- 1) авторизація;
- 2) відновлення пароля;
- 3) управління користувачами (просмотр, додавання, видалення, зміна);
- 4) управління замовленнями (просмотр, додавання, видалення, зміна).

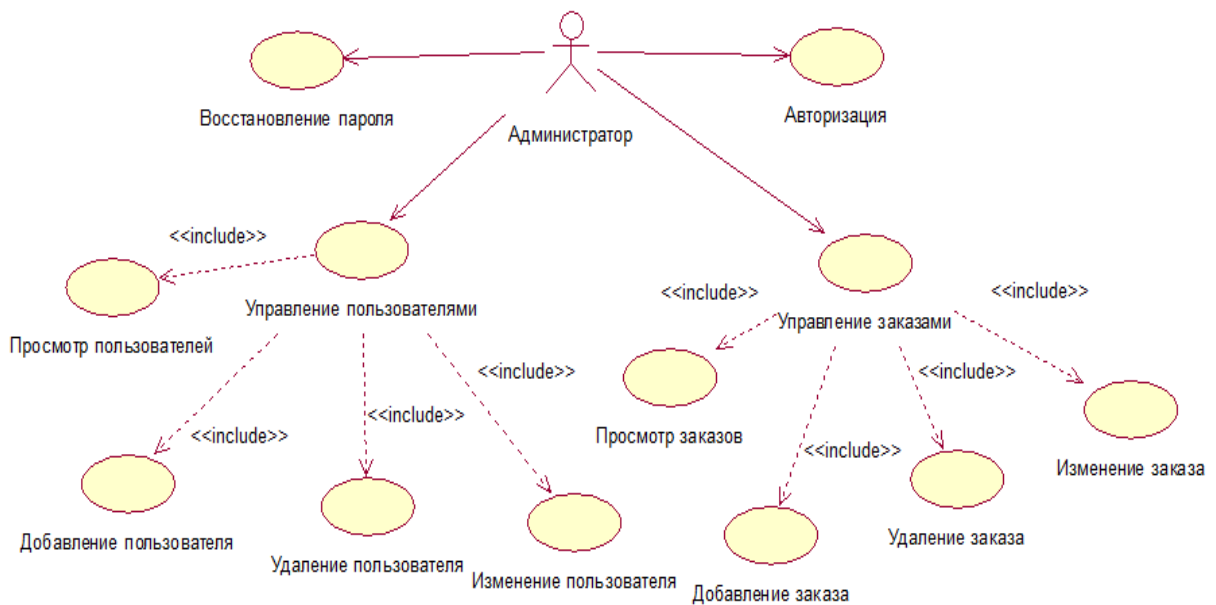


Рисунок 2.4 – Функції адміністратора

Для успішного функціонування майбутньої системи необхідно спроектувати і розробити такі компоненти [9]:

- 1) базу даних, необхідну для зберігання всієї інформації, що циркулює в системі;
- 2) адміністративну панель, що дозволяє реалізувати основні функції адміністратора сайту;
- 3) сайт, що дозволяє користувачам використовувати систему в своїх потребах (як виконавців, так і замовників);
- 4) API, необхідне для мобільної версії сайту (для замовників).

3 КОНТЕНТ-АНАЛІЗ ЯК МЕТОД АНАЛІЗУ МЕДІАКОНТЕНТУ

3.1 Контент-аналіз, переваги і недоліки методу

Контент-аналіз - техніка укладання, виробленого завдяки об'єктивному і систематичному виявленню відповідних завданням дослідження характеристик тексту. Процедура контент-аналізу - підрахунок частоти (або обсягу) згадок тих чи інших смислових одиниць аналізованого тексту. Контент-аналіз використовують для вивчення чималої кількості однотипних документів, релевантних темі дослідження. Він є цільовим (досліджується вузька сторона документів) і дає неповний, екстенсивний опис документів. Контент-аналіз використовують: при необхідності високої точності або об'єктивності аналізу; при наявності великого і несистематизованого матеріалу по темі дослідження; коли значущий мову (стиль) документа; коли важливо довести існування прихованого сенсу в документі. Контент-аналіз дозволяє дати точні характеристики досліджуваного об'єкта, з'ясувати причини, що породили повідомлення, оцінити ефект впливу повідомлення [10]. Сама процедура зводиться до двох операцій: 1) формулюються ключові, сутнісні поняття на основі конкретного тексту; 2) фіксується частота згадки і обсяг цих категорій, на підставі зібраних даних кількісні показники піддаються обробці. Отримані результати дозволяють судити і про латентний зміст тексту. Тому контент-аналіз часто розглядається як якісно-кількісний аналіз. Особливо зручно піддавати контент-аналізу матеріали засобів масової інформації. Контент-аналіз може використовуватися як допоміжний метод (наприклад, при аналізі відповідей на відкриті питання), засіб взаємоперевірки і контролю даних, отриманих за допомогою інших методів (опитування, спостереження), один з етапів емпіричного дослідження, головний і єдиний метод. Контент-аналіз був розроблений в США. Перші проби систематичного кількісного аналізу газетних текстів відносять до кінця XIX - початку XX століття, хоча перші відомі випадки використання контент-аналізу мали місце в XVIII в., Коли в Швеції на підставі підрахунку частоти появи теми Христа в тексті приймалося рішення про єретичні тій чи іншій книги. Можна виділити п'ять

історичних етапів становлення контент-аналізу як самостійного методу. Нині контент-аналіз активно використовується для вивчення комунікації в області державних структур, специфіки «подачі» матеріалів різними партійними виданнями, оцінок тих чи інших політичних подій, текстів виступів політичних лідерів (визначення їх психологічних типів, мотивації поведінки), справжнього сенсу інформації, переданої в ЗМІ чуток як найбільш ефективного каналу комунікації.

Б.Берельсон сформулював сімнадцять цілей, які можуть бути досягнуті з допомогою контент-аналізу. Серед них [11]:

- опис тенденцій при зміні змісту комунікативних процесів;
- опис розходжень у змісті комунікативних процесів для різних країн;
- порівняння різних ЗМІ;
- виявлення використовуваних пропагандистських прийомів;
- визначення намірів і інших характеристик учасників комунікації;
- визначення психологічного стану індивідів і/або груп;
- виявлення настанов, інтересів і цінностей (ширше – систем переконань і "моделей світу") різних груп населення і суспільних інститутів;
- виявлення фокусів уваги індивідів, груп і соціальних інститутів.

Дещо пізніше О.Ольсті виділяє сім запитань, на які намагається відповісти кількісний контент-аналіз [12]:

- опис тенденцій (на основі змісту тексту);
- виявлення взаємозв'язків між характеристиками і повідомленням тексту;
- оцінка відповідності стандартам комунікацій;
- аналіз методів переконань (пропаганди);
- аналіз стилю;
- виявлення взаємозв'язків між відомими характеристиками аудиторії і повідомленнями, підготовленими для неї;
- опис зразків (шаблонів) комунікації.

К. Нойсндорф указує на чотири головні ролі контент-аналізу, до яких належить описова, формування висновків, психометрична, прогностична.

Основна претензія до контент-аналізу стосувалася того, що він досліджує явно виражені параметри в тексті, а це істотно звужує пізнавальні можливості методу і фактично зводить його до опису.

Наступне, що закидають контент-аналізу, – це втрата контексту і, як наслідок, принципова неможливість кількісно виміряти роль комунікації в ньому. Статистичний аналіз отриманих частотних таблиць ще більше спрощує ситуацію і віддаляє від реальності. Іншими словами, мова розуміється спрощено: вона уподібнюється словнику, у якому представлені лише слова, ізольовані від контексту. Витягнуті з цілого і проаналізовані окремо, елементи тексту не дають повноцінної картини [10].

Третій аргумент висувається проти положення, згідно якого частота появи понять фіксує певний феномен, ступінь його важливості. Хоча, як правило, вагомі елементи або зразки зустрічаються частіше, проте частота категорії та важливість теми для автора не завжди пов'язані лінійно. Наприклад, при розгляді болючої теми лише її одноразове згадування може дати більше для роздумів, ніж згадування шістнадцять разів порівняно з п'ятнадцятьма.

Але головною силою контент-аналізу є його здатність аналізувати масштабні дані задля вивчення змін, які мають місце в них, і специфіки домінант дискурсу. Уже за своїм змістом не всі види контент-аналізу є об'єктивними, і суб'єктивність з'являється як у процесі кодування та здійснення вибірки, так і читання текстів. Запровадження жорстких правил проведення аналізу може зменшити до певної міри недоліки контент-аналізу.

3.2 Методи контент-аналізу

У науковій літературі зустрічаються різні класифікації видів контент-аналізу [13]. Залежно від характеру гіпотези виділяють пошуковий і контрольний контент-аналіз, розрізняють спрямований і ненаправлений контент-аналіз. Існує навіть вельми специфічне поділ контент-аналізу на безпосередній і непрямий, але частіше в літературі зустрічаються поняття «кількісний» і «якісний» контент-аналіз. На наш

погляд, найкращою є класифікація методів контент-аналізу, розроблена Р. Мертоном, яка включає шість варіантів [14] :

- 1) підрахунок символів (простий підрахунок певних ключових слів);
- 2) класифікацію символів по відношенню (баланс позитивних і негативних висловлювань з приводу об'єкта дослідження); використовується для аналізу ефективного розташування символів для пропаганди, для виявлення контрастних, суперечливих суджень і для визначення намірів комунікатора;
- 3) аналіз за елементами (вибір головних і другорядних частин тексту, визначення тем, пов'язаних з основними і периферійними інтересами аудиторії);
- 4) тематичний аналіз (виявлення явних і прихованих тем);
- 5) структурний аналіз (з'ясування характеру співвідношення різних матеріалів: взаємодоповнююче, об'єднаного, стикаються);
- 6) аналіз взаємини різних матеріалів (поєднання структурного аналізу з вивченням послідовності публікації матеріалів, обсягу і часу виходу їх у світ).

При обробці даних кількісного (частотного) виду контент-аналізу зазвичай фіксують як традиційні відсотки або частотні розподілу, різноманітні коефіцієнти (процедура підрахунку аналогічна стандартним прийомам класифікації по виділеним угрупованням), так і особливі методи кількісної обробки відомостей, отриманих в контент-аналітичному дослідженні. До останніх відносяться, наприклад, формула коефіцієнта Яніса для обчислення співвідношення позитивних і негативних по відношенню до прийнятої позиції суджень, формула оцінки «питомої ваги» смислових категорій в загальному обсязі тексту, методика аналізу залежності елементів змісту для розрахунку спільної виникнення різних елементів в тексті [15].

Методика аналізу залежності елементів змісту для розрахунку спільної виникнення різних елементів в тексті. Процедура цієї методики полягає в тому, що після підрахунку числа спільної зустрічальності одиниць аналізу розраховується квадратна матриця можливих і фактичних спільних появ цих одиниць в тексті.

Структурний контент-аналіз передбачає дослідження того, яким чином подається повідомлення про будь-якому політичному процесі (наприклад, про висвітлення в пресі ходу виборчої кампанії або процесу прийняття в парламенті того

чи іншого законопроекту). Супроводжується дане повідомлення фотографією (підвищує привабливість тексту), який розмір заголовка, надруковано воно на першій шпальті або «поховано» в купі реклами (тобто для нас важливо не зміст, а спосіб презентації матеріалу).

Якісний варіант методики контент-аналізу орієнтується не на кількісний вимір, а враховує співвідношення кількісних і якісних показників, роль попередніх гіпотез зведена до мінімуму, результат аналізу заздалегідь невідомий [14].

Одиниці аналізу в цьому випадку виділяються на підставі критерію «цілісність судження про предмет», а потім в ході аналізу структуруються у вигляді проблемного графа, що включає кілька заданих рівнів: змістовно-інсталяційний (на цьому рівні фіксується назва текстового масиву, підданого аналізу), тематичний (виділяються основні теми інформаційного масиву), проблемний (виділяються проблеми в рамках окремих тем), атрибутивний (вказуються характеристики, ознаки ситуації), цільової (дається інформація про бажане або реальний стан даних об'єктів) і т.д.

Після заповнення матричного класифікатора виробляються квантифікація інформаційного масиву і математична обробка результатів. Даний варіант обробки текстових матеріалів складний і вимагає високої кваліфікації дослідників [15].

3.3 Одиниці контент-аналізу

Одиниці тексту - це окрема група слів, що розглядається як ціле. Виділяється кілька типів одиниць.

Одиниці аналізу - це одиниці, що становлять основу аналізу; одиниці, які дослідник прагне охарактеризувати [12].

Одиниці вибірки - частини спостерігається реальності або потоку інформації, які розглядаються як незалежні один від одного. Вони мають ясно помітні кордону, їм можуть бути присвоєні унікальні номери, і вони можуть включатися до вибірки із заздалегідь відомою ймовірністю.

Одиниці кодування - це окремі сегменти інформації, що поміщаються в ту чи іншу категорію. Для кожної одиниці кодування дослідник приймає рішення, чи має

вона ті чи інші атрибути, які цікавлять його в даному дослідженні, чи належать вони до теми дослідження.

Одиниці контексту - це та сукупність контенту, яку необхідно взяти до уваги, характеризуючи одиницю кодування. Вони формують контекст, який визначає значення, сенс одиниць кодування в тому випадку, якщо цей сенс контекстно-залежний [13].

При аналізі текстів без застосування комп'ютера контекст зазвичай легко розпізнавати. У комп'ютерному аналізі контекст, як правило, визначається через аналіз слів, що оточують в тексті одиницю кодування.

Одиниці рахунку - це ті одиниці, за допомогою яких квантифікуються атрибути тексту. Вони збігаються з одиницями кодування, якщо дослідник зацікавлений в підрахунку кількості слів або інших елементів тексту. Іншими словами, одиниці рахунку - це саме те, що підраховується в процесі дослідження; то, до чого відносяться числа в матриці даних. Приклади: 5 слів були ідентифіковані як такі, що до агресії (що потрапляють в дану категорію). У матрицю ставиться число 5 - в даному випадку одиниця кодування збігається з одиницею рахунку. Приклад розбіжності цих одиниць: аналіз простору на сторінках газети, відданого під освітлення певної теми. Стаття, ідентифікована як відноситься до теми, - це одиниця кодування, а число квадратних сантиметрів (в яких виміряна площа статті і отриманий результат занесений в матрицю) - одиниця рахунку.

Фізичні одиниці мають окрему фізичну форму (наприклад, окремий номер газети).

Синтаксичні одиниці - ті, які є природними для граматики відповідного засобу комунікації (наприклад, слово у фразі або окрема новина у фразі випуску новин).

Одиниці референції - ті, які описують різними словами один і той же об'єкт (наприклад, «глава держави», «президент», в певному контексті - просто «він»).

Пропозиційні одиниці - це частини складних речень, що мають власну структуру, опису конкретних положень справ (ситуацій) [15].

Одиниці різного роду можуть перетинатися і включати один одного. Наприклад, при аналізі книг перша одиниця аналізу - це книга, друга - голови в

книгах, третя - параграфи або абзаци. У разі, якщо параграф - найменша з одиниць, на які дослідник розбив текст, він також служить і одиницею кодування. Однак можна продовжити ділити текст далі аж до пропозицій або граматичних частин пропозицій. В такому випадку одиницею вибірки може стати абзац. Кожна одиниця, яка більше, ніж складові її частини, може служити одиницею контексту: фраза - для слова, глава - для параграфа і т.д.

Крім одиниць контент-аналізу існують і концептуальні категорії. Концепт - це одиниця сенсу, окрема ідея.

3.4 Удосконалений метод аналізу медіаконтенту

На основі розглянутих методів аналізу медіаконтенту було розроблено модифікацію методу, опис якого наведено нижче.

1) На першому етапі потрібно визначити сукупності досліджуваних джерел або повідомлень за допомогою набору заданих критеріїв, яким має відповідати кожне повідомлення [14]:

- а) заданий тип джерела;
- б) один тип повідомлень;
- в) задані боки, які беруть участь в процесі комунікації (відправник, отримувач (реципієнт));
- г) можна порівняти розмір повідомлень;
- д) частота появи повідомлень;
- е) спосіб поширення повідомлень;
- ж) місце поширення повідомлень;
- з) час появи повідомлень.

2) Другий етап полягає у формуванні вибіркової сукупності повідомлень. Іноді можна вивчати всю визначену на першому етапі сукупність джерел, оскільки підлягають аналізу випадки часто обмежені за кількістю і добре доступні. Але також контент-аналіз повинен спиратися на обмежену вибірку, яка взята з більшого масиву інформації.

3) на третьому етапі необхідно виявити одиниці аналізу. Цими одиницями можуть бути слова або теми. Правильний вибір одиниць аналізу - це важлива складова всієї роботи.

Існують досить чіткі вимоги до вибору можливої одиниці аналізу:

- а) вона повинна бути досить великою для того, щоб висловлювати значення;
- б) вона повинна бути досить малою, щоб не висловлювати багато значень;
- в) вона повинна легко ідентифікуватися;
- г) число одиниць має бути настільки великим, щоб з них можна було робити вибірку.

4) Четвертий етап - це виділення одиниць рахунку, які можуть збігатися зі смисловими одиницями або носити специфічний характер.

5) На п'ятому етапі відбувається процедура підрахунку. Вона в загальному вигляді схожа на стандартні прийоми класифікації по виділеним угрупованням. Застосовується складання спеціальних таблиць, спеціальних формул, статистичних розрахунків.

б) Інтерпретація отриманих результатів відповідно до цілей і завдань дослідження. На цьому етапі виявляються і оцінюються такі характеристики текстового матеріалу, які дозволяють робити висновки про те, що хотів підкреслити або приховати автор.

Нижче розглянемо діаграму активності на рисунку 3.1. Ця діаграма показує весь процес, який проходить на сайті користувач для того, щоб метод був застосований. Спочатку користувачу необхідно авторизуватися, оскільки метод аналізу використовується на сайті з фільмами, де авторизований користувач може після перегляду фільму на сторінці з ним відповісти на питання, показані на мал. 4.8. Якщо авторизація пройшла успішно та дані були введені вірно користувач переходить до наступного кроку – вибір фільму для оцінки, де він може обрати фільм для перегляду та його оцінити, якщо дані при авторизації були введені невірно, то користувачу необхідно авторизуватись повторно. Коли користувач обере фільм для проведення оцінки він може взяти участь у проведенні аналізу – відповісти на питання, потім зберегти свої результати.



Рисунок 3.1 - Діаграма активності

4 ПРОГРАМНА РЕАЛІЗАЦІЯ ВДОСКОНАЛЕНОГО МЕТОДУ АНАЛІЗУ МЕДІАКОНТЕНТУ

У розробці програми необхідно розглянути такі аспекти як: використовуване програмне забезпечення, вимоги до додатка, схему взаємодії основних структурних елементів системи, технології, використовувані в додатку, а також реалізацію програми.

4.1 Інструментарій для розробки

Для написання даного сайту була використана об'єктно-орієнтована мова програмування Java, що розробляється компанією Sun Microsystems з 1991 року і офіційно випущений 23 травня 1995 року. Спочатку нову мову програмування називали Oak (James Gosling) і розробляли для побутової електроніки, але згодом вона була перейменована в Java і стала використовуватися для написання аплетів, додатків і серверного програмного забезпечення. Програми на Java можуть бути трансльовані в байт-код, що виконується на віртуальній java-машині (JVM) програмою, обробній байт-код і передавальній інструкції обладнанню, як інтерпретатор, але з тією відмінністю, що байт-код, на відміну від тексту, обробляється значно швидше. Java - це технологія, яку використовують для розробки онлайн-додатків, тобто програм, що запускаються і працюють прямо в вашому браузері [16].

Подібні додатки істотно розширюють можливості інтернету і за допомогою технології Java ми можемо грати в онлайн-ігри, спілкуватися в різних онлайн-чатах, завантажувати фото і відео на різні сайти і багато-багато іншого.

Але для того, щоб скористатися всіма цими перевагами необхідно, щоб на вашому комп'ютері була встановлене середовище Java. Тобто, це спеціальна програма, що дозволяє запускати Java-контент прямо в вашому браузері. Якщо середовище не встановлене на вашому комп'ютері, то багато онлайн-додатки і веб-сайти просто не будуть працювати. Зазвичай в цьому випадку, коли ви заходите на веб-сторінку, яка містить Java-контент, з'являється повідомлення про необхідність завантажити і

встановити на ваш комп'ютер Java.

Також після установки на комп'ютер, Java буде відстежувати появу нових оновлень на офіційному сайті і як тільки такі з'являться, ви отримаєте повідомлення з пропозицією їх встановити.

Так як Java-додатка досить популярні і застосовуються в інтернеті повсюдно, то їх можуть використовувати зловмисники в своїх цілях. Це означає, що варто завжди оновлювати Java на своєму комп'ютері, так як в оновленнях усуваються виявлені вразливості і підвищується безпека і стабільність роботи.

IntelliJ IDEA - інтегроване середовище розробки програмного забезпечення для багатьох мов програмування, зокрема Java, JavaScript, Python, Розроблена компанією JetBrains. Перша версія з'явилася в січні 2001 року і швидко набула популярності як перша середа для Java з широким набором інтегрованих інструментів для рефакторінга, які дозволяли програмістам швидко реорганізувати вихідні тексти програм. Дизайн середовища орієнтований на продуктивність роботи програмістів, дозволяючи сконцентруватися на функціональних завданнях, в той час як IntelliJ IDEA бере на себе виконання рутинних операцій. Починаючи з шостої версії продукту IntelliJ IDEA надає інтегрований інструментарій для розробки графічного інтерфейсу користувача [17]. Серед інших можливостей, середа добре сумісна з багатьма популярними вільними інструментами розробників, такими як CVS, Subversion, Apache Ant, Maven і JUnit. У лютому 2007 року розробники IntelliJ анонсували ранню версію плагіна для підтримки програмування на мові Ruby. Починаючи з версії 9.0, середа доступна в двох редакціях: Community Edition і Ultimate Edition. Community Edition є повністю вільною версією, доступною під ліцензією Apache 2.0, в ній реалізована повна підтримка Groovy, Scala, а також інтеграція з найбільш популярними системами управління версіями. В редакції Ultimate Edition, доступною під комерційною ліцензією, реалізована підтримка Java EE, UML-діаграм, підрахунок покриття коду, а також підтримка інших систем управління версіями, мов і фреймворків.

4.2 Структура реалізації веб-проекту

Цей Maven-проект має певну структуру.

Як бачимо на рисунку 4.1 в корені лежать:

- 1) директорія `idea`, в якій знаходяться налаштування ідеї для поточного проекту;
- 2) директорія `src`, в якій ми створюємо наші вихідні дані.



Рисунок 4.1 -Структура майбутньої програми

За базу даних була використана SQL.SQL (Structured Query Language - Структурована мова запитів) - мова управління базами даних для реляційних баз даних. Сам по собі SQL не є повною мовою програмування, але його стандарт дозволяє створювати для нього процедурні розширення, які розширюють його функціональність до повноцінної мови програмування. Мова була створена в 1970-х роках під назвою "SEQUEL" для системи управління базами даних (СКБД) System R. Пізніше він був перейменований в "SQL", щоб уникнути конфлікту торгових марок. У 1979 році SQL був вперше опублікований у вигляді комерційного продукту

Oracle V2. Перший офіційний стандарт мови був прийнятий ANSI в 1986 році і ISO - в 1987. З тих пір були створені ще кілька версій стандарту, деякі з них повторювали попередні з незначними варіаціями, інші брали нові суттєві риси. Незважаючи на існування стандартів, більшість поширених реалізацій SQL відрізняються так сильно, що код рідко може бути перенесений з однієї СУБД в іншу без внесення істотних змін. Це пояснюється великим обсягом і складністю стандарту, а також браком в ньому специфікацій в деяких важливих областях реалізації. SQL створювався як простий стандартизований спосіб вилучення і управління даними, що містяться в реляційній базі даних. Пізніше він став складніше, ніж замислювався, і перетворився в інструмент розробника, а не кінцевого користувача [18]. В даний час SQL (здебільшого в реалізації Oracle) залишається найпопулярнішим з мов управління базами даних, хоча і існує ряд альтернатив. що код рідко може бути перенесений з однієї СУБД в іншу без внесення істотних змін. Це пояснювалось великим обсягом і складністю стандарту, а також браком в ньому специфікацій в деяких важливих областях реалізації. SQL створювався як простий стандартизований спосіб вилучення і управління даними, що містяться в реляційній базі даних. Пізніше він став складніше, ніж замислювався, і перетворився в інструмент розробника, а не кінцевого користувача. В даний час SQL (здебільшого в реалізації Oracle) залишається найпопулярнішим з мов управління базами даних, хоча і існує ряд альтернатив. що код рідко можна перенести з однієї СУБД в іншу без внесення істотних змін. Це пояснюється великим обсягом і складністю стандарту, а також браком в ньому специфікацій в деяких важливих областях реалізації. SQL створювався як простий стандартизований спосіб вилучення і управління даними, що містяться в реляційній базі даних. Пізніше він став складніше, ніж замислювався, і перетворився в інструмент розробника, а не кінцевого користувача. В даний час SQL (здебільшого в реалізації Oracle) залишається найпопулярнішим з мов управління базами даних, хоча і існує ряд альтернатив. SQL створювався як простий стандартизований спосіб вилучення і управління даними, що містяться в реляційній базі даних. Пізніше він став складніше, ніж замислювався, і перетворився в інструмент розробника, а не кінцевого користувача. В даний час SQL (здебільшого в реалізації Oracle)

залишається найпопулярнішим з мов управління базами даних, хоча і існує ряд альтернатив. SQL створювався як простий стандартизований спосіб вилучення і управління даними, що містяться в реляційній базі даних. Пізніше він став складніше, ніж замислювався, і перетворився в інструмент розробника, а не кінцевого користувача. В даний час SQL (здебільшого в реалізації Oracle) залишається найпопулярнішим з мов управління базами даних, хоча і існує ряд альтернатив [18].

SQL складається з чотирьох окремих частин:

1) мова визначення даних (DDL) використовується для визначення структур даних, що зберігаються в базі даних. Оператори DDL можна створювати, редагувати і видаляти окремі об'єкти в БД. Допустимі типи об'єктів залежать від використовуваної СУБД і зазвичай включають бази даних, користувачів, таблиці і ряд дрібніших допоміжних об'єктів, наприклад, ролі і індекси;

2) мова маніпуляції даними (DML) використовується для вилучення і зміни даних в БД. Оператори DML дозволяють витягати, вставляти, змінювати і видаляти дані в таблицях. Іноді оператори select вилучення даних не розглядаються як частина DML, оскільки вони не змінюють стан даних. Всі оператори DML носять декларативний характер;

3) мова визначення доступу до даних (DCL) використовується для контролю доступу до даних в БД. Оператори DCL застосовуються до привілеїв і дозволяють видавати і відбирати права на застосування певних операторів DDL і DML до певних об'єктів БД;

4) мова управління транзакціями (TCL) використовується для контролю обробки транзакцій в БД. Зазвичай оператори TCL включають commit для підтвердження змін, зроблених в ході транзакції, rollback для їх скасування та savepoint для розбиття транзакції на кілька менших частин.

Слід зазначити, що SQL реалізує декларативну парадигму програмування: кожен оператор описує тільки необхідну дію, а СУБД приймає рішення про те, як його виконати, тобто планує елементарні операції, необхідні для виконання дії і виконує їх. Проте, для ефективного використання можливостей SQL розробнику

необхідно розуміти те, як СУБД аналізує кожен оператор і створює його план виконання.

4.3 Технічні особливості побудови веб-додатків

Суттєва перевага побудови веб-додатків для підтримки стандартних функцій браузера полягає в тому, що функції повинні виконуватися незалежно від операційної системи даного клієнта. Замість того, щоб писати різні версії для Microsoft Windows, Mac OS X, GNU / Linux та інших операційних систем, додаток створюється один раз для довільно обраної платформи і на ній розгортається.

Однак різна реалізація HTML, CSS, DOM і інших специфікацій в браузерах може викликати проблеми при розробці веб-додатків і подальшої підтримки. Крім того, можливість користувача налаштовувати багато параметрів браузера (наприклад, розмір шрифту, кольору, відключення підтримки сценаріїв) може перешкоджати коректної роботи програми [8].

Інший (менш універсальний) підхід полягає у використанні Adobe Flash, Silverlight або Java - аплетів для повної або часткової реалізації призначеного для користувача інтерфейсу.

Оскільки більшість браузерів підтримує ці технології (як правило, за допомогою плагінів), Flash- або Java - додатки можуть виконуватися з легкістю. Так як вони надають програмісту більший контроль над інтерфейсом, вони здатні обходити багато несумісності в конфігураціях браузерів, хоча несумісність між Java- або Flash - реалізаціями на стороні клієнта може призводити до різних ускладнень.

У зв'язку з архітектурним схожістю з традиційними клієнт-серверними додатками, в деякому роді «товстими» і клієнтами, існують суперечки щодо коректності віднесення подібних систем до веб-додатків.

4.3.1 Структура веб-додатків

Веб-додатки являють собою особливий тип програм, побудованих за

архітектурою "клієнт-сервер". Особливість їх полягає в тому, що саме веб-додаток знаходиться і виконується на сервері - клієнт при цьому отримує тільки результати роботи. Робота програми ґрунтується на отриманні запитів від користувача (клієнта), їх обробці та видачі результату. Передача запитів і результатів їх обробки відбувається через протокол HTTP [9].

Відображенням результатів запитів, а також прийомом даних від клієнта і як вони будуть передані на сервер зазвичай займається спеціальний додаток - браузер (Internet Explorer, Mozilla, Opera і т. Д.). Як відомо, однією з функцій браузера є відображення даних, отриманих з Інтернету, у вигляді сторінки, описаної на мові HTML, отже, результат, який передається сервером клієнтові, повинен бути представлений на цій мові.

На стороні сервера веб-додаток виконується спеціальним програмним забезпеченням (веб-сервером), який і приймає запити клієнтів, обробляє їх, формує відповідь у вигляді сторінки, описаної на мові HTML, і передає його клієнту.

В процесі обробки запиту користувача веб-додаток компонує відповідь на основі виконання програмного коду, що працює на стороні сервера, веб-форми, сторінки HTML, інший вміст, включаючи графічні файли. В результаті, як вже було сказано, формується HTML-сторінка, яка і відправляється клієнту. Виходить, що результат роботи веб-додатки ідентичний результату запиту до традиційного веб-сайту, проте, на відміну від нього, веб-додаток генерує HTML-код в залежності від запиту користувача, а не просто передає його клієнту в тому вигляді, в якому цей код зберігається в файлі на стороні сервера. Тобто веб-додаток динамічно формує відповідь за допомогою коду, - так званої виконуваної частини.

Протокол HTTP (Hypertext Transfer Protocol, протокол передачі гіпертекстів) - це мережевий протокол прикладного рівня для розподілених гіпертекстових інформаційних систем.

Цей протокол визначає способи передачі даних між веб-браузером і веб-сервером. основною областю його використання є World Wide Web [7].

HTTP є протоколом типу запит / відгук. Клієнт посилає серверу запит, що починається з рядка, що містить тип запиту (метод), URI необхідного ресурсу і

номера версії протоколу.

Сервер відповідає рядком стану, що містить версію протоколу і код стану, за якою слідує набір заголовків і, можливо, тіло повідомлення, як і в разі запиту відокремлене від заголовків порожнім рядком.

Визначено такі методи - GET, POST, HEAD, PUT, DELETE, TRACE і OPTION, причому найбільш часто використовуються прикладними програмами перші два.

Метод GET використовується для запиту конкретного ресурсу.

Всі параметри запиту передаються в URI запитаного ресурсу у вигляді пар <ключ> = <значення>, поділюваних знаком '&' і відокремлених від імені самого ресурсу знаком '?'.

Метод POST використовується для передачі значних обсягів даних сервера в тілі запиту, так як довжина URI обмежена.

Метод HEAD аналогічний методу GET за винятком того, що у відповіді на нього міститься тільки набір заголовків, а сам ресурс не передається (тіло відповіді відсутній).

Методи PUT і DELETE призначені для зміни ресурсів засобами HTTP, а методи TRACE і OPTION є службовими [9].

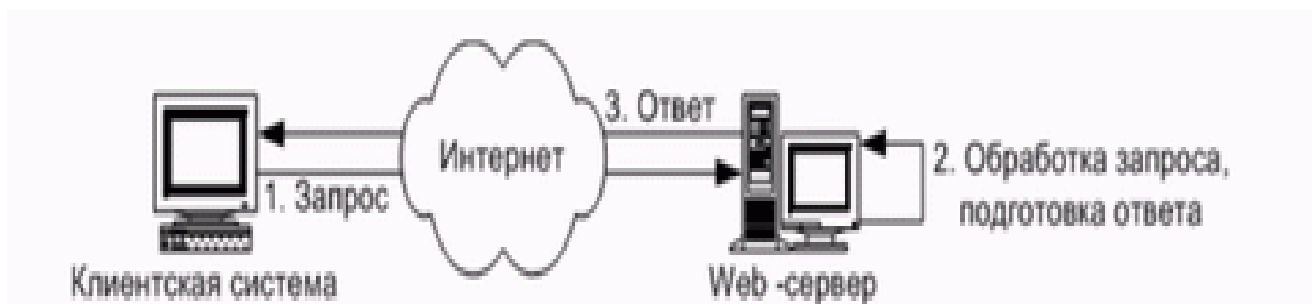


Рисунок 4.2 - Взаємодія сервера і клієнта

Як видно з рисунка 4.2, робота серверних додатків відбувається в три основних етапи:

- 1) запит. Клієнт, використовуючи веб - браузер, ініціює запит до сервера;
- 2) обробка запиту, підготовка відповіді. Після отримання запиту веб - сервер проводить обробку запитуваного ресурсу. У разі, якщо запитується статичний ресурс,

такий як HTML сторінка, малюнок, документ, ця інформація форматується для протоколу HTTP і передається клієнтові як відповідь. Якщо ж запитується динамічний ресурс, запит передається на обробку відповідного контейнеру веб-додатків, де і відбувається подальша робота;

3) після формування, дані передаються клієнту за допомогою протоколу HTTP в якості відповіді. Відповідь містить дані (зазвичай HTML код, або двійкові дані), а також додаткові параметри, що передаються в заголовках HTTP відповіді.

Робота додатків серверної сторони завжди відбувається за описаним вище сценарієм. Очевидно, що такий підхід створює складності при створенні веб-додатків, основною з яких є відсутність стану у веб-додатки (так зване *stateless programming*). Це означає, що додаток працює виключно в режимі запит-відповідь, не маючи даних про попередні кроки користувача або будь-якої іншої постійної інформації. Для вирішення цієї проблеми застосовується поняття користувальницької сесії, яка дозволяє зберігати дані на сервері протягом сеансу роботи користувача [8].

Однак наявністю сесій складності при створенні веб-додатків повністю не усуваються. Чим більше можливостей надає платформа реалізації для додатків серверної сторони в подоланні цих труднощів, тим швидше і ефективніше може вестися розробка.

4.4 Графічний інтерфейс

В ході реалізації поставленого завдання були використані результати бакалаврського диплому для аналізу медіаконтенту за допомогою методу контент-аналізу.

На головній сторінці ми спостерігаємо наступні функціональні кнопки: Головна, ТОП 100, Анонси, Вхід / Реєстрація, як можна побачити на рисунку 4.3.

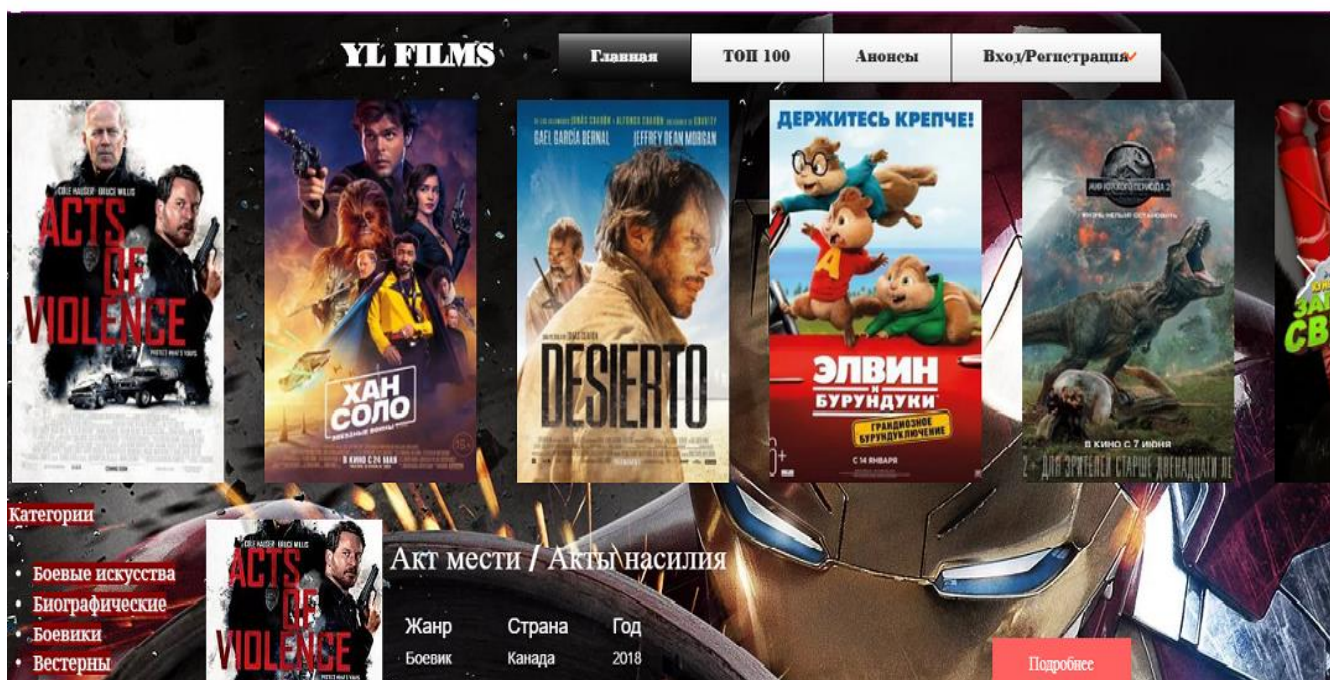


Рисунок 4.3 - Головна сторінка сайту

Також вгорі знаходяться фільми, які були рекомендовані глядачем.

Коли ми натискаємо кнопку «ТОП 100», що представлено на рисунку 4.4, ми переходимо на іншу мандрівницю, де ми бачимо топ фільмів, які увійшли у всесвітній рейтинг фільмів.

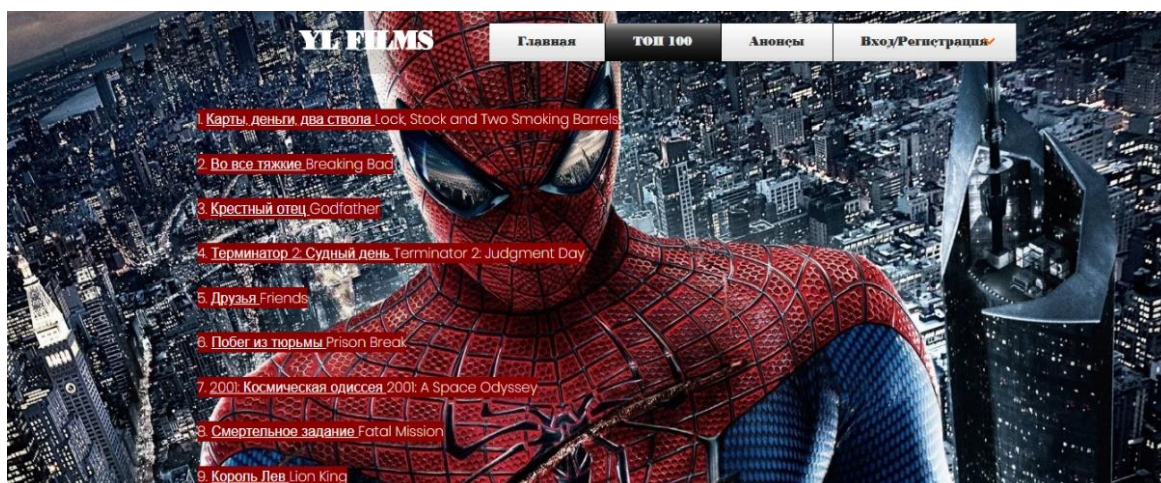


Рисунок 4.4 - Сторінка «ТОП 100»

Коли ми натискаємо кнопку «Анонсы», що зображено на рисунку 4.5, ми переходимо на сторінку з фільмами які виходять в прокат.

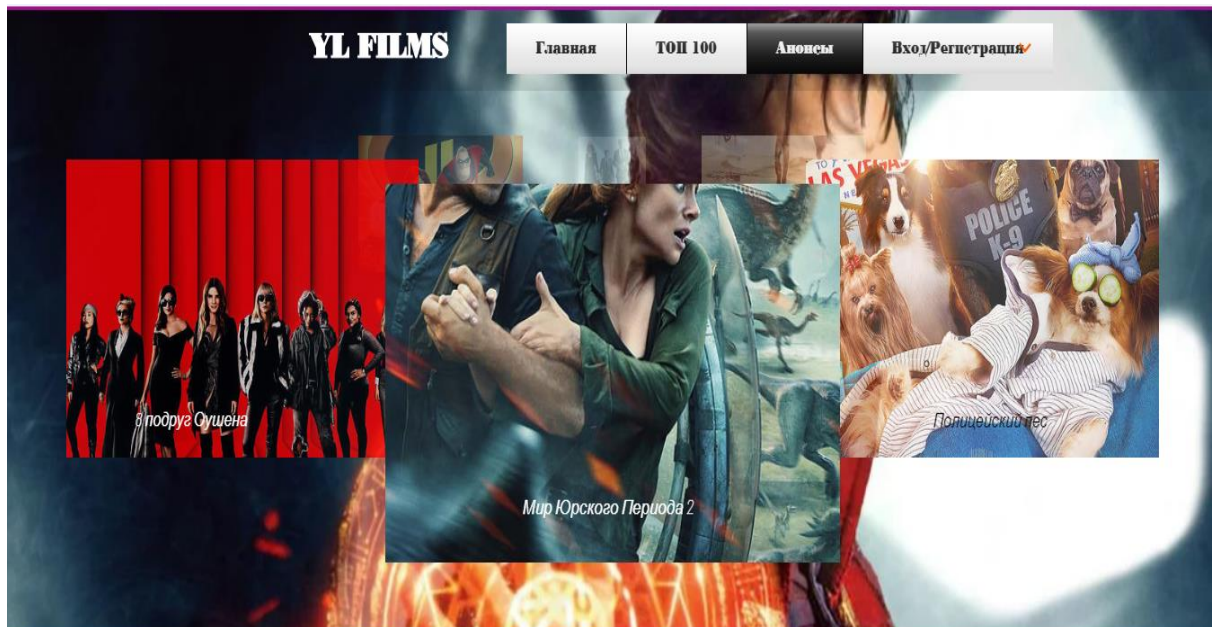


Рисунок 4.5 - Сторінка «Анонси»

Коли ми натискаємо кнопку «Вхід / Реєстрація», ми переходимо на сторінку реєстрації, як можна побачити з рисунку 4.6.

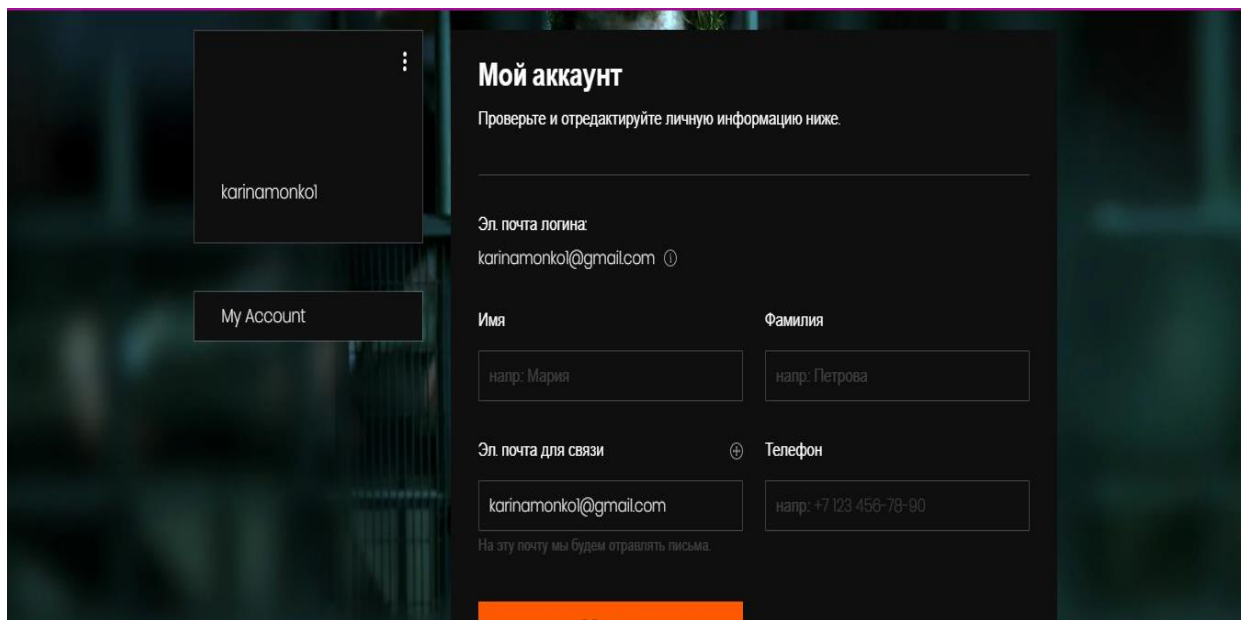


Рисунок 4.6 - Сторінка для входу та реєстрації

Також є можливість задати питання адміністратору. Для цього потрібно пройти на головну сторінку і спуститися вниз. Цей функціонал представлений на рисунку 4.7.

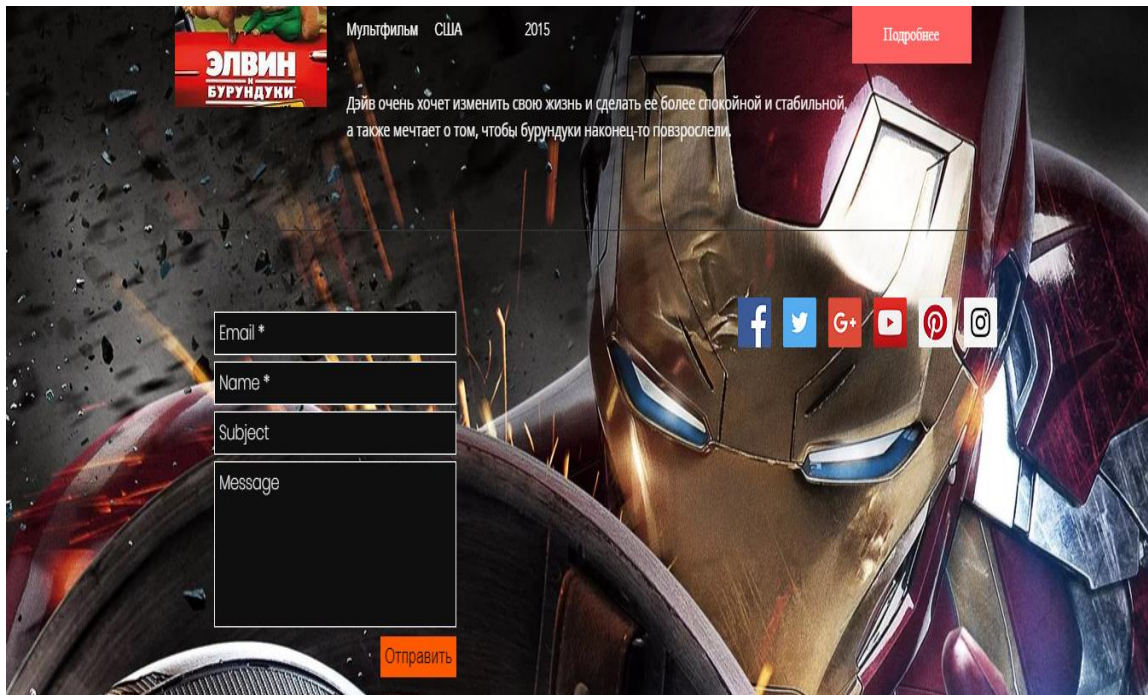


Рисунок 4.7 - Функціонал для зв'язку з адміністратором сайту

На рисунку 4.8 ми бачимо поле, яке знаходиться на сторінці кожного фільму для того, щоб ми могли отримати фідбек від користувача та використати метод контент-аналізу для аналізу медіаконтенту.

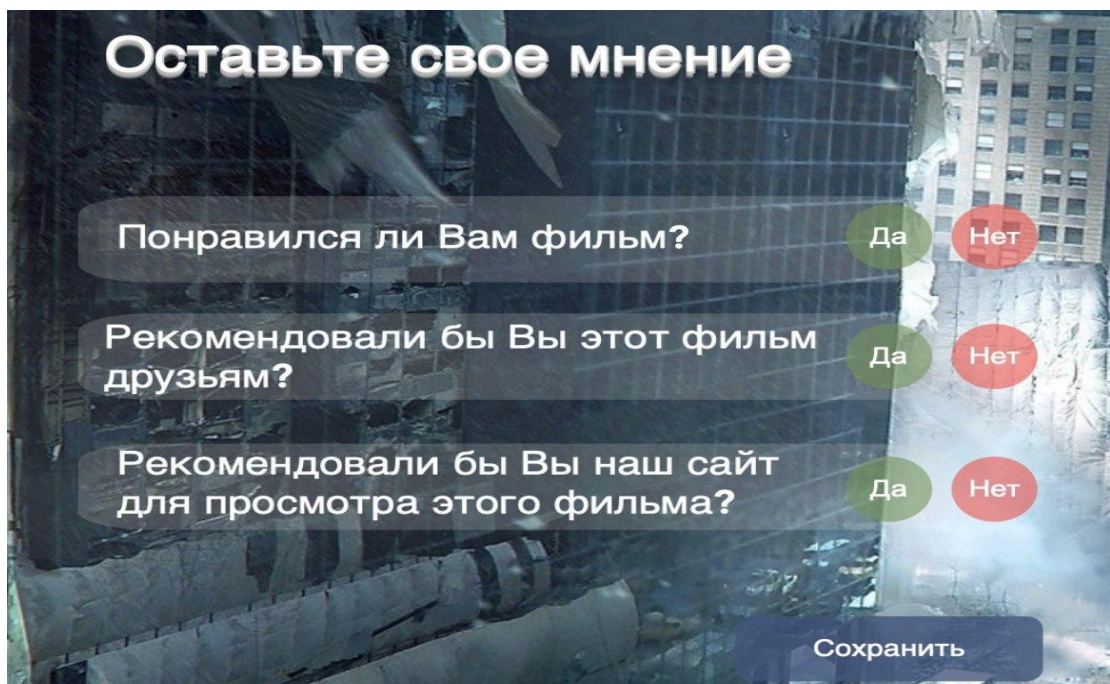


Рисунок 4.8 – Використання методу контент-аналізу

За допомогою Java був успішно створений веб - сервіс з рекомендаційної системою для задоволення потреб користувачів сайту і для того, щоб спростити і прискорити вибір користувачами фільму, шляхом надання їм рекомендацій. Також було додано функціонал аналізу лояльності користувачів до сервісу за допомогою проведення контент-аналізу медіаконтенту.

ВИСНОВКИ

У сучасному світі кожен користувач стикається з величезною кількістю інформації і він не здатен пропустити її всю через себе в пошуках того, що йому подобається. Це ускладнює вибір і робить цей процес більш тривалим. Спростити цю задачу покликані інформаційно-аналітичні системи. Зараз більшість масштабних комерційних і соціальних сайтів використовують механізми, які сортують значні обсяги даних для визначення потенційних переваг користувачів.

При виконанні магістерської роботи був вдосконалений метод аналізу медіаконтенту та розроблена його програмна реалізація на створеному сайті. Основна перевага методу контент-аналізу, який було модифіковано в даній роботі, в тому, що його можна застосовувати в умовах, при яких користувач може самостійно ввести інформацію, необхідну для аналізу і, отже, відсутня необхідність тривалого її збору.

Ще одна перевага даного підходу в тому, що він може бути застосован в іншій предметній області. Для цього необхідно лише замінити або змінити відповідний контент. Це дозволить аналізувати різний контент.

Для реалізації мети атестаційної роботи, а саме вдосконалення методу аналізу медіаконтенту та його реалізації на основі веб-додатку були вирішені наступні завдання:

- був проведений аналіз предметної області;
- проаналізовано існуючі методи аналізу медіаконтенту та характеристики інформаційно-аналітичних систем, де вони реалізовані;
- був вдосконалений метод аналізу медіаконтенту та розроблена його програмна реалізація на створеному сайті.

СПИСОК ЛІТЕРАТУРИ:

- 1) Медіа-контент [Електронний ресурс] - Режим доступу: <http://mp3fun.ru/news/300-media-kontent>
- 2) Види та методи аналізу медіаконтенту [Електронний ресурс] - Режим доступу: <http://www.dslib.net/zhurnalistika/informacionnyj-mediakontent-v-internete-sovremennaja-spezifika-i-kljuचेvye.html>
- 3) Приклади рекомендаційних систем [Електронний ресурс] - Режим доступу: <http://www.hr-portal.ru/>- Загл. з екрану.
- 4) Рекомендаційні системи [Електронний ресурс] - Режим доступу: <https://logic.pdmi.ras.ru/>
- 5) Жернакова О. Інформаційно-аналітичні системи Телемультимедіа, 2012. – 324 с.
- 6) Інформаційно-аналітичні системи [Електронний ресурс] - Режим доступу: <https://ru.wikipedia.org/wiki/>- Загл. з екрану.
- 7) Загальні поняття про веб-додатках [Електронний ресурс] - Режим доступу: <http://daхnow.narod.ru> - Загл. з екрану.
- 8) Технології веб-додатків [Електронний ресурс] - Режим доступу: <http://www.intuit.ru/>- Загл. з екрану.
- 9) Всі про веб-системах і веб-додатках [Електронний ресурс] - Режим доступу: <https://iit.com.ua>- Загл. з екрану.
- 10) Daniel Lemire, Anna Maclachlan. Content-Analysis, In SIAM Data Mining (SDM'05), Newport Beach, California, April 21-23, 2005. (англ.)
- 11) Воронцов, К. В. Поняття контент-аналізу Збірник матеріалів семінару [Текст] /Б.Г.Міркина, М: ВШЕ, 2008. - 54 с.
- 12) Деньжаков А.Ю., Шибанов С.В., Хмелівської Б.Г. «Контент-аналіз», 2015.- 257 с.
- 13) Melville, P. Content-Boosted Content-Analysis types [Text] / P. Melville, R. Mooney, R. Nagarajan University of Texas, USA .: AAAI-02, Austin, TX, USA, 2002. - P. 187 .

- 14) P. Melville, V. Sindhwani Content analysis today. Encyclopedia of Machine Learning. 2010. – 173 p.
- 15) F. Ricci, L. Rokach, B. Shapira, PB Kantor How to analyse? Handbook. Springer, 2011. 842 P.
- 16) Перрі, Б. Java сервлети і JSP: [Текст]: пер. з англ. - М.: Кудіц-Пресс, 2006. - 768 с
- 17) Шілдт, Г. Java. керівництво [Текст]: 8-е вид.: пер. з англ. - М.: ТОВ «І. Д. Вільямс», 2012. - 1104 с.
- 18) Снарський, А.А. Моделювання складних сітей [Текст]: навчальний посібник / А. А. Снарський, Д. В. Ланде. - К.: НТУУ «КПІ», 2015. - 212 с.