

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук  
(повна назва)

Кафедра Системотехніки  
(повна назва)

## КВАЛІФІКАЦІЙНА РОБОТА

### Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження особливостей використання методів визначення рекомендацій в  
системах електронної комерції з обмеженим асортиментом товарів

(тема)

Виконала:

студентка 2 курсу, групи ІТІМ-22-1

Дмитренко А. О.

(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки

(код і повна назва спеціальності)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма: Інформаційні технології проектування

(повна назва спеціалізації)

Керівник доц. Коваленко А.І

(прізвище, ініціали)

Допускається до захисту

Зав. кафедри СТ

(підпис)

проф. Гребеннік І.В.

(прізвище, ініціали)

2024 р.

Я, як студент ХНУРЕ, розумію і підтримую політику закладу із академічної доброчесності. Я не надавала і не одержувала недозволену допомогу під час підготовки кваліфікаційної роботи. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

16.01.2024



Дмитренко А. О.

Атестаційна робота не містить відомостей заборонених до відкритого опублікування.

Атестаційна робота виконана у відповідності до стандартів, що діють в Україні.

Попередній захист проведений «15» січня 2024 р.

Керівник кваліфікаційної роботи

доц. Коваленко А.І.

Харківський національний університет радіоелектроніки

Факультет \_\_\_\_\_ *Комп'ютерних наук* \_\_\_\_\_

Кафедра \_\_\_\_\_ *Системотехніки* \_\_\_\_\_

Рівень вищої освіти \_\_\_\_\_ *другий (магістерський)* \_\_\_\_\_

Спеціальність \_\_\_\_\_ *122 – Комп'ютерні науки* \_\_\_\_\_

Тип програми \_\_\_\_\_ *освітньо-професійна* \_\_\_\_\_

Освітня програма: \_\_\_\_\_ *Інформаційні технології проектування* \_\_\_\_\_

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_ *СТ* \_\_\_\_\_

\_\_\_\_\_ *проф. Гребеннік І.В.* \_\_\_\_\_

" \_\_\_\_\_ " \_\_\_\_\_ *2024 р.*

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентці \_\_\_\_\_ *Дмитренко Ангеліні Олександрівні* \_\_\_\_\_  
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження особливостей використання методів визначення рекомендацій в системах електронної комерції з обмеженим асортиментом товарів

\_\_\_\_\_ затверджена наказом університету від 20.11.2023 р. № 1373СТ.

2. Термін подання студентом роботи до екзаменаційної комісії 16.01.2024р.

**3. Вихідні дані до роботи:** Провести дослідження рекомендаційних методів у системах електронної комерції з обмеженим асортиментом товарів. Перелік використовуваних програмних засобів: серверна частина повинна являти собою базу даних, розробленої для платформи СУБД MySQL Server. Клієнтська частина повинна являти собою веб-інтерфейс доступу до БД з використанням технології Django. Операційна система – Windows 10 або вище, інтегроване середовище розробки – PyCharm 2021.3.2 або вище, мова програмування – Python.

**4. Перелік питань, що потрібно опрацювати в роботі:** 4.1 Огляд і аналіз сучасного стану реалізації рекомендаційних функцій e-систем. 4.2 Постановка задачі. 4.3 Теоретичні підходи (методи) для визначення рекомендацій в системах електронної комерції. 4.4 Методи визначення рекомендацій на основі контенту (Content Based). 4.5 Подання даних про об'єкт. 4.6 Зважування даних TF-IDF. 4.7 Методи визначення рекомендацій за спільною фільтрацією (Collaborative Filtering). 4.8 Кластеризація. 4.9 Методи обчислення подібності (Similarity Computation). 4.10 Методи визначення рекомендацій на основі знань (Knowledge-Based). 4.11 Використання жорстких обмежень. 4.12 Вибір близьких об'єктів. 4.13 Висновки за другим розділом. 4.14 Вибір предметної області для проведення експериментальних досліджень. 4.15 Вибір архітектури для реалізації системи електронної комерції. 4.16 Розробка бази даних системи електронної комерції для досліджень квітів. 4.17 Дослідження варіантів реалізації методів на основі контенту (Content-Based). 4.18 Дослідження варіантів реалізації методів, основаних на знаннях (Knowledge-Based). 4.19 Дослідження варіантів реалізації методів спільної фільтрації (Collaborative Filtering). 4.20 Рекомендації до реалізації методів визначення рекомендацій у системах електронної комерції з обмеженим асортиментом

**5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій:** 1. Класифікація рекомендаційних методів. 2. Аналіз предметної області (фактор № 1). 3. Аналіз предметної області (фактори № 2,3). 4. Архітектура е-системи. 5. EER-модель даних е-системи. 6. Головна сторінка з рекомендаціями для клієнта. 7. Інтерфейс каталогу з обраним фільтром. 8. Інтерфейс користувача – форма для відгуку. 9. Інтерфейс детальної інформації про товар з оцінкою і відгуком. 10. EER-модель даних удосконаленої е-системи.

**6. Консультанти розділів роботи:**

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Аналіз методів і підходів	Доц. Коваленко А.І.		17.10.2023
Дослідження варіантів реалізації методів	Доц. Коваленко А.І.		16.11.2023

**КАЛЕНДАРНИЙ ПЛАН**

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання на виконання роботи	16.10.2023	Виконано
2	Аналіз предметної області	17.10.2023-30.10.2023	Виконано
3	Огляд літератури	30.10.2023-01.11.2023	Виконано
4	Аналіз методів і підходів	02.11.2023-16.11.2023	Виконано
5	Дослідження варіантів реалізації методів	16.11.2023-27.12.2023	Виконано
6	Формування висновків	28.12.2023-02.01.2024	Виконано
7	Оформлення пояснювальної записки	03-11.01.2024	Виконано
8	Представлення роботи на рецензування	13.01.2024	Виконано
9	Підготовка презентації до роботи	13.01.2024-16.01.2024	Виконано
10	Захист роботи	17.01.2024	

Дата видачі завдання 16 жовтня 2023 р.

Студент \_\_\_\_\_  
(підпис)

Дмитренко А. О.

Керівник роботи \_\_\_\_\_  
(підпис)

доцент Коваленко А.І.

## РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи магістра: 81 с., 9 табл., 24 рис., 2 дод., 22 джерел.

Е-СИСТЕМА, РЕКОМЕНДАЦІЙНІ СИСТЕМИ, ОБМЕЖЕНИЙ АСОРТИМЕНТ ТОВАРІВ, СПІЛЬНА ФІЛЬТРАЦІЯ, ФІЛЬТРАЦІЯ НА ОСНОВІ ЗНАНЬ, ФІЛЬТРАЦІЯ НА ОСНОВІ КОНТЕНТУ

Об'єкт дослідження – процес електронної комерції, що реалізується за допомогою інформаційних систем з реалізованими рекомендаційними функціями.

Предмет дослідження – інформаційні технології проектування та реалізації методів визначення рекомендацій у системах електронної комерції з обмеженим асортиментом товарів.

Мета дослідження – розглянути особливості практичної реалізації методів визначення рекомендацій в е-системах з обмеженим асортиментом товарів.

Методи дослідження: системний підхід; методи отримання даних для підготовки рекомендацій: за пам'яттю (Memory-Based), за моделлю (Model-Based); рекомендаційні методи, що засновані на описі характеристик товару та даних користувачів (Content-Based); рекомендаційні методи, що засновані на знаннях (Knowledge-Based); рекомендаційні методи, що засновані на сумісній (колаборативній) фільтрації (Collaborative Filtering): метод порівняння користувачів (User-Based), метод порівняння елементів (Item-Based), гібридні методи (Hybrid Filtering).

У роботі розглянути особливості практичної реалізації методів визначення рекомендацій в е-системах з обмеженим асортиментом товарів. Для дослідження обрана й розглядається рекомендаційна система електронної комерції з продажу квітів.

Галузь застосування – системи електронної комерції з обмеженим асортиментом товарів.

## ABSTRACT

Explanatory note contains: 81 p., 24 fig., 9 tabl., 22 sources.

E-SYSTEM, RECOMMENDER SYSTEMS, LIMITED PRODUCT ASSORTMENT, SHARE FILTERING, KNOWLEDGE BASED FILTERING, CONTENT BASED FILTERING

The object of the study is the process of electronic commerce, which is implemented with the help of information systems with implemented recommendation functions.

The subject of the research is information technologies for the design and implementation of methods for determining recommendations in electronic commerce systems with a limited assortment of goods.

Research methods: systematic approach; methods of obtaining data for preparing recommendations: by memory (Memory-Based), by model (Model-Based); recommendation methods based on the description of product characteristics and user data (Content-Based); recommendation methods based on knowledge (Knowledge-Based); recommendation methods based on collaborative (collaborative) filtering (Collaborative Filtering): method of comparing users (User-Based), method of comparing items (Item-Based), hybrid methods (Hybrid Filtering).

The paper examines the peculiarities of the practical implementation of methods for determining recommendations in e-systems with a limited assortment of goods. A recommender system for e-commerce selling flowers is selected and considered for the study.

The field of application is e-commerce systems with a limited range of products.

## ЗМІСТ

ВСТУП.....	8
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ДОСЛІДЖЕНЬ ТА ПОСТАНОВКА ЗАДАЧІ .....	10
1.1 Огляд і аналіз сучасного стану реалізації рекомендаційних функцій в системах електронної комерції .....	10
1.2 Постановка задачі.....	15
2 АНАЛІЗ МЕТОДІВ ДЛЯ ВИЗНАЧЕННЯ РЕКОМЕНДАЦІЙ У СИСТЕМАХ ЕЛЕКТРОННОЇ КОМЕРЦІЇ.....	17
2.1 Теоретичні підходи (методи) для визначення рекомендацій в системах електронної комерції.....	17
2.2 Методи визначення рекомендацій на основі контенту (Content Based).....	22
2.2.1 Подання даних про об'єкт .....	23
2.2.2. Зважування даних TF-IDF .....	23
2.3 Методи визначення рекомендацій за спільною фільтрацією (Collaborative Filtering).....	25
2.3.1. Кластеризація .....	27
2.3.2 Методи обчислення подібності (Similarity Computation).....	28
2.3.3 Альтернативні алгоритми та покращення (метод порівняння елементів)	31
2.4. Методи визначення рекомендацій на основі знань (Knowledge-Based) .....	31
2.4.1 Використання жорстких обмежень .....	32
2.4.2 Вибір близьких об'єктів.....	33
2.5 Висновки за другим розділом .....	34
3 АНАЛІЗ ОСОБЛИВОСТЕЙ ВИКОРИСТАННЯ МЕТОДІВ ВИЗНАЧЕННЯ РЕКОМЕНДАЦІЙ У СИСТЕМАХ Е- КОМЕРЦІЇ З ОБМЕЖЕНИМ АСОРТИМЕНТОМ ТОВАРІВ .....	36
3.1 Вибір предметної області для проведення експериментальних досліджень ..	36
3.2 Вибір архітектури для реалізації системи електронної комерції .....	38
3.3 Розробка бази даних системи електронної комерції для досліджень квітів ...	41
3.4 Дослідження варіантів реалізації методів на основі контенту (Content-Based).....	47
3.5 Дослідження варіантів реалізації методів, оснований на знаннях (Knowledge-Based).....	57
3.6 Дослідження варіантів реалізації методів спільної фільтрації (Collaborative Filtering) .....	62
3.7 Рекомендації до реалізації методів визначення рекомендацій у системах електронної комерції з обмеженим асортиментом .....	69
ВИСНОВКИ.....	74
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....	76
Додаток А Графічний матеріал кваліфікаційної роботи.....	79
Додаток Б Сертифікат учасника конференції .....	84

## ВСТУП

У мирний час в Україні, як і у всьому світі, стрімко зростає попит та пропозиція на системи електронної комерції. Україна – молода країна, яка переживає важкі економічні часи. Але темпи росту е-комерції у країн, що розвиваються, набагато вищі, ніж у розвинених. Визначено, що електронна комерція в Україні тільки недавно пережила етап зародження і перемістилася на етап зростання [2]. Через війну в країні великі скупчення людей, як зазвичай відбувається в торгових центрах та супермаркетах, є смертельно небезпечними, саме тому альтернатива зробити покупки онлайн із безпечного місця є дуже привабливою для більшості громадян. Зі зростання попиту на системи електронної комерції пропорційно зростає потреба у персоналізації цих систем.

Стандартні системи електронної комерції звичайно виконують свою основну функцію – продаж товарів та послуг, але для успішного розвитку бізнесу цього не достатньо. Покращення стандартної системи електронної комерції для збільшення продажів може бути досягнуто за допомогою різних стратегій і підходів. Ось кілька кроків, які можуть допомогти вам досягти цієї мети.

По-перше, автоматизація обліку продажів є документування й аналіз фінансової інформації, пов'язаної з продажами, з метою визначення прибутку та контролю над фінансовими показниками підприємства, що допоможе зменшити витрати бізнесу.

По-друге, реклама та акційні пропозиції грають важливу роль в успішній електронній комерції. Їх правильне планування та використання може покращити ваші продажі.

По-третє, CRM (Customer Relationship Management) допомагають підприємствам зберігати, організувати та аналізувати дані про клієнтів, а також вдосконалювати спілкування та відносини з ними [7].

По-четверте, рекомендаційні системи покращують користувацький досвід та збільшують продажі. Вони допомагають рекомендувати товари або послуги, які могли б бути цікаві конкретному клієнту на основі їхнього попереднього досвіду та поведінки.

Рекомендаційні системи можуть пропонувати споживачам товари, які доповнюють їхні поточні вибори або збільшують загальний обсяг покупок, що призводить до збільшення середнього чеку та здійсненню додаткових продажів. Це дозволяє оптимізувати маркетингові витрати, спрямовуючи ресурси на більш перспективні товари та цільові аудиторії. У цілому, рекомендаційні системи не тільки забезпечують персоналізацію, але і можуть бути стратегічним інструментом для підвищення ефективності та конкурентоспроможності в галузі електронної комерції.

Аналіз підручників та статей [4–5], з описом методів реалізації рекомендаційних функцій в е-системах, дозволяє зробити висновок, що у більшості з них розглядаються дві, обмежених за призначенням, предметні області – системи перегляду фільмів та системи продажу книжок (або бібліотечні системи). Тому є практичний інтерес з розгляду особливостей реалізації рекомендаційних функцій для систем електронної комерції з обмеженим асортиментом товарів за окремими категоріями, таких як продаж продуктів харчування, квітів, одягу, взуття, побутової техніки (що купується для використання на довгостроковий період) тощо.

Мета дослідження даної роботи – розглянути особливості практичної реалізації методів визначення рекомендацій в е-системах з обмеженим асортиментом товарів.

# 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ДОСЛІДЖЕНЬ ТА ПОСТАНОВКА ЗАДАЧІ

1.1 Огляд і аналіз сучасного стану реалізації рекомендаційних функцій в системах електронної комерції

Рекомендаційні функції застосовуються в різних галузях та сферах діяльності. Ось деякі з основних областей, де вони знаходять своє використання:

– електронна комерція – рекомендаційні функції інтернет-магазинів допомагають пропонувати користувачам товари або послуги, які можуть бути їм до вподоби чи корисними;

– медіа та розваги – використовуються для рекомендацій фільмів, музики, книг, статей, новин, відеоігор тощо. Наприклад, популярні платформи Netflix та Spotify використовують рекомендації для підбору контенту для користувачів;

– соціальні мережі – рекомендаційні функції допомагають знайомити користувачів з новими друзями, сторінками, групами та контентом, які вони можуть знайти цікавими;

– освіта – платформи для навчання та онлайн-курси використовують рекомендаційні функції для рекомендації навчальних матеріалів та курсів;

– фінансові послуги – банки та фінансові установи використовують рекомендаційні функції для рекомендацій фінансових продуктів та інвестицій.

Розглянемо детальніше область електронної комерції. Попит на системи електронної комерції продовжує зростати і залишається високим у сучасному світі. Це викликано рядом факторів, які підкреслюють їх важливість і актуальність:

– зростання онлайн-торгівлі набуває все більшої популярності серед споживачів. Усе більше людей прагнуть здійснювати покупки онлайн, що збільшує попит на системи електронної комерції;

– зручність і доступність систем електронної комерції відіграє важливу роль для споживачів, оскільки дозволяє здійснювати покупки, без необхідності фізичного відвідування магазину. Це особливо важливо для людей, які шукають зручні та доступні способи покупок. На сьогоднішній день важко знайти людину, яка не використовує смартфон, що робить системи електронної комерції ще більш доступними будь де та будь коли;

– інтернет-магазини можуть пропонувати розширений асортименту товарів і послуг, включаючи ті, які важко знайти в звичайних магазинах.;

– споживачі все більше очікують персоналізованих підходів до покупок, і системи електронної комерції можуть надавати рекомендації та персоналізовані пропозиції більш ефективно за людей-консультантів в звичайних магазинах;

– з інтернет-торгівлі легко виходити на міжнародний ринок, і це створює нові можливості для підприємств отримати товари, які можна знайти лише закордоном.

Класифікація електронної комерції за учасниками електронних відносин дозволяє виділити основних суб'єктів електронної комерції, що можуть бути зацікавлені в її використанні. У залежності від учасників електронних взаємовідносин виділяють такі системи електронної комерції:

– бізнес-до-бізнесу (B2B) – одна компанія закупає товари або послуги від іншої компанії;

– бізнес-до-споживача (B2C): продаж товарів або послуг кінцевим споживачам;

– споживач-до-споживача (C2C): один користувач продає товар або послугу іншому користувачеві, часто через онлайн-платформи або аукціони;

– споживач-до-бізнесу (C2B): споживачі пропонують свої товари або послуги компаніям [3].

Для всіх вищезгаданих типів систем електронної комерції важливі рекомендаційні функції. Споживачі очікують більш персоналізованих підходів до покупок, і системи електронної комерції можуть надавати рекомендації та

персоналізовані пропозиції задля покращення користувацького досвіду та підвищення прибутковості електронних торгових платформ для власників бізнесів.

Тому уточнено діапазон можливих ролей, які може виконувати рекомендаційні функції від імені постачальника послуг та від імені користувача. Існує багато причини, чому постачальники послуг можуть захотіти використовувати рекомендаційні функції:

– для збільшення кількості проданих товарів та доходу. Це основна функція для комерційної рекомендаційної функції – можливість продавати більше товарів порівняно з продажами без будь-яких рекомендацій. Некомерційні програми мають подібні цілі. Наприклад, новинний сайт збільшує кількість прочитаних новин. Загалом, що з точки зору постачальника послуг, основною метою впровадження рекомендаційних є збільшення кількості користувачів, які приймають рекомендацію та споживають товар, порівняно з кількістю простих відвідувачів, які просто переглядають інформацію.;

– продаж більш різноманітних товарів. Ще одна головна функція рекомендаційних функцій — це надання можливості користувачеві дізнатися про товари, які може бути важко знайти без точної рекомендації. Наприклад, в туристична рекомендаційна функції може запропонувати подорож в незнане користувачем місці;

– підвищення задоволеності користувачів. Добре розроблені рекомендаційні функції також може покращити досвід роботи користувача з сайтом або програмою та отримає задоволення від використання системи. Поєднання ефективних, точних рекомендацій і зручного інтерфейсу підвищить суб'єктивну оцінку системи користувачем. Це, у свою чергу, збільшить використання системи та ймовірність того, що рекомендації будуть прийняті;

– краще розуміння того, чого хоче користувач. Іншою важливою функцією рекомендаційної функції, яку можна використовувати для багатьох інших програм, є опис уподобань користувача, які збираються або явно, або передбачені системою.

Потім постачальник послуг може вирішити повторно використати ці знання для ряду інших цілей, такі як покращення управління запасами або виробництвом товару [8].

Тепер зрозуміло чому постачальники послуг запроваджують рекомендаційні функції. Але користувачам вони також стануть у нагоді, так як рекомендації підтримують їх основну ціль – отримати необхідний товар чи послугу. Система рекомендує всі предмети, які можуть задовольнити деякі потреби користувачів. У таких випадках недостатньо просто знайти хороші речі. Це особливо вірно, коли кількість елементів відносно невелика або коли RS є критично важливим, наприклад, у медичних або фінансових програмах. У цих ситуаціях, окрім переваг, отриманих від ретельного вивчення всіх можливостей, користувач також може отримати вигоду від ранжирування цих елементів у RS або від додаткових пояснень, які створює RS. Отже, рекомендаційні функції повинні збалансовувати потреби обох сторін.

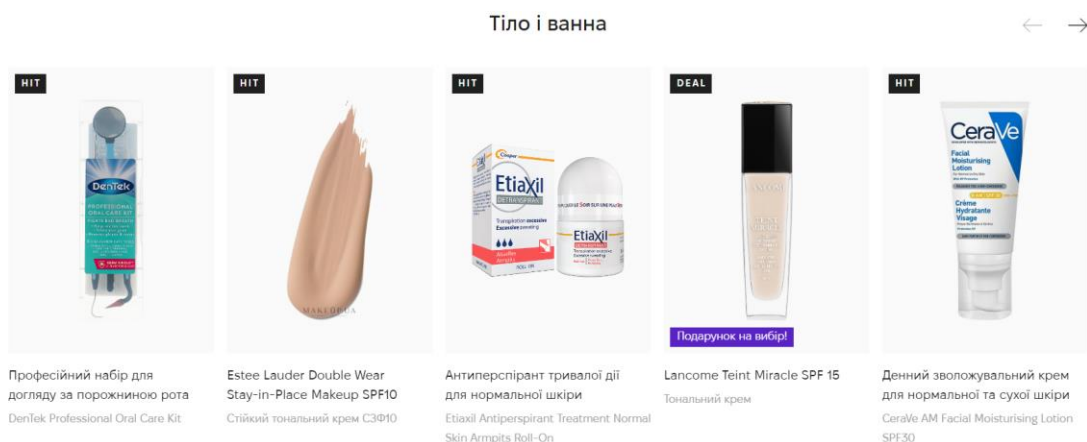


Рисунок 1.1 – Рекомендовані товари для незареєстрованого користувача

Розглянемо приклад існуючої рекомендаційної функції інтернет магазину [22]. На головній сторінці для незареєстрованого користувача ми одразу ж можемо спостерігати роботу рекомендаційної функції. Для деяких категорій товарів система пропонує нам хіти продаж або новинки. Наприклад, на рис. 1.1, для

категорії товарів «Тіло і ванна» пропонується п'ять найпопулярніших товарів. При відкритті сторінки детальної інформації будь-якого товару, маємо вже більш персоналізовану рекомендацію «Інші клієнти також купили» (рис. 1.2).

Інші клієнти також купили








 <p><b>DEAL</b></p> <p>Подарунок на вибір!</p>	 <p><b>DEAL</b></p> <p>Подарунок на вибір!</p>	 <p><b>DEAL</b></p> <p>Подарунок на вибір!</p>	 <p><b>DEAL</b></p> <p>Подарунок на вибір!</p>	 <p><b>DEAL</b></p> <p>Подарунок на вибір!</p>
<p>Комплексний коригувальний засіб проти недоліків і слідів La Roche-Posay Effaclar Duo (+) SPF30</p> <p>★★★★★ 86</p> <p>638 € 748 €</p>	<p>Гель-крем потрійної дії для корекції недоліків проблемної La Roche-Posay Effaclar Duo + M</p> <p>★★★★★ 1798</p> <p>418 € 522 €</p>	<p>Лосьйон для звуження пор з мікро-відлущувальним La Roche-Posay Effaclar Astringent Lotion</p> <p>★★★★★ 729</p> <p>481 € 526 €</p>	<p>Очищувальна себорегулювальна маска La Roche-Posay Effaclar Unclogging Purifying Sebo-Controlling Mask Anti-</p> <p>★★★★★ 343</p> <p>474 € 575 €</p>	<p>Очищувальний гель-мус для жирної та проблемної шкіри La Roche-Posay Effaclar Purifying Foaming Gel</p> <p>★★★★★ 2186</p> <p>467 € 512 €</p>

Рисунок 1.2 – Рекомендовані товари базуючись на зацікавленому користувачем товарі

Неточність рекомендаційних функції може бути викликана різними факторами. Ось декілька факторів, які можуть впливати на точність рекомендацій:

– недостатні дані про користувачів. Особливо ця проблема виникає в невеликих системах електронної комерції, де вибірка вподобань, історій покупок, рейтингів клієнтів тощо невелика, тому система має обмежену інформацію про користувачів, що може призвести до менш точних рекомендацій;

– нові користувачі й нові товари. Можуть виникнути проблеми з наданням рекомендацій для нових користувачів, визначити вподобання яких неможливо без їх взаємодії з системою, та для нових товарів, для яких ще немає відгуків від споживачів;

– зміна вподобань користувачів. Уподобання користувачів можуть змінюватися з часом, і рекомендаційні функції можуть не завжди враховувати ці зміни;

– обмежені алгоритми. Використовувані алгоритми рекомендацій можуть бути обмеженими в їхній здатності адаптуватися до складних сценаріїв;

– неправильний вибір алгоритму. Не завжди можливо знайти один універсальний алгоритм рекомендацій, і вибір неправильного алгоритму може призвести до неточних рекомендацій.

Для рекомендаційних функцій головне завдання не просто давати рекомендації – а давати їх ефективно та найбільш точно з урахуванням вхідних даних. Для покращення точності рекомендацій можна використовувати різні стратегії, такі як поліпшення збору та аналізу даних та розвиток більш складних алгоритмів.

## 1.2 Постановка задачі

Завданням кваліфікаційної роботи є дослідження особливостей використання методів визначення рекомендацій в системах електронної комерції. Робота спрямована на аналіз та порівняння різних підходів реалізації рекомендаційних методів для систем електронної комерції, з метою виявлення їхніх переваг та недоліків у конкретних умовах застосування.

Системи електронної комерції мають однакову мету – продати товар чи послугу, але вхідні дані для рекомендаційних систем, таких як інформація про користувача та інформація про товар, всі вони надають у різному вигляді та об'ємі. У залежності від вхідних даних підбирається метод рекомендацій, який найефективніше впорається з наданням рекомендацій.

Необхідно виконати такі завдання:

- провести аналіз методів визначення рекомендацій у системах електронної комерції;
- обрати предметну область електронної комерції з обмеженим асортиментом товарів для проведення дослідження;
- вибрати архітектуру реалізації системи електронної комерції з обмеженим асортиментом товарів;
- розробити базу даних системи електронної комерції з обмеженим асортиментом товарів;
- провести дослідження особливостей реалізації методів визначення рекомендацій в е-системах з обмеженим асортиментом товарів;
- експериментально дослідити реалізацію обраних методів з визначенням змін компонентів системи електронної комерції (база даних, web-інтерфейс доступу до бази даних);
- реалізувати базу даних і web-інтерфейс системи електронної комерції з визначеними методами рекомендацій.

## 2 АНАЛІЗ МЕТОДІВ ДЛЯ ВИЗНАЧЕННЯ РЕКОМЕНДАЦІЙ У СИСТЕМАХ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

### 2.1 Теоретичні підходи (методи) для визначення рекомендацій в системах електронної комерції

Рекомендаційні системи мають на меті надати користувачам рекомендацій, що допоможуть їм вибирати товари або послуги які найбільше відповідають їхнім потребам і вподобанням. Основною задачею рекомендаційних систем є полегшення та прискорення процесу прийняття рішень користувачем і підвищення задоволення від використання системи. Вони застосовуються в найрізноманітніших застосунках, включаючи електронну комерцію, медіа, соціальні мережі, музику, кіно, новини і багато інших.

У системах електронної комерції налічується багато методів і теоретичних підходів, які використовують для рекомендації товарів та послуг, та які допомагають покупцям знаходити саме те, що їх цікавить. Підходи до рекомендацій щодо товарів та послуг можна розділити на дві обширні категорії: персоналізовані та неперсоналізовані [8].

Для створення рекомендаційного алгоритму потрібно зібрати різноманітні дані, які дозволять системі розуміти і передбачувати вподобання та поведінку користувачів. Необхідні дані можна розділити на дві головні категорії:

– дані користувачів. Основна інформація про користувачів, така як вік, стать, місце проживання, освіта і інші персональні характеристики. А також історія взаємодії, наприклад інформація про те, як користувач взаємодіє з системою, включаючи перегляди, покупки, оцінки, пошуки, дії на сайті тощо;

– дані товарів або послуг. Описи об'єктів – інформація про товари, послуги, контент або інші об'єкти, включаючи назву, опис, категорії, зображення тощо. А також оцінки та відгуки користувачів про товари або послуги;

Загалом система рекомендацій класифікується за методом отримання даних:

1) за пам'яттю (Memory-Based) – коли використовується інформація, що зберігається в базі даних – це особисті дані клієнтів (здебільшого демографічні) і дані їх історії замовлень;

2) за моделлю (Model-Based) – коли використовується модель, що визначає рейтингові оцінки товарів, що надані клієнтами. Модель е-системи автоматизовано фіксує дій користувача під час його роботи з інтерфейсом для визначення оцінок рейтингу товарів. Ця модель реалізується декількома методами, наприклад:

– впровадження моделі відгуків на товар за допомогою визначення твердження «подобається/не подобається» або «рекомендую/не рекомендую», що реалізується у вигляді кнопок-ярликів інтерфейсу (лайків, дізлайків тощо);

– виставляння рейтингу товару за прийнятою в моделі е-системі шкалою;

– створення особистого профілю клієнта, в якому визначається рейтинг товарів за визначеними характеристиками (цей профіль заповнює клієнт під час реєстрації в е-системі);

– анкетування клієнтів за визначеними характеристиками й категоріями товарів;

– фіксація та розрахунок статистики використання параметрів (визначених ключових слів), що застосовувались клієнтом для пошуку товарів;

– фіксація та розрахунок статистики за найменуваннями товарів, які переглядали клієнти;

3) гібридний підхід, коли використовується методи отримання даних за пам'яттю (Memory-Based) і за моделлю (Model-Based).

Неперсоналізовані рекомендації – це рекомендації, які однакові для всіх користувачів і не враховують особистих характеристик. Вони виконують свою роль, коли відсутні необхідні дані про користувачів для створення персоналізованих рекомендацій. А також можуть бути використані для просування нових товарів. Однак користувачі жадають більш індивідуальних рекомендацій від систем, якими вони користуються. Тому неперсоналізовані рекомендації не завжди ефективні для забезпечення задоволення потреб користувачів.

Персоналізовані рекомендації – це рекомендації, які надаються користувачам на основі їхніх індивідуальних властивостей, інтересів і попередньої взаємодії з системою або платформою. Основна особливість персоналізованих рекомендацій – це індивідуальний підхід до кожного користувача. Системи враховують інтереси, історію взаємодії, лайки та інші індивідуальні дані, щоб створити рекомендації, які відповідають потребам конкретного користувача. Для створення персоналізованих рекомендацій системи використовують різні джерела даних. Це можуть бути явні дані, такі як оцінки, відгуки, покупки, та неявні дані – перегляди, час проведений на сторінці.

На рис. 2.1 зображено класифікацію за джерелом отримання даних, обрана класифікація рекомендаційних методів та рекомендаційні алгоритми.

1) методи, що засновані на описі характеристик товару та даних користувачів (Content-Based):

– зважування даних «TF-IDF». Для визначення рейтингу використовується значення статистичної міри «TF-IDF» (Term Frequency – Inverse Document Frequency), який зазвичай застосовується для текстових документів (книжок). Коефіцієнт TF використовується для оцінки важливості ключового слова й знаходиться як відношення числа його входжень до загального числа слів документа. Коефіцієнт IDF – це інверсія частоти, з якою ключове слово зустрічається в документах. Статистична міра TF-IDF є добутком двох співмножників:  $TF \times IDF$ . Велику вагу за TF-IDF отримують ключові слова з

високою частотою в межах конкретного документа та з низькою частотою використання інших документів;

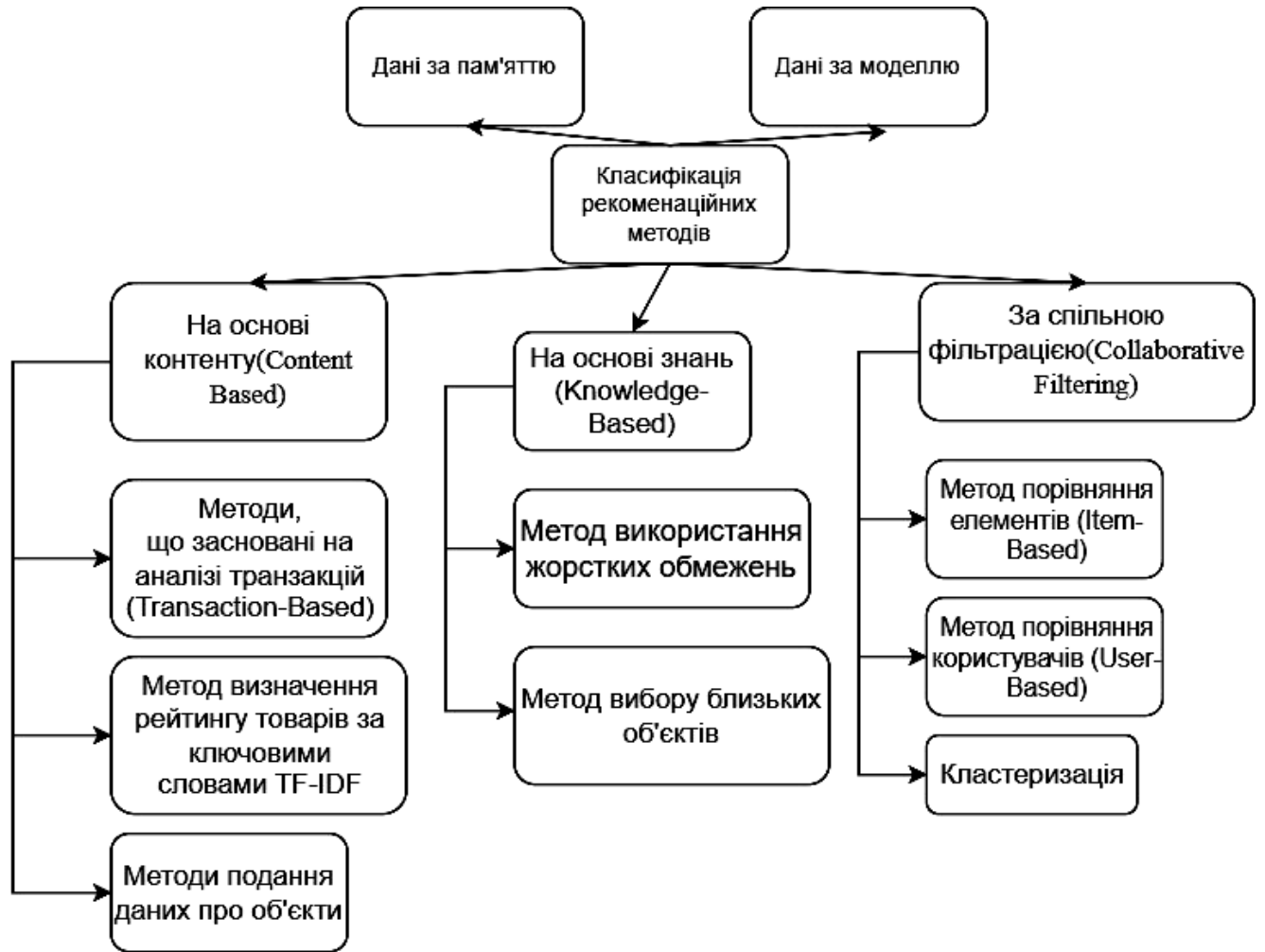


Рисунок 2.1 – Варіант класифікації рекомендаційних методів та рекомендаційні алгоритми

– методи подання даних про об'єкти. За цими методами будується профілі клієнтів і товарів. Для створення профілю клієнта використовуються анкети або інші види опитувань з метою отримання інформації про його інтереси. Порівняння (аналіз) профілів товарів і профілю клієнтів дозволяє зробити висновок, який товар йому подобається;

2) методи, що засновані на знаннях (Knowledge-Based);

– метод вибору близьких об'єктів. Для цього методу інформація про товар подається як набір параметрів, що описують властивості товару. Для системи реалізуються функції пошуку або фільтрації товарів за обраними параметрами. Завдання функцій системи – допомогти клієнту відібрати товар, який задовольняє його запит за параметрами (властивостями);

– метод використання жорстких обмежень, призначений для визначення вимог користувача. Вимоги користувача визначаються за допомогою функції системи, яка реалізує задану кількість та послідовність запитань. За допомогою отриманих відповідей будується запит до бази даних і подається перелік товарів;

3) методи, що засновані на сумісній (колаборативній) фільтрації (Collaborative Filtering):

– метод порівняння користувачів (User-Based). Цей метод дозволяє подати обраному клієнту рекомендовану прогнозну оцінку товару, для якого він не виставив оцінку. Прогнозна оцінка товару розраховується на основі оцінок інших клієнтів, які відібрані за мірою схожості (збігу) оцінювання всіх товарів. Під час генерації рекомендацій особиста інформація клієнтів про придбанні товари не використовується;

– метод порівняння елементів (Item-Based). Цей метод також дозволяє подати обраному клієнту рекомендовану прогнозну оцінку товару, для якого він не виставив оцінку. Прогнозна оцінка товару в цьому випадку розраховується на основі оцінок товарів, відібраних за мірою схожості (збігу) їх оцінювання іншими клієнтами. Під час генерації рекомендацій особиста інформація клієнтів про придбанні товари також не використовується;

4) гібридні методи (Hybrid Filtering), які використовують методи пунктів «1)», «2)» і «3)» сумісно.

## 2.2 Методи визначення рекомендацій на основі контенту (Content Based)

Метод визначення рекомендацій на основі контенту (Content-Based Filtering) використовує інформацію про характеристики товарів або послуг і вподобання користувачів для створення персоналізованих рекомендацій. Дані системи порівнюють опис товарів, що у продажу з інтересами користувача, отриманими з його попередніх оцінок.

Цей підхід фільтрації на основі контенту зображено на рисунку 2.2.

Перевага даного підходу в тому, що модель не потребує даних про інших користувачів, оскільки рекомендації є індивідуальними. Також модель може враховувати конкретні інтереси користувача. Недоліки в тому, що цей метод вимагає наявності опису для кожного товару та те, що рекомендації можна створити лише на основі існуючих інтересів клієнта, так як для нових клієнтів недостатньо вхідних даних.

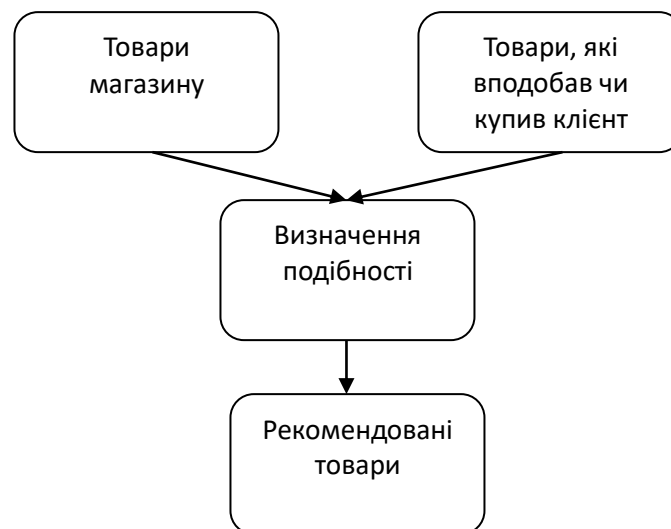


Рисунок 2.2 - Структура підходу фільтрації на основі контенту для рекомендацій товарів

### 2.2.1 Подання даних про об'єкт

Кожен об'єкт у будь-якій системі рекомендацій має назву, а також іншу текстову інформацію, яку можна використовувати. Користувачі системи мають свій профіль. Там вони вказують інформацію про себе та свої інтереси. Маючи ці дані та інформацію про товари можна зробити висновки, чи цікавий конкретний товар користувачу. Але, в цьому випадку міра схожості інтересів користувача та книги набуває двох можливих значень: 1 – якщо подібний товар описаний у профілі користувача, 0 – в іншому випадку.

Для більш точних рекомендацій використовується інший підхід: знаючи інформацію про товари, куплені людиною, йому можна порадити схожі товари. Схожість двох товарів вимірюється за допомогою порівняння наявних даних.

Наприклад, якщо  $keywords(B_i)$  – множина ключових слів товару  $B_i$ , то порівняти її з книгою  $B_j$  можна так:

$$dice(B_i B_j) = \frac{|keywords(B_i) \cap keywords(B_j)|}{|keywords(B_i)| + |keywords(B_j)|} \quad (2.1)$$

Формула (2.1) визначає коефіцієнт Дайса. Він набуває значень з відрізка  $[0,1]$  і показує, наскільки близькі дві множини. Але множина ключових слів майже ніколи не є явно. Їх можна одержати, аналізуючи безпосередньо зміст об'єктів.

### 2.2.2. Зважування даних TF-IDF

Коефіцієнт TF-IDF показує, наскільки слово характеристики товару важливо. Основна ідея методу в тому, що слово вважається більш важливим, чим частіше воно зустрічається в описі характеристики одного товару, і рідко в інших.

Нехай набір текстів (документів). Усі тексти розбиті слова. Коефіцієнт TF-IDF показує, наскільки слово важливе у тексті. Основна ідея в тому, що слово вважається важливішим, якщо воно часто зустрічається в одному тексті та рідко в інших. TF (term frequency) – це нормалізована частота слова в тексті: Метод описує формула (2.2).

$$TF(x, D) = \frac{freq(x, D)}{\max_{y \in D} freq(y, D)} \quad (2.2)$$

де  $freq(x, D)$  – це кількість слів  $x$  у тексті  $D$ . TF приймає значення з відрізка  $[0, 1]$ .

IDF (inverse document frequency) – зворотна частота вжитих слів. Формула (2.3)

$$IDF(x) = \frac{N}{n(x)} \quad (2.3)$$

$N$  – це кількість текстів у наборі,  $n(x)$  – кількість текстів, у яких трапляється слово  $x$ .

Коефіцієнт TF-IDF обчислюється як добуток TF на IDF. Для порівняння двох текстів, їх можна подати у вигляді векторів у евклідовому просторі. Кожному виміру відповідає слово, а значення кожного компонента вектора дорівнює TF-IDF. А потім порахувати косинус між цими векторами.

Для зменшення обчислювальних витрат, використовуються різні евристики:

– стоп-слова. Деякі слова як прийменники, сполучники зустрічаються в текстах дуже часто, хоча не містять ніякої корисної інформації. Тому ці слова заносяться до списку стоп-слів і не розглядаються при побудові векторної моделі;

– зменшення кількості компонентів вектора. Одне слово може приймати множина форм (різні відмінки, часи тощо.). Тому існують алгоритми, що дозволяють перевести слово до нормальної форми, щоб ототожнювати слова;

– для зменшення кількості компонентів вектора, можна зменшити кількість різних слів, залишивши лише з найбільшою вагою TF-IDF. Різних слів у документі дуже багато.

До переваг даного методу можемо віднести:

– враховує не тільки частоту зустрічання слова в документі, але і його важливість для всього корпусу документів;

– завдяки стоп-словам зменшує кількість неважливих слів, підсилюючи значущі;

– допомагає виявляти ключові слова в текстах, які найбільше характеризують сутність об'єктів.

До недоліків методу можемо віднести:

– обчислення TF-IDF може бути ресурсозатратним завданням, особливо для великих об'ємів документів;

– вимагає точності обробки тексту та лінгвістичної підготовки для ефективної роботи [4].

### 2.3 Методи визначення рекомендацій за спільною фільтрацією (Collaborative Filtering)

Колаборативна фільтрація (Collaborative filtering) – це метод рекомендації, у якому аналізується лише реакція користувачів на об'єкти: оцінки, які виставляють користувачі об'єктам. Оцінки можуть бути як явними (користувач явно вказує, на скільки «зірочок» він оцінює об'єкт), так і неявними (наприклад, кількість переглядів одного ролика). Чим більше оцінок збирається, тим точніше вийдуть рекомендації. Виходить, що користувачі допомагають один одному у фільтрації об'єктів. Тому такий метод називається також спільною фільтрацією. Є два основних етапів для створення рекомендацій: подібність клієнтів/товарів та

прогнозування рейтингу товарів. Цей підхід спільної фільтрації зображено на рисунку 2.3.

Рекомендації зі спільною фільтрацією поділяються на дві основні категорії: фільтр колаборативного споживача (User-Based Collaborative Filtering) і фільтр колаборативного предмета (Item-Based Collaborative Filtering).



Рисунок 2.3 – Структура підходу на основі спільної фільтрації для рекомендацій товарів

У методі колаборативного споживача схожість між користувачами визначається на основі схожості їхньої поведінки в системі. Наприклад, якщо два користувачі мають подібні замовлення, оцінки, або лайки то їм можуть сподобатися ті самі товари чи послуги.

У методі колаборативного предмета схожість між товарами визначається на основі взаємодії користувача з ними. Якщо багато користувачів купували два

певних товари разом або давали їм високі оцінки, то ці товари вважаються схожими.

### 2.3.1. Кластеризація

Мета алгоритму – віднести користувача до кластеру, що містить найбільш схожих на нього клієнтів. Використовуючи метрику подібності, алгоритм кластеризації групує найбільш схожих клієнтів разом, щоб сформувати кластери. Ці алгоритми зазвичай починаються з початкового набору кластерів, кожен з яких часто містить одного випадково вибраного клієнта. Потім вони підбирають клієнтів до існуючих кластерів, з можливістю для створення нових кластерів.

Найкраще вирішення проблеми кількості порівнянь є методи кластеризації, такі як k-means.

K-means – це один з найпопулярніших методів кластеризації. Для реалізації алгоритму k-means обирається кількість кластерів і призначаються початкові значення центрів кластерів. Кожному об'єкту з набору даних призначається кластер з найближчим центром, після чого обчислюються нові центри для кожного кластера. Кроки повторюються ітераційно. Алгоритм:

- обрати k точок – центри майбутніх кластерів;
  - кожному об'єкту присвоюють кластер так, щоб відстань від цього об'єкту до відповідного центру кластера була меншою, ніж відстань до інших кластерів;
  - базуючись на отриманих елементах кластерів, провести перерахування нових центрів кластерів, як середнього значення всіх елементів, що належать кластерові;
  - ітераційно, кроки 2 та 3 повторюють доти, поки система не стабілізується.
- Необхідно звернути увагу, що:
- необхідно спрогнозувати кількість кластерів до початку роботи алгоритму.
  - алгоритм досягає не глобальних, а локальних мінімумів;

– на результат суттєво впливає розподіл початкових центрів кластерів, але алгоритм не надається рекомендацій для коректного їх вибору [11].

### 2.3.2 Методи обчислення подібності (Similarity Computation)

Розрахунок схожості між користувачами (User-based Collaborative filtering) – для кожної пари користувачів/товарів обчислюється ступінь схожості на основі їхніх дій, таких як покупки, оцінки або перегляди товарів. Ступінь схожості можна розрахувати за формулою косинуса схожості або коефіцієнт кореляції [9].

Нехай у системі є користувачі та об'єкти. Нехай деякі користувачі оцінили деякі об'єкти. Нехай маємо користувач  $a$ . Наше завдання – передбачити, яку оцінку поставив би користувач  $a$  об'єкту  $i$ . Розглянемо тільки користувача  $a$  та тих користувачів, які оцінили об'єкт  $i$ . Алгоритм включає 3 кроки:

- для кожного користувача  $u$  обчислимо, наскільки його інтереси збігаються з інтересами користувача  $a$ ;
- після цього виберемо множину користувачів, найбільш близьких до  $a$ ;
- передбачимо оцінку на основі оцінок об'єкта  $i$  "сусідами" з попереднього кроку.

Перший крок. Кожному користувачеві в матриці  $R$  відповідає один рядок. Тому обчислюватимемо близькість векторів-рядків користувачів. Існує множина способів підрахунку близькості векторів. Один із найпростіших – порахувати косинус між цими векторами:

$$\text{sim}(u, a) = \frac{\sum_{i=1}^m r_{a,i} \times r_{u,i}}{\sqrt{\sum_{i=1}^m r_{a,i}^2} \times \sqrt{\sum_{i=1}^m r_{u,i}^2}} \quad (2.4)$$

Тут  $\text{sim}(u, a)$  – міра близькості (схожості) користувачів  $a$  та  $u$ .  $r_{u,i}$  - значення матриці  $R$ :  $u$  рядок,  $i$  стовпець.  $\text{sim}(u, a)$  набуває значення з відрізка  $[0,1]$ . Якщо

користувач не вказав оцінку для якогось об'єкта, відповідне значення матриці дорівнює 0.

Косинусна подібність не залежить від абсолютних значень векторів чи їхньої розмірності. Це робить метод придатним для порівняння векторів різних розмірностей. Але коли вектори мають малі або нульові абсолютні значення, косинусна подібність може бути нестійкою.

Існує інший захід близькості між векторами – коефіцієнт кореляції Пірсона:

$$sim(u, a) = \frac{\sum_{i \in I} (r_{a,i} - \bar{r}_a) \times (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i \in I} (r_{a,i} - \bar{r}_a)^2 \times (r_{u,i} - \bar{r}_u)^2}} \quad (2.5)$$

Значення р-Пірсона знаходиться в межах від -1 до 1, що дозволяє легко порівнювати ступінь кореляції між двома векторами.

Тут  $I$  – множина об'єктів, які оцінив як користувач  $a$ , і  $r_{a,i}$  і  $r_{u,i}$  - це середні оцінки користувачів  $a$  і  $u$ , користувач  $u$ . відповідно.

У чисельнику підраховується добуток відхилень оцінок двох користувачів від середніх значень одного об'єкта. Знаменник необхідний у тому, щоб ця величина приймала значення з відрізка  $[-1,1]$ . Чим сильніше збігаються інтереси, тим ближчі значення близькості до 1. Якщо коефіцієнт кореляції Пірсона негативний, то інтереси користувачів протилежні.

Вирази (2.4) та (2.5) – найпоширеніші заходи близькості, які використовуються в даному алгоритмі. Коефіцієнт кореляції Пірсона (2.5) буде залишатися високим, якщо всі оцінки користувачів, що порівнюються, відрізняються на постійну величину (один користувач оцінює більш строго, інший – ставить більш високі оцінки). Косинус (2.4) у разі видасть менше значення. Але якщо оцінки двох користувачів відрізняються пропорційно, то косинус (2.4) буде більше, кореляція (2.5) – менше. В обох випадках оцінки користувачів схожі. Тому, яку з цих метрик використати залежить від конкретного завдання.

Другий крок. Тепер потрібно вибрати множину  $K$  найбільш схожих користувачів  $a$ . Можна вибрати всіх користувачів. Але оскільки користувачів з несхожими чи несильними інтересами, що перетинаються, досить багато, то вони будуть негативно впливати на точність передбачення оцінки для об'єкта  $i$ . Крім того, кількість користувачів впливає обсяг обчислень на третьому кроці.

Одне з можливих рішень – встановити поріг міри близькості, обчисленої першому кроці: (2.4), (2.5). Користувачі з мірою близькості, що перевищує поріг, увійдуть у множину  $K$ . Інші – ні. Але найчастіше вибирається ціла константа  $k$ . Потім усі користувачі сортуються за зменшенням міри близькості. І до множини  $K$  входять користувачі, найбільш близькі до  $a$ .

Третій крок. Маючи множину  $K$  близьких користувачів, потрібно обчислити оцінку, яку поставив би користувач  $a$  об'єкту  $i$ . Нагадаємо, що розглядаються лише ті користувачі, які оцінили об'єкт  $i$ . Потрібна оцінка обчислюється за такою формулою:

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u \in K} (r_{u,i} - \bar{r}_u) \times \text{sim}(a, u)}{\sum_{u \in K} |\text{sim}(a, u)|} \quad (2.6)$$

де  $p_{a,i}$  – це очікувана оцінка користувача  $a$  для об'єкта  $i$ . За основу береться його середня оцінка  $\bar{r}_a$  а потім додається середнє відхилення оцінки інших користувачів з множини  $K$  для об'єкта  $i$  від їх середньої оцінки. Чим ближче користувач  $u$  до користувача  $a$  (з мірою близькості  $\text{sim}(a, u)$ , обчисленої на першому кроці), тим сильніший його внесок у прогноз оцінки [10].

Таким чином, описаний алгоритм передбачає оцінки для об'єктів, які поточний користувач ще не оцінив. Для того, щоб зробити рекомендацію для даного користувача, достатньо передбачити оцінки для всіх неоцінених об'єктів і вибрати об'єкти з найбільшою оцінкою .

### 2.3.3 Альтернативні алгоритми та покращення (метод порівняння елементів)

Існує також інший підхід: замість користувачів порівнювати об'єкти (Item-based Collaborative filtering). У цьому випадку алгоритм майже такий самий, включає 3 кроки:

- для кожного об'єкта  $j$  обчислимо, наскільки він схожий на об'єкт  $i$ , для якого передбачається оцінка;
- виберемо множину об'єктів, найближчих до  $i$ ;
- передбачимо оцінку на основі оцінок обраних на другому кроці об'єктів користувачем  $a$ .

Порівняння об'єктів першому кроці відбувається як і (2.4) і (2.5). Другий крок аналогічний до другого кроку алгоритму, заснованого на порівнянні користувачів. Основна відмінність у тому, що  $K$  – множина близьких об'єктів. На третьому кроці рахується виважена середня оцінка схожих об'єктів, отримана користувачем  $a$ .

### 2.4. Методи визначення рекомендацій на основі знань (Knowledge-Based)

Методи визначення рекомендацій на основі знань передбачає використання експертних знань або правил для генерації рекомендацій, навіть якщо система не має достатньо потрібної кількості інформації про товари, що купуються користувачем, на основі яких можна було б надавати рекомендації. Тобто, цей метод найкраще підходить для рекомендацій товарів, які купуються нерегулярно [12].

Для вирішення вище описаного завдання використовуються методи, що базуються на знаннях, їх можна поділити на дві групи:

- використання жорстких обмежень;
- вибір близьких об'єктів.

Ідея обох методів закладається у наступному: клієнти чітко формулюють свої вимоги до товару, для якого важко отримати оцінки, коли система намагається знайти потрібний товар.

У першому випадку алгоритм знаходить та рекомендує лише ті товари, які точно відповідають усім вимогам користувача. У другому випадку знаходить товари з характеристиками, близькими до вимог.

#### 2.4.1 Використання жорстких обмежень

Основна ідея методу це знайти товари, які повністю задовольняють вимоги користувача. Ключове завдання полягає в ефективній взаємодії з користувачем, задати правильні запитання задля отримання його вимог до бажаного товару.

Для того щоб обмежити кількість запитань, що задаються користувачеві, використовуються значення параметрів за замовчуванням. Вони поділяються на 3 групи:

- статичні замовчування – базові, необхідні атрибути;
- залежні замовчування – атрибути, які застосовуються в парі з іншим атрибутом;
- похідні замовчування – визначаються після взаємодії системи з клієнтом.

Користувачі вказують не всі необхідні параметри, а лише ті, які вважаються найбільш суттєвими. Якщо користувач вказав занадто жорсткі вимоги, то може виявитися, що шуканих товарів немає в системі. У цьому випадку система послідовно та автоматично послаблює обмеження користувача на товари.

## 2.4.2 Вибір близьких об'єктів

Методи вибору близьких товарів рекомендують товари, які є найбільш близькими до вимог користувачів. У цьому випадку можлива рекомендація товарів, які не повністю відповідають вимогам.

Міра близькості між вимогами та властивостями обчислюється за формулою:

$$\text{similarity}(p, REQ) = \frac{\sum_{r \in REQ} w_r \times \text{sim}(p, r)}{\sum_{r \in REQ} w_r} \quad (2.7)$$

Тут  $p$  – об'єкт,  $REQ$  – множина вимог користувача.  $r$  – вимога.  $w_i$  – вага даної вимоги,  $\text{sim}(p, r)$  – міра близькості товару  $p$  до вимоги  $r$ .

Чисельні властивості поділяються на три групи:

- «більше – краще» (наприклад, чим більший товар тим краще);

$$\text{sim}(p, r) = \frac{\varphi_r - \min(r)}{\max(r) - \min(r)} \quad (2.8)$$

- «менше – краще» (наприклад, чим нижча ціна, тим краще);

$$\text{sim}(p, r) = \frac{\max(r) - \varphi_r}{\max(r) - \min(r)} \quad (2.9)$$

- «ближче – краще» (наприклад, користувачеві потрібен монітор з певною діагоналлю, більше чи менше – гірше).

$$sim(p, r) = 1 - \frac{|\varphi_r(p) - r|}{\max(r) - \min(r)} \quad (2.10)$$

Тут  $sim(p, r)$  – міра близькості товару  $p$  до вимоги  $r$ ;  $\min(r)$ ,  $\max(r)$  – мінімальні та максимальні значення властивості, відповідно. в формулі (2.10)  $r$  – це точне значення вимоги користувача;  $\varphi_r(p)$  – значення відповідної властивості товару  $p$  [13].

## 2.5 Висновки за другим розділом

Отже, у другому розділі було класифіковано рекомендаційні системи електронної комерції за підходами – персоналізовані та неперсоналізовані, за методами отримання даних, та за способом генерації персоналізованих рекомендацій. За методами отримання даних класифікується за пам'яттю (Memory-Based) та за моделлю (Model-Based). Серед персоналізованих підходів рекомендаційні методи класифікуються на системи рекомендацій на основі контенту (Content-Based), методи, що засновані на знаннях (Knowledge-Based), системи рекомендацій із спільною фільтрацією (Collaborative Filtering). Загальне порівняння мети методів наведено в таблиці 2.1. Визначено, що для створення ефективного рекомендаційного алгоритму потрібно зібрати такі дані: рейтинги користувача, рейтинги користувачів схожих на нього, атрибути товарів чи послуг та вимоги користувача.

Таблиця 2.1 – Мета різних методів рекомендацій

Метод	Мета	Вхідні дані
Collaborative	Надати рекомендацію, основану на спільному підході, що зважає рейтинги та дії користувача, та схожих на нього користувачів.	Рейтинги користувача + рейтинги користувачів схожих на нього
Content-Based	Надати рекомендацію, основану на контенті	Рейтинги користувачів +

	(атрибутах контенту), що був вподобаний користувачем у минулих оцінках, рейтингах, діях тощо.	атрибути товарів чи послуг
Knowledge-based	Надати рекомендацію, основу на явних вимогах користувача щодо того, що він хоче.	Вимоги користувача + атрибути товарів чи послуг

### 3 АНАЛІЗ ОСОБЛИВОСТЕЙ ВИКОРИСТАННЯ МЕТОДІВ ВИЗНАЧЕННЯ РЕКОМЕНДАЦІЙ У СИСТЕМАХ Е- КОМЕРЦІЇ З ОБМЕЖЕНИМ АСОРТИМЕНТОМ ТОВАРІВ

#### 3.1 Вибір предметної області для проведення експериментальних досліджень

Рекомендаційні системи є дуже популярними і навіть необхідними доповненнями до майже кожної системи, яка пропонує деякі товари чи послуги. І, відповідно, їх дослідження, покращення та розробка нових методів є актуальними. Існує безліч областей систем електронної комерції, де рекомендаційні системи впроваджують найчастіше, це кіно-прокат, продаж продуктів харчування, книжок, одягу, квітів, тощо. Огляд апробованих статей сучасного стану теоретичної розробки методів для реалізації рекомендаційних систем [14, 15, 16, 17, 18, 19, 20] дозволяє зробити висновок що ці системи найчастіше розробляються для двох предметних областей систем електронної комерції – прокат кінофільмів та продаж книжок. А для систем електронної комерції з обмеженим асортиментом товарів, таких як продаж одягу, взуття та квітів, статей не знайшлося. З усіх розглянутих областей, квіти – це товари не першої необхідності і є елементами розкоші, на відміну від одягу та продуктів харчування. Саме тому рекомендаційні системи тут дуже ефективні – система може порекомендувати користувачу більш цінні букети що в результаті призведе до задоволення обох сторін. Таким чином підприємство отримає більший дохід та задоволеного клієнта, який повернеться знову. Тому за результатами аналізу доцільно зробити дослідження рекомендаційних методів області систем електронної комерції продажу квітів.

Головною особливістю предметної області систем електронної комерції, що спеціалізуються на продажу квітів для букетів, є обмежений асортимент. Це пов'язано з особливостями квіткового бізнесу в Україні.

Серед факторів, що обмежують асортимент квітів на ринку України слід визначити:

– низький рівень технологій культивування (наявності високосортного насіння) та обмежена база вирощування квітів (теплиць), що пов'язана з неможливістю конкурувати з імпортом. Ці причини визначають перший фактор – сезонність пропозиції квітів. За даними аналітики ринку [6] в українських теплицях вирощують в основному троянди. Друге місце в структурі займають тюльпани, але їх доля не більше 2 %. Запропонована кількість квітів вітчизняних виробників найбільша в літньо-осінній період. Кількість імпортованих квітів більша в зимово-весняний період, що веде за собою збільшення ціни на ті види квітів, що вирощуються в літньо-осінній період (рис. 1);

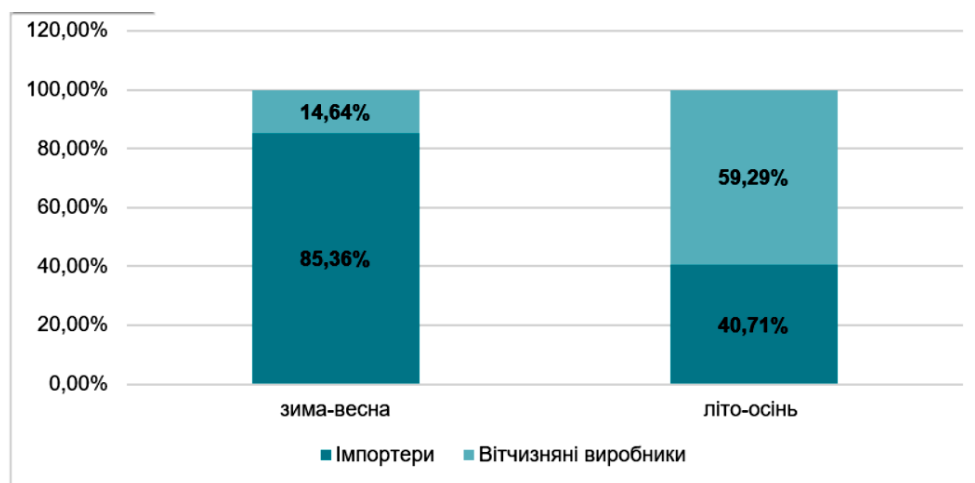


Рисунок 3.1. Сезонна структура ринку квітів України в 2020 році

– другий фактор, пов'язаний з підвищенням статистики продажів у святкові дні. Найбільшу виручку продавцям квітів за аналітикою 2020 року [6] приносять свята 8 березня й 14 лютого, на частку яких припадає 13,5 % і 9,4 % річного

обсягу продажів відповідно (рис. 2). День початку занять у школах (1 вересня) додає ринку ще 5,6 % річного обсягу продаж;

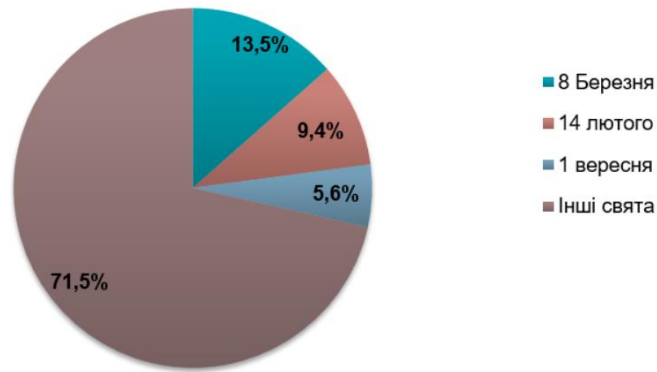


Рисунок 3.2. Структура продажів зрізаних квітів в Україні в 2020 році

– третій фактор, пов'язаний обмеженою покупною спроможністю клієнтів, які чітко поділяються на дві крайні позиції за витраченими коштами: «VIP-клієнтів» та «бізнес-клієнтів» (середня позиція займає малий відсоток). Додатковий до цього фактор – сезонності пропозиції, обумовлює обмежений топ квітів. Наприклад, для свята 8 березня топ двох позицій продажів складають імпортні та вітчизняні тюльпани (1) і троянди (2), що визначають 96 % пропозиції, останні 4 % – це «рідкі» квіти з обмеженою пропозицією (асортиментом): орхідеї, іриси, гладіолуси, півонії, лілії, хризантеми, гербери, гортензії тощо.

### 3.2 Вибір архітектури для реалізації системи електронної комерції

При розробці компонентів рекомендаційної системи з продажу квітів була обрана триланкова архітектура «клієнт-сервер» для глобальної мережі інтернет (рис. 3.3) [21].

Два головні компоненти:

- Клієнт – комп'ютер на стороні користувача, який відправляє запит до сервера для надання інформації або виконання певних дій;

- сервер – більш потужний комп'ютер або обладнання, призначене для вирішення певних завдань з виконання програмних кодів, виконання сервісних функцій за запитом клієнтів, надання користувачам доступу до певних ресурсів, зберігання інформації і баз даних.

Триланкова клієнт-серверна архітектура, розвиток якої спостерігаємо з середини 90-х, реалізує відділення прикладного рівня від рівня управління даними. Основна ідея архітектури «клієнт-сервер» полягає в поділі мережевого додатка на кілька компонентів, кожен з яких реалізує деякий набір сервісів.

Модель такої системи полягає в тому, що клієнт відправляє запит на сервер, де він обробляється, і готовий результат відправляється клієнтові. Сервер може обслуговувати кілька клієнтів одночасно. Якщо одночасно приходить більше одного запиту, то вони встановлюються в чергу і виконуються сервером послідовно. Компоненти такого додатку можуть виконуватися на різних комп'ютерах, виконуючи серверні і клієнтські функції. Це дозволяє підвищити надійність, безпеку і продуктивність мережевих додатків і мережі в цілому.

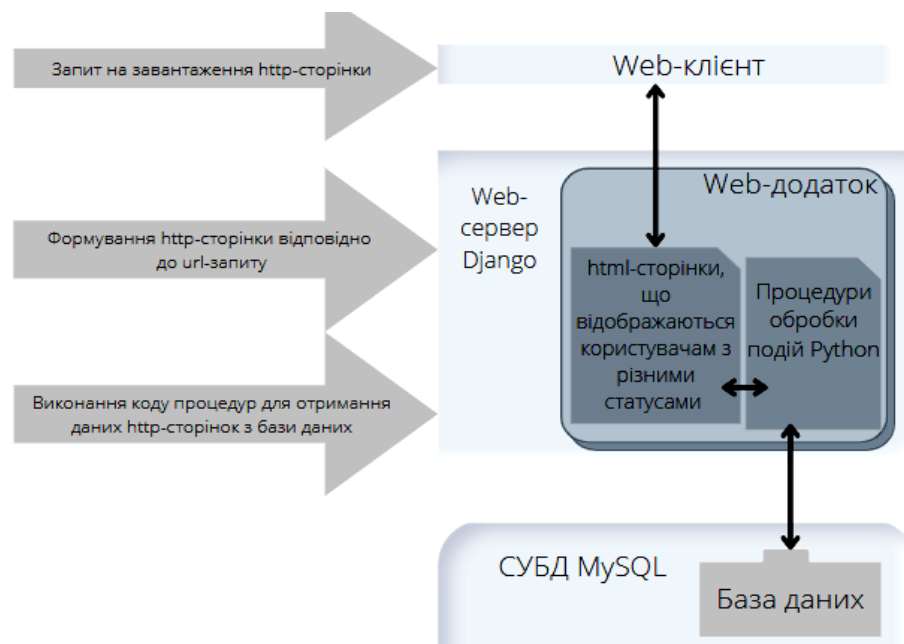


Рисунок 3.3 – Триланкова архітектура «клієнт-сервер»

Триланкового архітектура складніша, ніж дволанкова, завдяки тому, що функції розподілені між серверами другого і третього рівня. Але саме ця особливість дозволяє отримати такі переваги:

- високий ступінь гнучкості і масштабованості;
- високий рівень безпеки через те, що захист визначається для кожного рівня окремо;
- висока продуктивність через розподілення завдання між серверами.

Дворівнева клієнт-серверна архітектура передбачає взаємодію двох програмних модулів – клієнтського та серверного. В залежності від того, як між ними розподіляються функції, розрізняють:

- модель тонкого клієнта, в якій вся логіка застосування та управління даними зосереджена на сервері. Клієнтська програма забезпечує тільки функції рівня представлення (тонкими клієнтами часто також називають пристрої з обмеженою потужністю: кишенькові комп'ютери, мобільні телефони тощо);
- модель товстого клієнта, в якій сервер тільки керує даними, а обробка інформації та інтерфейс користувача зосереджені на стороні клієнта.

Типовим прикладом клієнт-серверної взаємодії є WWW. У тривіальній формі ця інформація це набір веб-сторінок, які можуть зберігатися на сервері у вигляді файлів, реалізованих за допомогою мови розмітки HTML.

Як перша ланка було обрано СУБД MySQL Server. Дане програмне забезпечення відповідає за доступ до даних файлів бази даних. Основна ідея :

- дані зберігаються в таблицях (двовірних структурах), які складаються із стовбців (полів) та рядків (записів);
- кожне поле може містити тільки один тип даних;
- записи містять значення, що відповідають кожному із стовбців — це т.зв. кортеж, сукупність взаємозв'язаних даних.

– один із стовбців (або комбінація з кількох стовбців) — первинний ключ (primary key) - однозначно ідентифікує кожен рядок з таблиці, забезпечує можливість відрізнити один рядок від іншого

– певним чином можна зв'язувати запис однієї таблиці із записом (записами) з інших.

У якості другої ланки – тонкого клієнта – виступає веб-сервер, який зберігає сторінки, які використовують користувачі з різними статусами: «zareєстрований користувач», «nezareєстрований користувач», «адмін». Було обрано веб-сервер «Django», що реалізується мовою програмування Python. Також зберігається код на обробку подій, наприклад клік мишки чи прокрутка веб-сторінки.

Третя ланка – веб-клієнт, представляє собою будь-який браузер, який використовується клієнтом для роботи з системою. Коли користувач вводить в адресний рядок посилення, то веб-сервер відображає йому html-сторінку в вікні веб-браузера, де знаходиться відповідний інтерфейс рекомендаційної системи магазину квітів, який дозволяє йому взаємодіяти з системою та формувати запити до неї. Код інтерфейсу зберігається на стороні веб-сервера. Веб-сервер відповідає за зв'язок веб-браузера з клієнтом та СУБД.

Користувач надсилає запит до сервера, який обробляє запит та, за необхідності, звертається до сервера СУБД MySQL. Сервер оперує з даними та повертає результат на запит користувачеві до веб-клієнта, тобто браузера, яким користується клієнт.

### 3.3 Розробка бази даних системи електронної комерції для досліджень квітів

Облік персональних даних клієнтів, їх історії покупок та замовлень квітів, облік квітів на складі реалізується за допомогою п'яти базових таблиць і трьох підставкових, створених за допомогою інтегрованого середовища розробки

Workbench системи адміністрування баз даних платформи СУБД MySQL. EER-модель даних рекомендаційної системи квітниць подана на рис. 3.4

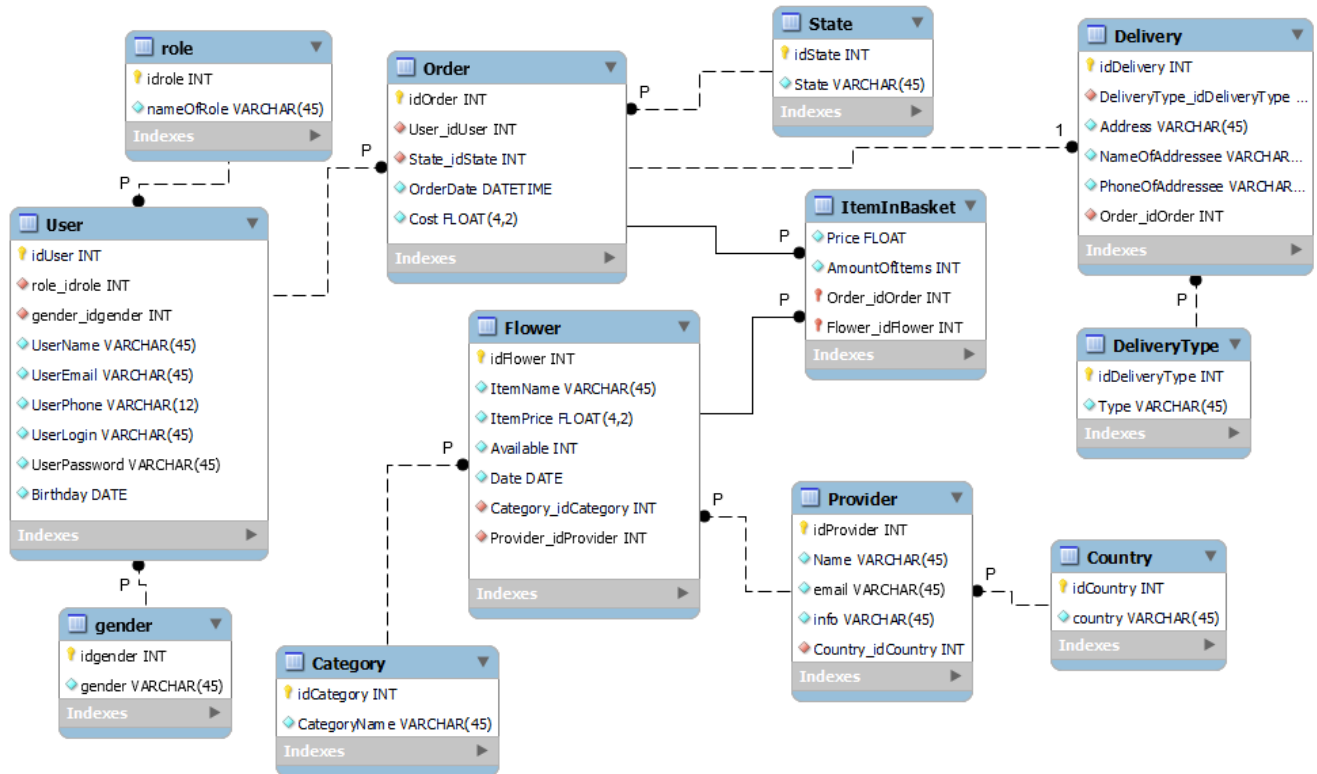


Рисунок 3.4 Структура таблиць бази даних е-системи продажу квітів

Були визначені такі сутності, атрибути та їх типи для бази даних е-системи з продажу квітів.

Були визначені такі таблиці, атрибути та їх типи для бази даних е-системи продажу квітів:

Базова таблиця «User» зберігає дані про користувачів е-системи продажу квітів. Для сутності «User» визначені такі атрибути:

- ключовий атрибут «idUser» (домен «числовий», тип «INT») – первинний ключ сутності «User».

- атрибут «gender\_idgender» (домен «числовий», тип «INT») – зовнішній ключ сутності «Gender», визначає стать користувача е-системи продажу квітів;

– атрибут «UserName» (домен «рядковий», тип «STRING») – ім'я користувача е-системи продажу квітів;

– атрибут «UserEmail» (домен «рядковий», тип «STRING») – електронна пошта користувача е-системи продажу квітів

– атрибут «UserPhone» (домен «рядковий», тип «STRING») – номер телефону користувача е-системи продажу квітів

– атрибут «UserLogin» (домен «рядковий», тип «STRING») – логін користувача е-системи продажу квітів;

– атрибут «UserPassword» (домен «рядковий», тип «STRING») – пароль користувача е-системи продажу квітів;

– атрибут «Birthday» (домен «дата», тип «DATETIME») – дата народження користувача е-системи продажу квітів.

Базова таблиця «Order» зберігає інформацію про е-системи продажу квітів. Для сутності «Order» визначені такі атрибути:

– ключовий атрибут «idOrder» (домен «числовий», тип «INT») – первинний ключ сутності «Order»;

– атрибут «User\_idUser» (домен «числовий», тип «INT») – зовнішній ключ сутності «User», визначає користувача е-системи продажу квітів, який створив замовлення.

– атрибут «orderDate» (домен «дата», тип «DATETIME») – описує дату створення замовлення е-системи продажу квітів;

– атрибут «Cost» (домен «числовий», тип «FLOAT») – визначає вартість е-системи продажу квітів;

Сутність «Role» зберігає роль користувача. Для сутності «Role» визначені такі атрибути:

– ключовий атрибут «idRole» (домен «числовий», тип «INT») – первинний ключ сутності «Role»;

– атрибут «idRole» (домен «рядковий», тип «STRING») – визначає роль користувача е-системи продажу квітів.

Сутність «Sate» зберігає стан замовлень е-системи продажу квітів. Для сутності «Sate» визначені такі атрибути:

– ключовий атрибут «idState» (домен «числовий», тип «INT») – первинний ключ сутності «State»;

– атрибут «state» (домен «рядковий», тип «STRING») – описує стан замовлення е-системи продажу квітів.

Базова таблиця «Flower» описує товар е-системи продажу квітів. Для сутності «Flower» визначені такі атрибути:

– ключовий атрибут «idFlower» (домен «числовий», тип «INT») – первинний ключ сутності «Flower».

– атрибут «Category\_idCategory» (домен «числовий», тип «INT») – зовнішній ключ сутності «Category»;

– атрибут «Provoder\_idProvider» (домен «числовий», тип «INT») – зовнішній ключ сутності «Provider»;

– атрибут «ItemName» (домен «рядковий», тип «STRING») – описує назву е-системи продажу квітів;

– атрибут «ItemPrice» (домен «числовий», тип «FLOAT») – описує ціну товару е-системи продажу квітів;

– атрибут «Available» (домен «числовий», тип «INT») – описує кількість товару на складі е-системи продажу квітів;

– атрибут «Date» (домен «дата», тип «DATETIME») – описує дату отримання товару на склад е-системи продажу квітів;

Базова таблиця «Provider» зберігає дані про постачальника товарів е-системи продажу квітів. Для сутності «Provider» визначені такі атрибути:

– ключовий атрибут «idProvider» (домен «числовий», тип «INT») – первинний ключ сутності «Provider».

– атрибут «idCountry» (домен «числовий», тип «INT») – зовнішній ключ сутності «Country»;

– атрибут «Name» (домен «рядковий», тип «STRING») – назва компанії-постачальника е-системи продажу квітів;

– атрибут «Email» (домен «рядковий», тип «STRING») – електронна пошта постачальника товару е-системи продажу квітів;

– атрибут «Phone» (домен «рядковий», тип «STRING») – номер телефону постачальника товару е-системи продажу квітів;

– атрибут «Info» (домен «рядковий», тип «STRING») – додаткова інформація про постачальника товару е-системи продажу квітів.

Сутність «Delivery» зберігає дані про доставку замовлень е-системи продажу квітів. Для сутності «Delivery» визначені такі атрибути:

– ключовий атрибут «idDelivery» (домен «числовий», тип «INT») – первинний ключ сутності «Delivery»;

– атрибут «address» (домен «рядковий», тип «STRING») – описує адресу отримувача замовлення е-системи продажу квітів;

– атрибут «nameOfAddressee» (домен «рядковий», тип «STRING») – описує ім'я отримувача замовлення е-системи продажу квітів;

– атрибут «phoneOfAddressee» (домен «рядковий», тип «STRING») – описує номер телефону отримувача замовлення е-системи продажу квітів;

– атрибут «idOrder» (домен «числовий», тип «INT») – зовнішній ключ сутності «Order» е-системи продажу квітів.

Сутність «DeliveryType» зберігає типи доставок замовлень е-системи продажу квітів. Для сутності «DeliveryType» визначені такі атрибути:

– ключовий атрибут «idDeliveryType» (домен «числовий», тип «INT») – первинний ключ сутності «DeliveryType»;

– атрибут «type» (домен «рядковий», тип «STRING») – описує тип доставки замовлень е-системи продажу квітів.

Базова таблиця «ItemInBasket» зберігає товари додані користувачем е-системи продажу квітів. Для сутності «ItemInBasket» визначені такі атрибути:

- ключовий атрибут «Flower\_idFlower» (домен «числовий», тип «INT») – первинний ключ сутності «Flower»;
- ключовий атрибут «Order\_idOrder» (домен «числовий», тип «INT») – первинний ключ сутності «Order»;
- атрибут «amountOfItems» (домен «числовий», тип «INT») – кількість товару, додана до кошика користувачем;

Підставкова таблиця «Gender» зберігає стать користувача е-системи продажу квітів. Для сутності «Gender» визначені такі атрибути:

- ключовий атрибут «idGender» (домен «числовий», тип «INT») – первинний ключ сутності «Gender»;
- атрибут «gender» (домен «рядковий», тип «STRING») – стать користувача е-системи продажу квітів.

Підставкова таблиця «Category» зберігає дані про наявні категорії товарів у е-системи продажу квітів. Для сутності «Categorie» визначені такі атрибути:

- ключовий атрибут «idCategory» (домен «числовий», тип «INT») – первинний ключ сутності «Category».
- атрибут «categoryName» (домен «рядковий», тип «STRING») – назва категорії товару е-системи продажу квітів.

Підставкова таблиця «Country» зберігає інформацію про країну постачання товару до е-системи продажу квітів. Для сутності «Country» визначені такі атрибути:

- ключовий атрибут «idCountry» (домен «числовий», тип «INT») – первинний ключ сутності «Country».
- атрибут «country» (домен «рядковий», тип «STRING») – країна постачання товару до крамниці е-системи продажу квітів.

Для усіх зв'язків між таблицями бази даних рекомендаційної системи крамниці квітів обраний тип посилальної цілісності – «RESTRICT», але для зв'язку між таблицями «Order» та «ItemsInBasket» та таблицями «Flower» та «ItemsInBasket» для операцій «ON DELETE» та «ON UPDATE» вибрано «CASCADE» (табл. 3.1).. Тому що коли клієнт відмовляється від замовлення, адміністратор видаляє його, і інформація про замовлення повинна видалитись повністю. При видаленні замовлення з таблиці «Order» каскадом видаляються пов'язані з його зовнішнім ключом записи.

Таблиця 3.1 – Посилальна цілісність зв'язків між таблицями

Таблиця 1	Зовнішній ключ	Таблиця 2	Первинний ключ	Тип посилальної цілісності ON DELETE	Тип посилальної цілісності ON UPDATE
User	Role_idrole	Role	idrole	RESTRICT	RESTRICT
	Gender_idgender	Gender	idgender	RESTRICT	RESTRICT
Flower	Category_idcategory	Category	idcategory	RESTRICT	RESTRICT
	Provider_idprovider	Provider	idprovider	RESTRICT	RESTRICT
Provider	Country_idcountry	Country	idcountry	RESTRICT	RESTRICT
Order	User_iduser	User	iduser	RESTRICT	RESTRICT
	State_idstate	State	idstate	RESTRICT	RESTRICT
ItemsInBasket	Order_idOrder	Order	idOrder	CASCADE	CASCADE
	Product_idProduct	Product	idProduct	CASCADE	CASCADE

### 3.4 Дослідження варіантів реалізації методів на основі контенту (Content-Based)

Методи на основі контенту ефективно працює навіть тоді, коли не завжди доступна необхідна кількість оцінок. Крім того, система не знає, що

рекомендувати новому користувачеві і кому рекомендувати новий об'єкт. У цьому випадку доцільно використовувати інформацію про об'єкти.

Кожен об'єкт у будь-якій системі має назву, опис та ключові слова, які можна використовувати для створення рекомендацій. Наприклад, е-система книгарні. Користувачі системи мають профіль, де вказують інформацію про свої інтереси: жанри, автори тощо. Маючи дані та інформацію про книги можна зробити висновки, чи цікава конкретна книга користувачу. Міра схожості інтересу користувача до книги набуває двох можливих значень: 1 – якщо жанр книги вказаний у профілі користувача, 0 – в іншому випадку. Схожість двох книг вимірюється за допомогою порівняння. Наприклад, якщо  $keywords(V_i)$  – множина ключових слів книги  $V_i$ , порівняти її з книгою  $V_j$  можна за формулою коефіцієнта Дайса (2.1). Дісе набуває значень з відрізка  $[0,1]$  і показує, наскільки близькі дві множини. Для виділення ключових слів використовуються різні алгоритми, наприклад TF-IDF.

Вище розглянутий метод визначення рейтингу товарів за ключовими словами (з використанням статистичної міри «TF-IDF») призначений для продажу книжок. Порівнявши об'єкти книжка та квітка очевидно, що вибірка ключових слів для книжки значно більша, ніж для квітки. Для прикладу, розглянемо ключові слова для деяких книжок в таблиці 3.2.

Розглядаючи такі товари як квіти, для їх опису використовують загально визначені характеристики, таких як колір, розмір, кількість бутонів. Немає такого розмаїття ключових слів для квітів. Цей недолік є суттєвим для реалізації методу для продажу квітів – обмеження з визначення ключових слів, що призведе до хибних результатів. Тому даний метод недоцільно впроваджувати для обраної предметної області з обмеженим асортиментом квітів.

Використання методів аналізу транзакцій (Transaction-Based), зокрема методу визначення асоціацій (Data Mining) за аналізом історії покупок (кошиків замовлень) клієнтів теж не впроваджується.

Таблиця 3.2 – Ключові слова для деяких книжок

Назва	Автор	Жанр	Ключові слова
Таємна Історія	Донна Тартт	Психологічний трилер	Англія, студенти, друзі, смерть, антична історія, трагедія, підлітки, таємниця, бідний серед заможних.
Анна	Емі Оделл	Біографії і мемуари	Величні жінки, мода, історія успіху, бізнес-акули, історії компаній, життя справжніх людей, інтерв'ю, конкуренція, головний редактор, “Диявол носить Prada”.
Двір срібного полум'я	Сара Маас	Фентезі	Кохання, гордість, монстри, війна, дружба, королеви, інтриги, феєрі, небезпека, втрата сім'ї.

Це обумовлено тим, що в рекомендаційних системах з обмеженим асортиментом недостатньо обсягу даних для ефективного застосування методів дата-майнінгу, які вимагають великої кількості тренувальних даних для навчання моделей. Якщо певні товари в асортименті магазину продаються рідко, то рекомендаційні моделі можуть мати багато пропусків у даних, що призводить до низької точності рекомендацій. Коли користувачі недостатньо взаємодіють із системою або не залишають достатньо велику кількість оцінок або відгуків, то алгоритми дата-майнінгу можуть мати недостатньо інформації для створення точних рекомендацій. Схожий сценарій застосовується й у випадку продажу квітів – зазвичай клієнт купляє квіти не для власного використання, а на подарунок, тому оцінка товару часто стає для них неактуальним заняттям. Немає сенсу використовувати складні алгоритми пошуку асоціацій, якщо результат виявлення вподобань клієнтів (кількість придбаних квітів одного виду) можна відразу визначити запитом SELECT з угрупованням даних з таблиць «User», «Order», «Iteminbasket», «Flower» (рис. 3.6) – тобто квітки які клієнт вже замовив.

Приклади вибірок даних подаються в табл. 3.3 і табл. 3.4; На рис. 3.6 показано інтерфейс для адміністратора для перегляду цих даних.

Таблиця 3.3 – Вибірка з угрупованням для отримання інформації покупок усіх клієнтів

ID клієнта	Назва квітки	Кількість замовлень
Клієнт 1	Троянда	50
Клієнт 2	Тюльпан	47
...	...	...
Клієнт N	Троянда	35

Таблиця 3.4 – Вибірка з угрупованням для отримання інформації покупок одного клієнта

ID клієнта	Назва квітки	Кількість замовлень
Клієнт 1	Троянда	50
Клієнт 1	Тюльпан	5
...	...	...
Клієнт 1	Хризантема	1

Одним зі способів отримання даних про вподобання клієнтів є збереження історії перегляду квітів. Для цього використовується додаткова таблиця «Control». За допомогою функцій інтерфейсу, під час перегляду інформації про квіти, до таблиці «Control» додаються дата й час перегляду, ідентифікатори квітів та клієнтів (рис. 3.5).

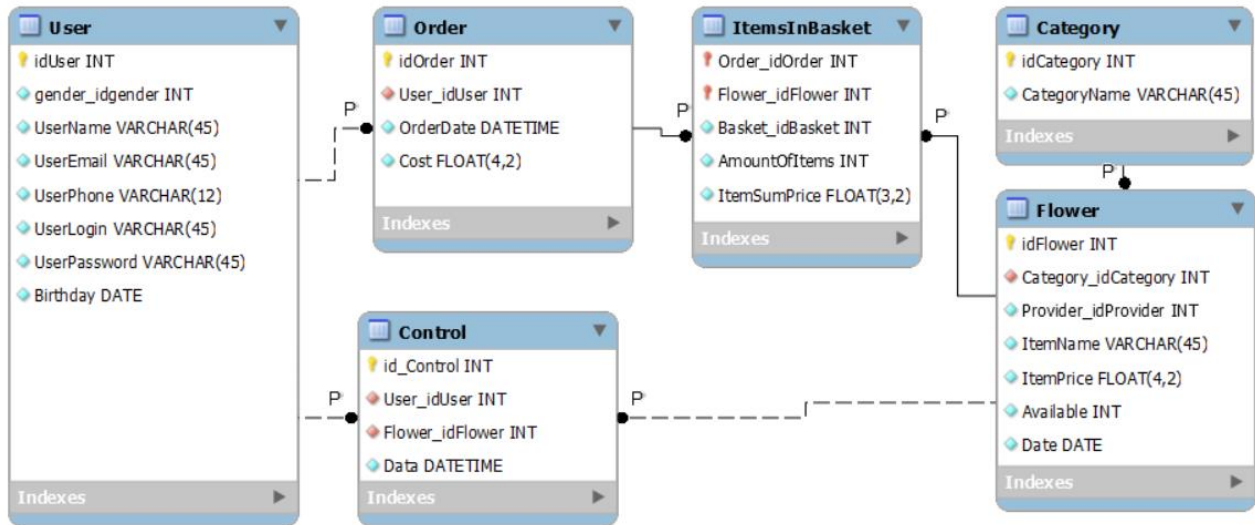


Рисунок 3.5 Зміна структури бази даних (додання таблиці «Control»)

За допомогою запиту до таблиці «Control» можна отримати повну інформацію або узагальнену інформацію (за допомогою запиту на угруповання) про пошуки квітів клієнтом (табл. 3.5). На рис. 3.7 показано інтерфейс для адміністратора для перегляду інформації з таблиці «Control».

Таблиця 3.5 – Вибірка для отримання інформації пошуків клієнта

ID клієнта	Назва квітки	Узагальнена кількість	АБО	ID клієнта	Дата	Час	Назва квітки
Клієнт 1	Троянда	15		Клієнт 1	23.11.23	21:15	Троянда
Клієнт 1	Тюльпан	5		Клієнт 1	23.11.23	21:15	Тюльпан
...	...	...		...			
Клієнт 1	Хризантема	1		Клієнт 1	18.11.23	21:15	Троянда

Головна>Користувачі>Клієнти>0325>Історія замовлень

Дата	Час	Назва товару	id товару	id замовлення
04.12.23	21:15	Троянда	001	001
04.12.23	21:16	Гортензія	012	001

Показати 11 на сторінку ▾

Рисунок 3.6 – Історія замовень користувача з id 0325

Головна>Користувачі>Клієнти>0325>Історія перегляду

Дата	Час	Назва товару	id товару
23.11.23	21:15	Троянда	001
23.11.23	21:16	Хризантема	015
23.11.23	21:19	Лілія	003
23.11.23	21:22	Гортензія	012
23.11.23	21:25	Гербера	021
23.11.23	21:28	Піон	013
23.11.23	21:34	Гладіолус	005
24.11.23	13:51	Троянда	001
04.12.23	16:01	Троянда	001
04.12.23	16:05	Хризантема	015
04.12.23	16:10	Гортензія	012

Показати 11 на сторінку ▾

Рисунок 3.7 – Історія перегляду товарів користувача з id 0325

На рис. 3.8 зображено інтерфейс головної сторінки е-системи з продажу квітів зі сторони клієнта з рекомендаціями на основі його попередньої переглядів та покупок.

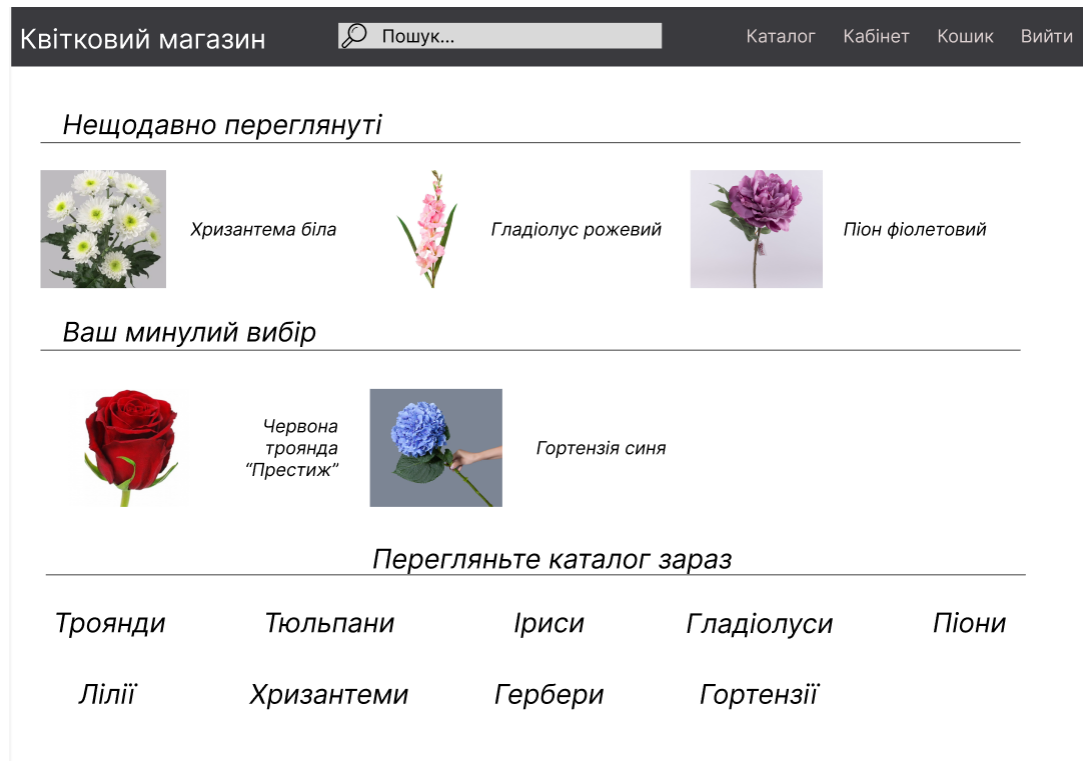


Рисунок 3.8 – Головна сторінка з рекомендаціями для клієнта

Використання методів подання даних про об'єкти, зокрема, створення профілів клієнтів і товарів, теж має особливості під час їх реалізації. За результатами аналізу визначено, що створювати окремі профілі квітів не має сенсу. Найкращий варіант – це реалізація уявлення (view), що містить інформацію історії покупок усіх клієнтів і об'єднує дані таблиць: «User», «Gender», «Order», «Iteminbasket», «Flower», «Category», «Provider».

Запити на вибірку, у тому числі і з угрупованням, до створеного уявлення (view) дозволяє створити профіль обраного клієнта або провести класифікацію клієнтів (створити профілі для груп клієнтів) за визначеними ознаками(представлено графічно на рис. 3.9):

– стать клієнта:

- а) жінка;
- б) чоловік;
- в) не зазначено.

– вік покупця:

- а) молодий – до 35 років;
- б) дорослий – з 36 до 60 років;
- в) літній – з 61 року.

– сума замовлення:

- а) дешеві – до 500 грн.;
- б) дорогі – більш ніж 500 грн..

– дата створення замовлення:

- а) святкові дні – дні загальних свят;
- б) звичайні дні – дні, під час яких немає загальних свят.

Рекомендації клієнту визначаються за аналізом статистики покупок.

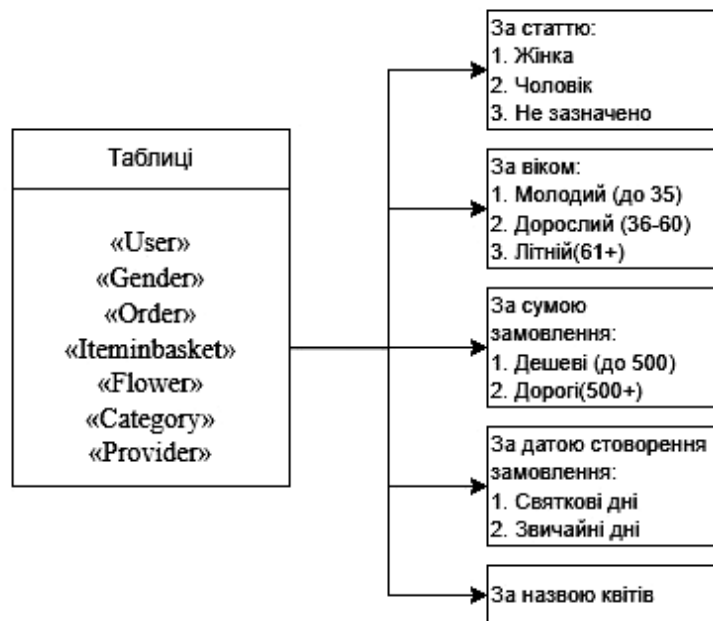


Рисунок 3.9 – Схема розподілення зібраної статистичної інформації про клієнтів та їх замовлення для уявлення за категоріями

Представлення «order\_info» (лістинг 3.1) створене для того, щоб відобразити інформацію про клієнта е-системи з продажу квітів та його замовлення. Це необхідно для збору статистичної інформації для подальшого аналізу майбутнього вибору клієнтів в залежності від його характеристик та уподобань.

Поля були вибрані з таких таблиць: mydb.`user`, mydb.`order`, mydb.`state`, mydb.gender, mydb.itemsinbasket, mydb.product, mydb.category, mydb.provider. Об'єднання таблиць було виконано за допомогою ключових полів вказаних вище таблиць.

### Лістинг 3.1 – Представлення «order\_info» е-системи з продажу квітів

```
CREATE VIEW order_info as
SELECT `user`.Name as 'Ім'я користувача',
       `user`.birthday as 'Дата народження', `user`.UserPhone as 'Мобільний
номер',
       `user`.UserEmail as 'Електронна пошта', gender.gender as 'Стать',
       `order`.idOrder as 'Номер замовлення',
       `order`.OrderDate as 'Дата замовлення', itemsinbasket.AmountOfItems as
'Кількість товарів',
       product.ItemName as 'Назва товару', product.ItemPrice as 'Ціна',
       category.categoryName as 'Категорія',
       provider.Name as 'Виробник'
FROM mydb.`user`, mydb.`order`, mydb.gender, mydb.itemsinbasket,
mydb.flower, mydb.category,
mydb.provider
where
(`order`.User_idUser = user.idUser) AND
(`order`.idOrder = itemsinbasket.Order_idOrder) AND
(product.idFlower = itemsinbasket.Flower_idFlower) AND
(provider.idProvider = product.Provider_idProvider) AND
(category.idCategory = product.Category_idCategory) AND
(`user`.gender_idgender= gender.idgender)$
```

На рис. 3.10 зображено отримання статистичних даних про усі замовлення клієнтів по категоріям. На рис. 3.11 зображено інтерфейс рекомендацій клієнту на основі класифікації.

Сторінка Адміністратора Відкрити сайт Вийти

Головна>Користувачі>Клієнти>Профілі клієнтів

Критерії, за якими потрібно отримати статистичну інформацію:

Frame 24

стать клієнта жінка ▾

вік покупця молодий ▾

сума замовлення дорогі ▾

дата замовлення звичайні дні ▾

назва квітів троянда ▾

[Отримати дані](#)

Результат:

Frame 13


	Ім'я	Стать	Назва товару	Ціна
0325	Аліна	жінка	Троянда	2 499
0325	Аліна	жінка	Гортензія	897
0326	Оксана	жінка	Гербера	699
0326	Оксана	жінка	Хризантема	765
0326	Оксана	жінка	Троянда	1367
0326	Оксана	жінка	Троянда	1367
0326	Оксана	жінка	Хризантема	765
0327	Катерина	жінка	Піон	630
0327	Катерина	жінка	Піон	630
0327	Катерина	жінка	Троянда	637
0327	Катерина	жінка	Троянда	637




Frame 14  
Показати 11 на сторінку ▾

Рисунок 3.10 – Інтерфейс статистичної інформації про замовлення клієнтів на основі класифікації для адміністратора

Квітковий магазин Пошук... Каталог Кабінет Кошик Вийти

Каталог>Троянди>Троянда "Престиж"



### Червона троянда "Престиж"

49 грн./шт. [До](#) [Додати до кошика](#)

Упаковка: [Без упаковки](#) [Крафт](#)

В наявності  ★★★★★ 5/5 4 відгука

Характеристика:

- Колір: Червоний
- Кількість бутонів на стеблі: 1
- Форма бутона: Шарова
- Висота стебля: 50 см.
- Аромат: Слабовиражений
- Сезонність: Несезнна
- Форма листків: Зубчаста, овальна

Перегляньте відгуки попередніх клієнтів:

Дякую, все було чудово. Рекомендую.

Рисунок 3.11 – Інтерфейс рекомендацій клієнту на основі профілю

### 3.5 Дослідження варіантів реалізації методів, основаних на знаннях (Knowledge-Based)

Методи, що засновані на знаннях працюють ефективніше за інші методи, коли покупка товарів відбувається нечасто, тобто система не має достатньо потрібної кількості інформації про товари, що купуються користувачем.

Для вирішення вище описаного завдання використовуються методи, що базуються на знаннях, їх можна поділити на дві групи:

- використання жорстких обмежень;
- вибір близьких об'єктів.

Ідея обох методів закладається у наступному: клієнти чітко формулюють свої вимоги до товару, для якого важко отримати оцінки, та система намагається знайти потрібний товар.

Інформація про товар надається у вигляді набору атрибутів. Для квітів це Колір бутона, кількість бутонів на стеблі, висота стебля, аромат, сезонність, форма листків. Наявні квіти в е-системі продажу квітів подаються в таблиці 3.6.

Для реалізації методу побудови запитів користувачів на основі знань про квіти до таблиці «Flower» додаються такі поля (рис. 3.12):

- «bud-colour» – колір бутона, може бути один чи декілька кольорів;
- «bud-shape» –форма бутона. Бутони квітів можуть мати різні форми в залежності від виду та видів рослин: «конусні» (лілії), «ліхтарні» (тюльпани), «плоскі» (гербера), «шарові» (троянда, півонія), «суцвіття» (гладіолус, гортензія);
- «num-flower» –можлива кількість бутонів на стеблі. Наприклад, на одній гілці хризантеми може бути 7-10 квітучих бутонів;
- «stem-height» – висота стебла, визначається у сантиметрах;
- «aroma» – наявність та інтенсивність аромату, може бути: «яскраво виражений», «слабо виражений», «без аромату»;

– «leaf-shape» – форма листів. Листочки у квітів можуть мати різні форми, розміри та текстури: «ланцетні» (тюльпан, лілія, ірис, гладіолус, хризантема), «периста» (півонії, гербера), «овальні» (гортензія), «зубчасті» (тройнда, гортензія);

– seasonality – сезонність. Сезонність квітів є характерною для конкретних періодів. Можуть бути сезонними («весна», «літо», «осінь», «зима») та не сезонними.

Таблиця 3.6 – Атрибути квітів в е-системі продажу квітів

Назва	Ціна	Колір бутона	Кількість бутонів на стеблі	Висота стебля	Аромат	Сезонність	Форма листків	Форма бутона
Троянди	40-120	білий, червоний, рожевий, жовтий, персиковий	1	до 90 см.	слабо виражений	несезонна	зубчаста, овальна	шарові
Тюльпани	35	червоний, жовтий, рожевий, білий	1	15-40см.	слабо виражений	весна	ланцетна	ліхтарні
Іриси	55	синій, пурпуровий, блакитний, фіолетовий, жовтий, помаранчевий, червоний, рожевий	1	до 60 см.	без аромату	весна	ланцетна	ліхтарні
Гладіолуси	50	білий, зелений, жовтий, кремовий, помаранчевий, малиновий, червоний, рожевий	5-8	40-50см.	слабо виражений	літо	ланцетна	сучвіття
Піони	80-210	бордовий, білий, рожевий	1	до 70 см.	яскраво виражений	весна	периста	шарові
Лілії	30-	білий,	1-5	до 70	яскраво	літо	ланцетна	конусні

	80	жовтий, рожевий, чорний, помаранчеви й		см.	виражени й		а	
Хризантеми	70-125	жовтий, червоний, білий, рожевий, зелений, помаранчеви й, синій	7-10	45-75 см.	яскраво виражени й	несезонн а	ланцетн а	плоскі
Гербери	40	помаранчеви й, жовтий, червоний, рожевий	1	до 45 см.	без аромату	несезонн а	периста	плоскі
Гортензії	200 - 320	білий, рожевий, синій, фіолетовий	1	65 см.	без аромату	літо- осінь	зубчаста , овальна	суцвітт я

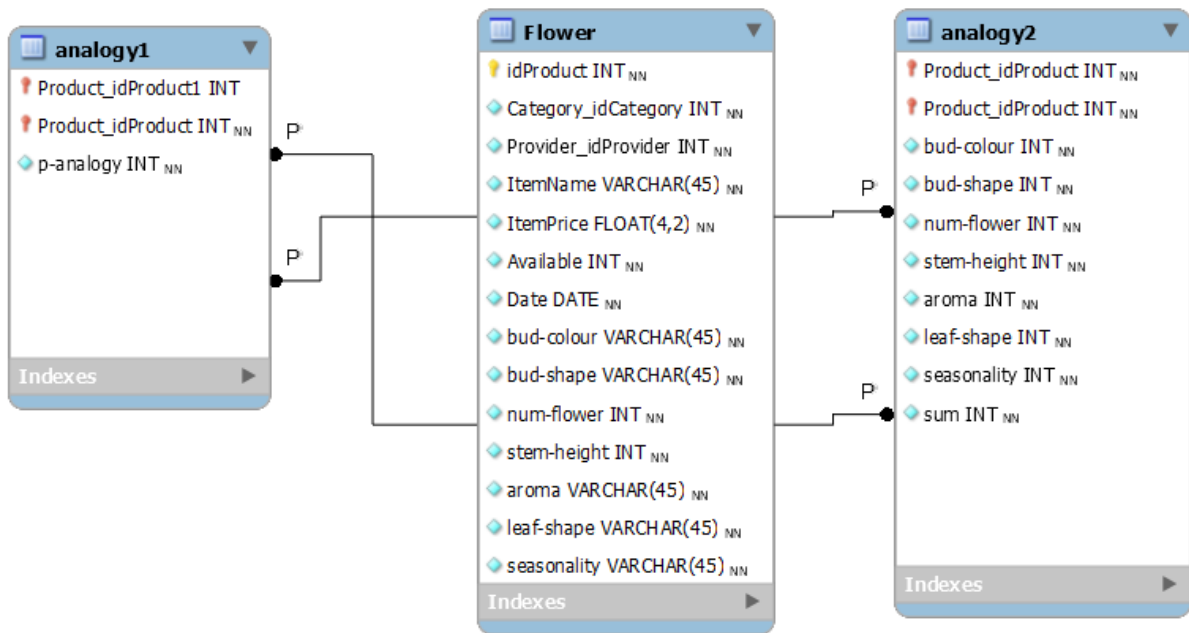


Рисунок 3.12 – Зміна структури бази даних (додання таблиць «Analogy»)

Крім того, до бази даних додається таблиця «Analogy1» або «Analogy2». Ці таблиці дозволяють заздалегідь визначити схожість квітів за визначеною шкалою (`Analogy1.p-analogy`) або визначити схожість квітів за кожним

параметром, з розрахунком підсумкової оцінки схожості («Analogy2.sum»). Розрахунки проводяться за формулами, описаними у пп. 2.4.2 вибір близьких об'єктів.

Упровадження параметрів опису квітів також дозволяє реалізувати метод використання жорстких обмежень. Для цього, за допомогою інтерфейсу, клієнту пропонується обрати параметри пошуку квітів (включаючи ціну) або обрати параметри пошуку за схожістю. Послідовність вибору параметрів визначає клієнт. Користувачі вказують не всі необхідні параметри, а лише ті, які вважаються найбільш суттєвими. Табл. 3.7 містить інформацію про те, у якій послідовності користувачі встановлювали атрибути.

Параметри передбачаються з урахуванням їхньої популярності. Перший передбачуваний параметр – найпопулярніший перший параметр серед попередніх користувачів (ціна). Якщо користувач вказав як перший параметр ціну, а як другий – колір бутона, то знаходиться найпопулярніший третій атрибут у рядках з номерами 1 і 2

Таблиця 3.7– Послідовність відповіді користувачів на питання системи

id	Атрибут 1	Атрибут 2	Атрибут 3	Атрибут 4
Клієнт 1	Ціна	Колір бутона	Аромат	Кількість бутонів
Клієнт 2	Ціна	Колір бутона	Кількість бутонів	Аромат
Клієнт 3	Колір бутона	Ціна	Аромат	Кількість бутонів
...	...	...	...	...
Клієнт m	Кількість бутонів	ціна	Колір бутона	Аромат

За результатом пошуку клієнт отримує список квітів, відфільтрований за визначеними параметрами. Якщо користувач вказав занадто жорсткі вимоги, то може виявитися, що таких товарів не існує.

Інтерфейс каталогу з доступним фільтрами для користувача зображено на рис. 3.13 для категорії квітів «Хризантеми». Обравши будь-який фільтр сторінка

автоматично оновиться показуючи доступний товар. На рис. 3.14 зображено інтерфейс каталогу товарів каталогу з обраним фільтром кольору бутона «Білий». Нижче відображеного відфільтрованого товару знаходиться список рекомендованого товару на основі обраного фільтру «Також можуть сподобатись».

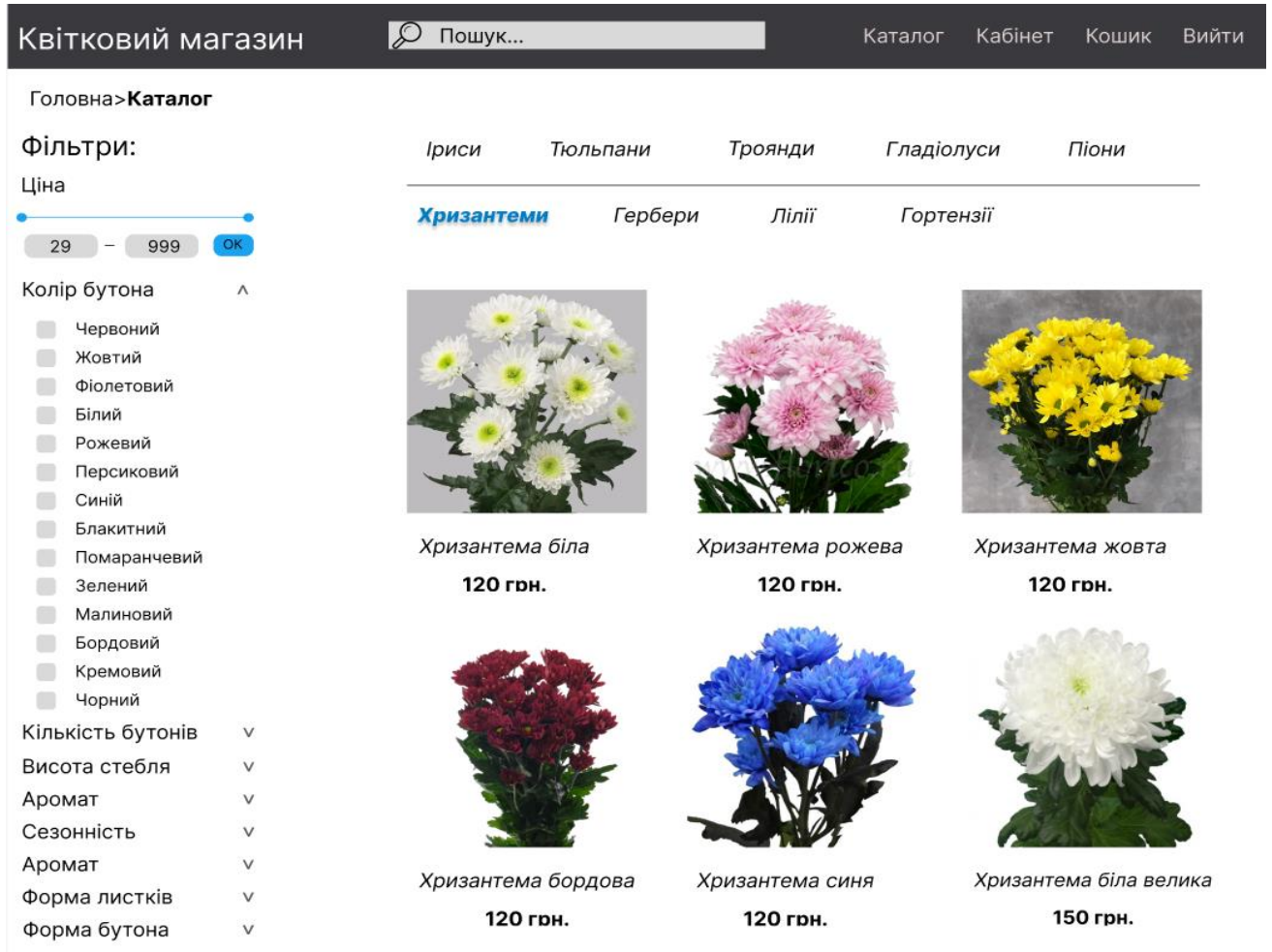


Рисунок 3.13 – Інтерфейс каталогу з фільтрами

Квітковий магазин  [Каталог](#) [Кабінет](#) [Кошик](#) [Вийти](#)

Головна > **Каталог**

Фільтри:

Ціна

29 – 999

Колір бутона

- Червоний
- Жовтий
- Фіолетовий
- Білий
- Рожевий
- Персиковий
- Синій
- Блакитний
- Помаранчевий
- Зелений
- Малиновий
- Бордовий
- Кремовий
- Чорний

Кількість бутонів

Висота стебля

Аромат

Сезонність


Аромат

Форма листків

Форма бутона


[Іриси](#) [Тюльпани](#) [Троянди](#) [Гладіолуси](#) [Піони](#)

**Хризантеми** [Герberi](#) [Лілії](#) [Гортензії](#)



Хризантема біла


**120 грн.**




Хризантема біла велика

**150 грн.**

*Також можуть сподобатись:*



Троянда біла



Гербера біла

Рисунок 3.14 – Інтерфейс каталогу з обраним фільтром кольору бутона «Білий» та відповідними рекомендаціями

### 3.6 Дослідження варіантів реалізації методів спільної фільтрації (Collaborative Filtering)

Метод кластеризації не підходить для створення рекомендацій у системах з обмеженим асортиментом товарів. На це є декілька причин:

– недостатньо кластерів: з обмеженого асортименту товарів – у нашому випадку квітів – важко створити достатню кількість значущих кластерів, що призведе до поганого відображення вимог користувачів;

– неоднорідність кластерів: товари всередині одного кластера будуть неоднорідними, що ускладнює точне рекомендування через можливе не врахування індивідуальних властивостей товарів;

– недостатньо персоналізації: кластеризація надає загальні рекомендації для всіх користувачів у межах певного кластера, не враховуючи індивідуальних вподобань та історії кожного конкретного користувача.

Для використання порівняння за методами «User-Based» і «Item-Based» в е-системі має бути реалізована модель (Model-Based) визначення рейтингових оцінок квітів клієнтами. Для цього до бази даних додається таблиця «Testimonial» (відгук), а до таблиці «Flower» – поле «average-rating» (рис. 3.9).

Таблиця «Testimonial» дозволяє зберегти текст відгуку на квітку (текстове поле «review») і визначити рейтинг (оцінку) за заданою шкалою (поле «rating»). Функція інтерфейсу е-системи підраховує значення середнього рейтингу квітки за всіма виставленими оцінками, зберігає його у полі «Flower.Average-rating» та подає під зображенням квітки під час перегляду інформації про неї.

Для реалізації методу порівняння користувачів (User-Based) розглядаються рейтингові оцінки товарів, які виставляють усі клієнти, тим самим висловлюючи свої особисті уподобання. Оціночна модель може бути сформована у вигляді тимчасової таблиці. Її дані отримуються за результатами запитів до таблиці «Testimonial» за допомогою SQL-оператора «Cursor» (рис. 3.15).

Під час генерації персоналізованої рекомендації обраному клієнту, обчислюються прогноз оцінки товару, який він не оцінив (приклад в табл. 3.8: клієнт 3 не оцінив квітку 2). Наше завдання – передбачити, яку оцінку поставив би клієнт 3 для квітки 2. Для прогнозу використовуються дані оцінок інших клієнтів, які за уподобаннями (за оцінками) найбільш близькі до нього.

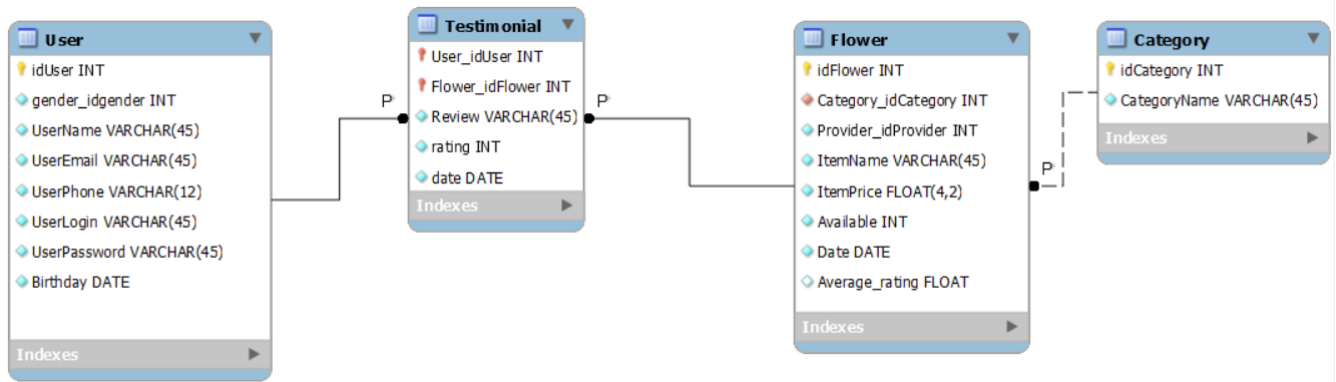


Рисунок 3.15 Зміна структури бази даних (додання таблиці «Testimonial»)

Таблиця 3.8 – Оціночна модель рейтингових оцінок

id	Квітка 1	Квітка 2	Квітка 3	Квітка 4	...	Квітка n
Клієнт 1	5	3	4	2	...	3
Клієнт 2	4	3	5	3	...	2
Клієнт 3	4	не визначена	2	1	...	4
...	...	...	...	...	...	...
Клієнт m	3	1	1	5	...	1

Алгоритм, що реалізує даний метод, реалізується за три кроки [1]:

- для кожного користувача  $u$  обчислимо, наскільки його інтереси збігаються з інтересами користувача  $a$ ;
- після цього виберемо множину користувачів, найбільш близьких до  $a$ ;
- передбачимо оцінку на основі оцінок квітці "сусідами" з попереднього кроку.

Крок 1. Оцінки обраного «клієнта  $a$ » попарно порівнюються з оцінками інших клієнтів. Міра збігу визначається розрахунками:

– косинуса кута між векторами:

$$sim(u, a) = \frac{\sum_{i=1}^m r_{a,i} \times r_{u,i}}{\sqrt{\sum_{i=1}^m r_{a,i} \times \sum_{i=1}^m r_{u,i}}}, \quad (3.1)$$

де « $r_{a,i}$ » – оцінка «квітки  $i$ », яка виставлена «клієнтом  $a$ » або «клієнтом  $u$ ».  $sim(u,a)$  набуває значення з відрізка  $[0,1]$ . Якщо користувач не вказав оцінку для якогось об'єкта, відповідне значення матриці дорівнює 0;

– коефіцієнта кореляції Пірсона:

$$sim(u, a) = \frac{\sum_{i \in I} (r_{a,i} - \bar{r}_a) \times (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i \in I} (r_{a,i} - \bar{r}_a)^2 \times (r_{u,i} - \bar{r}_u)^2}}, \quad (3.2)$$

де  $\bar{r}_a$  і  $\bar{r}_u$  – середні оцінки, що виставлені клієнтами « $a$ » і « $u$ ». Чим сильніше збігаються інтереси, тим ближче значення близькості до 1. Якщо коефіцієнт кореляції Пірсона негативний, то інтереси користувачів протилежні.

Крок 2. За визначеним порогом здійснюється відбір клієнтів « $u$ », які найбільш близькі за оцінками до клієнта « $a$ ». Для цього отриманий список клієнтів (множина  $K$ ) сортується за зменшенням значень  $sim(u, a)$ , а далі обмежується за допомогою SQL-оператора «Limit».

Крок 3. Для обраного користувача обчислюється прогноз – зважена середня оцінка товару на основі оцінок клієнтів (множина  $K$ ), що відібрані за кроком 2.

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u \in K} (r_{u,i} - \bar{r}_u) \times sim(a, u)}{\sum_{u \in K} |sim(a, u)|}, \quad (3.3)$$

де  $p_{a,i}$  – це очікувана оцінка користувача  $a$  для об'єкта  $i$ . До  $\bar{r}_a$  додається середнє відхилення від середньої оцінки інших користувачів з множини  $K$ .

Таким чином, описаний алгоритм передбачає оцінки для квітів, які поточний клієнт ще не оцінив. Для того, щоб зробити рекомендацію для даного клієнта,

достатньо передбачити оцінки для всіх неоцінених квітів і вибрати лише з найбільшою оцінкою.

На відміну від «User-Based», рекомендації за методом «Item-Based» генеруються шляхом порівняння схожості (збігу) оцінок рейтингу товарів, а не клієнтів. Алгоритм, що реалізує даний метод, реалізується за три кроки [1].

Крок 1. Оцінки обраного товару «i» попарно порівнюються з оцінками інших товарів «j», оцінених клієнтом «u». Міра збігу визначається розрахунками:

- косинуса кута між векторами (3.1);
- коефіцієнта кореляції Пірсона (3.2);

Крок 2. За визначеним порогом здійснюється відбір товарів, які найбільш близькі за оцінками. Для цього отриманий список товарів (множина K) сортується за зменшенням значення  $sim(i, j)$ , а далі обмежується за допомогою SQL-оператора «Limit».

Крок 3. Для обраного користувача «a» обчислюється прогноз – зважена середня оцінка товару на основі оцінок, що відібрані за кроком 2.

$$p_{a,i} = \frac{\sum_{j \in K} r_{u,j} \times sim(i,j)}{\sum_{j \in K} |sim(i,j)|}. \quad (3.4)$$

Для реалізації даного методу було створено систему збору відгуків. На рис 3.16 зображено інтерфейс для клієнта, який вже завершив замовлення, та перейшовши до особистого кабінету в пункті завершених замовлень він може обрати про який товар написати відгук, натиснувши кнопку «Залишити відгук про товар». Після цього відкривається форма, зображена на рис. 3.17, де клієнт може оцінити товар за п'ятибальною шкалою у форматі п'яти зірочок. А також за бажанням написати словесний відгук та натиснути кнопку «Відправити» для збереження рейтингу. Розрахувавши середній рейтинг, система відображає його на

сторінці опису товару, як показано на рис. 3.18. жовті зірочки та внизу сторінки відгуки попередніх клієнтів.

Квітковий магазин  [Каталог](#) [Кабінет](#) [Кошик](#) [Вийти](#)

Кабінет>Завершені замовлення>**Замовлення №001**

### Замовлення №001




	Червона троянда "Престиж" Кількість: 51 Упаковка: Без упаковки	2 499 грн. <b>Залишити відгук про товар</b>
	Гортензія синя Кількість: 3 Упаковка: Крафт	897 грн. <b>Залишити відгук про товар</b>

Рисунок 3.16 – Інтерфейс користувача для виконаного замовлення

Квітковий магазин  [Каталог](#) [Кабінет](#) [Кошик](#) [Вийти](#)

Кабінет>Завершені замовлення>Замовлення №001>**Відгук Червона троянда "Престиж"**

### Відгук

	Червона троянда "Престиж" Кількість: 51 Упаковка: Без упаковки
---	--

Оцініть цей товар: ☆☆☆☆☆

За бажанням, напишіть декілька слів, які описують отриманий вами товар.

**Відправити**

Рисунок 3.17 – Інтерфейс користувача форма для відгуку

Каталог&gt;Троянди&gt;Троянда "Престиж"



## Червона троянда "Престиж"

49 грн./шт.

Додати до кошика

Упаковка:

Без упаковки

Крафт

В наявності ✓

★★★★★ 5/5 4 відгука

Характеристика:

Колір: Червоний

Кількість бутонів на стеблі: 1

Форма бутона: Шарова

Висота стебля: 50 см.

Аромат: Слабовиражений

Сезонність: Несезнна

Форма листків: Зубчаста, овальна

Перегляньте відгуки попередніх клієнтів:

Дякую, все було чудово. Рекомендую.

Також можуть сподобатись:



Хризантема біла



Гладіолус рожевий



Піон фіолетовий

Рисунок 3.18 – Інтерфейс детальної інформації про товар з оцінкою і відгуком

### 3.7 Рекомендації до реалізації методів визначення рекомендацій у системах електронної комерції з обмеженим асортиментом

За проведеним дослідженням варіантів реалізації рекомендаційної функції в системах електронної комерції з обмеженим асортиментом можна зробити такі висновки.

Оновлена база даних (рис. 3.19) має 4 додаткові таблиці, та до таблиці «Flower» було додано 7 нових полів.

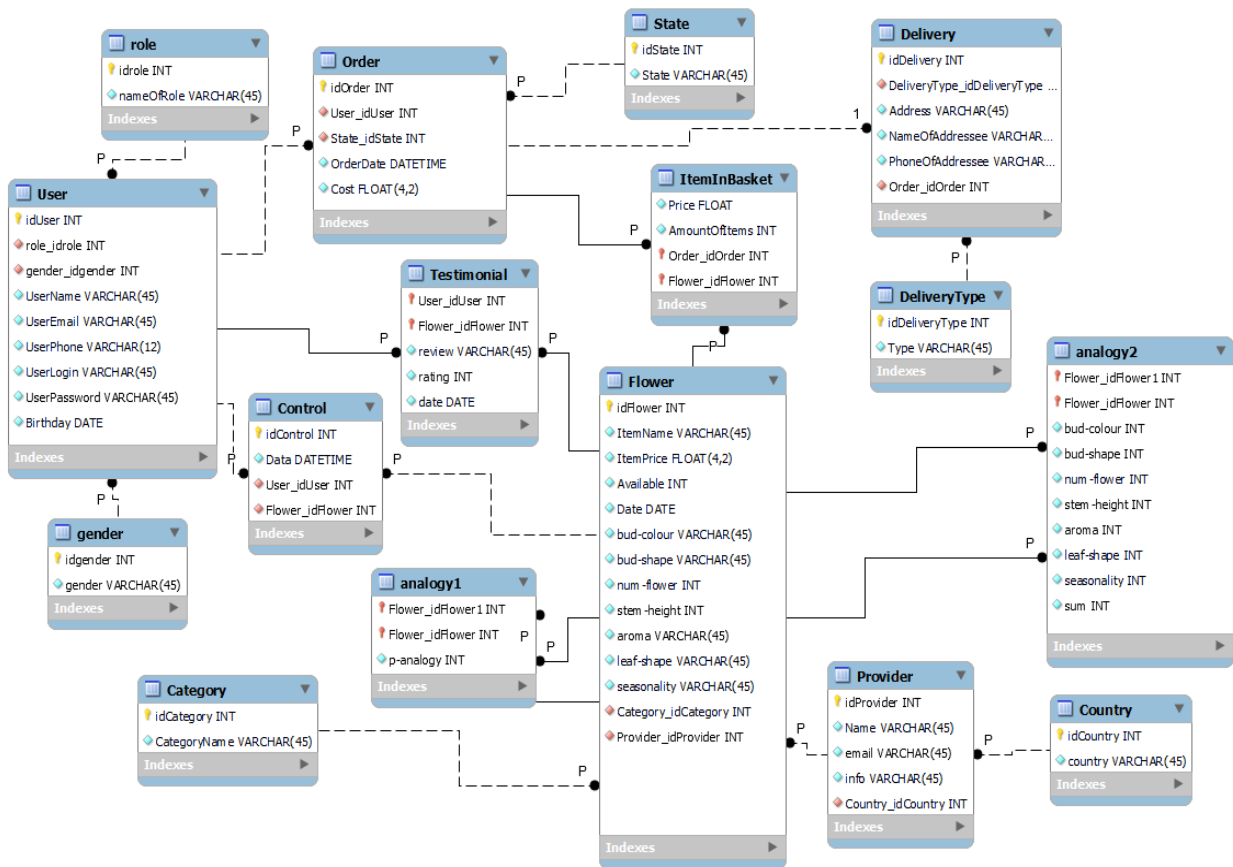


Рисунок 3.19 – Оновлена база даних

Таблиця «Testimonial» зберігає дані про відгуки клієнтів про товари е-системи продажу квітів. Для сутності «Testimonial» визначені такі атрибути:

- ключовий атрибут «Flower\_idFlower» (домен «числовий», тип «INT») – первинний ключ сутності «Flower»;

- ключовий атрибут «User\_idUser» (домен «числовий», тип «INT») – первинний ключ сутності «User»;

- атрибут «review» (домен «рядковий», тип «VARCHAR(45)») – відгук користувача е-системи продажу квітів;

- атрибут «review» (домен «рядковий», тип «VARCHAR(45)») – відгук користувача е-системи продажу квітів;

- атрибут «rating» (домен «числовий», тип «INT») – оцінка товару користувачем е-системи продажу квітів;

- атрибут «date» (домен «дата», тип «DATETIME») – описує дату створення відгуку клієнта е-системи продажу квітів.

Таблиця «Control» зберігає інформацію про товари, які переглянув конкретний користувач. Для сутності «Control» визначені такі атрибути:

- ключовий атрибут «idControl» (домен «числовий», тип «INT») – первинний ключ сутності «Control»;

- ключовий атрибут «Flower\_idFlower» (домен «числовий», тип «INT») – первинний ключ сутності «Flower»;

- ключовий атрибут «User\_idUser» (домен «числовий», тип «INT») – первинний ключ сутності «User»;

- атрибут «Data» (домен «дата», тип «DATETIME») – дата перегляду користувачем товару е-системи продажу квітів.

Таблиця «analogy1» зберігає дані про відгуки клієнтів про товари е-системи продажу квітів. Для сутності «analogy1» визначені такі атрибути:

- ключовий атрибут «Flower\_idFlower» (домен «числовий», тип «INT») – первинний ключ сутності «Flower»;

- ключовий атрибут «Flower\_idFlower1» (домен «числовий», тип «INT») – первинний ключ сутності «Flower»;

– атрибут «p-analogy» (домен «числовий», тип «FLOAT») – оцінка схожість квітів за визначеною шкалою е-системи продажу квітів.

Таблиця «analogy2» зберігає дані про відгуки клієнтів про товари е-системи продажу квітів. Для сутності «analogy2» визначені такі атрибути:

– ключовий атрибут «Flower\_idFlower» (домен «числовий», тип «INT») – первинний ключ сутності «Flower»;

– ключовий атрибут «Flower\_idFlower1» (домен «числовий», тип «INT») – первинний ключ сутності «Flower»;

– атрибут «bud-colour» (домен «числовий», тип «INT») – співпадає колір бутона порівнюваних квітів (1 – співпадає, 0 – ні);

– атрибут «bud-shape» (домен «числовий», тип «INT») – співпадає форма бутона порівнюваних квітів (1 – співпадає, 0 – ні);

– атрибут «num-flower» (домен «числовий», тип «INT») – співпадає можлива кількість бутонів на стеблі порівнюваних квітів (1 – співпадає, 0 – ні);

– атрибут «stem-height» (домен «числовий», тип «INT») – співпадає висота стебла порівнюваних квітів (1 – співпадає, 0 – ні);

– атрибут «aroma» (домен «числовий», тип «INT») – співпадає наявність та інтенсивність аромату порівнюваних квітів (1 – співпадає, 0 – ні);

– атрибут «leaf-shape» (домен «числовий», тип «INT») – співпадає форма листів порівнюваних квітів (1 – співпадає, 0 – ні);

– атрибут «seasonality» – співпадає сезонність порівнюваних квітів (1 – співпадає, 0 – ні).

До таблиці «Flower» було додано такі поля:

– атрибут «bud-colour» (домен «рядковий», тип «VARCHAR(45)») – колір бутона, може бути один чи декілька кольорів;

– атрибут «bud-shape» (домен «рядковий», тип «VARCHAR(45)») – форма бутона. Бутони квітів можуть мати різні форми в залежності від виду та видів

рослин: «конусні» (лілії), «ліхтарні» (тюльпани), «плоскі» (гербера), «шарові» (троянда, півонія), «суцвіття» (гладіолус, гортензія);

– атрибут «num-flower» (домен «числовий», тип «INT») – можлива кількість бутонів на стеблі. Наприклад, на одній гілці хризантеми може бути 7-10 квітучих бутонів;

– атрибут «stem-height» (домен «числовий», тип «INT») – висота стебла, визначається у сантиметрах;

– атрибут «aroma» (домен «рядковий», тип «VARCHAR(45)») – наявність та інтенсивність аромату, може бути: «яскраво виражений», «слабо виражений», «без аромату».

– атрибут «leaf-shape» (домен «рядковий», тип «VARCHAR(45)») – форма листів. Листочки у квітів можуть мати різні форми, розміри та текстури: «ланцетні» (тюльпан, лілія, ірис, гладіолус, хризантема), «периста» (півонії, гербера), «овальні» (гортензія), «зубчасті» (троянда, гортензія);

– атрибут seasonality (домен «рядковий», тип «VARCHAR(45)») – сезонність. Сезонність квітів є характерною для конкретних періодів. Можуть бути сезонними («весна», «літо», «осінь», «зима») та не сезонними.

За розглянутими методами рекомендаційних функцій прийнято рішення з реалізації:

– подання даних про об'єкти. За результатами аналізу визначено, що створювати окремі профілі для квітів не має сенсу. Найкращій варіант – це реалізація уявлення (view), що містить інформацію історії покупок усіх клієнтів і об'єднує дані таких таблиць: «User», «Gender», «Order», «Iteminbasket», «Flower», «Category», «Provider»;

- використання жорстких обмежень та вибір близьких об'єктів. До бази даних додається таблиця «Analogy1» або «Analogy2». Ці таблиці дозволяють заздалегідь визначити схожість квітів за визначеною шкалою (Analogy1.p-analogy) або визначити схожість квітів за кожним параметром, з розрахунком підсумкової

оцінки схожості («Analogy2.sum»). Клієнти чітко формулюють свої вимоги до товару, для якого важко отримати оцінки, та система намагається знайти потрібний товар. Інформація про товар надається у вигляді набору атрибутів. Параметри передбачаються з урахуванням їхньої популярності. Перший передбачуваний параметр – найпопулярніший перший параметр серед попередніх користувачів. Якщо користувач вказав занадто жорсткі вимоги, то може виявитися, що таких товарів не існує;

– методи обчислення подібності. Має бути реалізована модель визначення рейтингових оцінок квітів клієнтами. Для реалізації методу порівняння користувачів (User-Based) розглядаються рейтингові оцінки товарів, які виставляють усі клієнти, тим самим висловлюючи свої особисті уподобання. Під час генерації персоналізованої рекомендації обраному клієнту, обчислюються прогноз оцінки товару, який він не оцінив. Для прогнозу використовуються дані оцінок інших клієнтів, які за уподобаннями (за оцінками) найбільш близькі до нього. На відміну від «User-Based», рекомендації за методом «Item-Based» генеруються шляхом порівняння схожості (збігу) оцінок рейтингу товарів, а не клієнтів.

За розглянутими методами рекомендаційних функцій прийнято рішення не реалізовувати:

– зважування даних TF-IDF. Головний недолік є суттєвим для реалізації методу для продажу квітів – обмеження з визначення ключових слів, що призведе до хибних результатів. Тому даний метод недоцільно впроваджувати для обраної предметної області з обмеженим асортиментом квітів;

– метод кластеризації. Причинами є недостатня кількість кластерів та їх неоднорідність у системах з обмеженим асортиментом товарів що ускладнює точне рекомендування через можливе не врахування індивідуальних властивостей товарів що призводить до недостатнього рівня персоналізації, що суперечить бажані цілі.

## ВИСНОВКИ

За проведеним дослідженням варіантів реалізації методів рекомендацій в системах е-комерції з обмеженим асортиментом товарів визначено такі вхідні дані: персональні дані користувачів (вік, стать, місце проживання), історія взаємодії користувачів (перегляди, покупки, оцінки, пошуки, дії на сайті), інформація про товари (назва, опис, характеристики).

За розглянутими методами рекомендаційних функцій прийнято рішення з реалізації таких методів:

- подання даних про об'єкти. Найкращій підхід – це створення профілю клієнта. Знаючи інформацію про товари, куплені, переглянуті чи вподобані клієнтом, порадити схожі товари за допомогою реалізації уявлень (view), що містить інформацію історії покупок та дій усіх клієнтів;

- використання жорстких обмежень. Ідея методу: клієнти чітко формулюють свої вимоги до товару, для якого важко отримати оцінки, та система намагається знайти потрібний товар. Метод реалізується за допомогою упродовження параметрів опису квітів;

- вибір близьких об'єктів. Ідея методу: клієнти чітко формулюють свої вимоги до товару, система рекомендує товари, найближчі до вимог користувачів. У цьому випадку можлива рекомендація товарів, які не повністю відповідають вимогам;

- методи обчислення подібності. Для реалізації методу порівняння користувачів (User-Based) розглядаються рейтингові оцінки товарів, які виставляють усі клієнти. Під час генерації персоналізованої рекомендації обраному клієнту, обчислюються прогноз оцінки товару, який він не оцінив. Для прогнозу використовуються дані оцінок інших клієнтів, які за уподобаннями (за

оцінками) найбільш близькі до нього. На відміну від «User-Based», рекомендації за методом «Item-Based» генеруються шляхом порівняння схожості (збігу) оцінок рейтингу товарів, а не клієнтів.

За розглянутими методами рекомендаційних функцій прийнято рішення не реалізовувати такі методи:

– зважування даних TF-IDF. Головний недолік є суттєвим для реалізації методу для продажу квітів – обмеження з визначення ключових слів, що призведе до хибних результатів. Тому даний метод недоцільно впроваджувати для обраної предметної області з обмеженим асортиментом квітів;

– метод кластеризації. Причинами є недостатня кількість кластерів та їх неоднорідність у системах з обмеженим асортиментом товарів що ускладнює точне рекомендування через можливе не врахування індивідуальних властивостей товарів що призводить до недостатнього рівня персоналізації, що суперечить бажані цілі.

У рамках виконання кваліфікаційної роботи для наочного представлення використання рекомендованих методів було створено веб-застосунок з триланковою архітектурою «клієнт-сервер» для глобальної мережі інтернет.

Результати дослідження вказують на великий потенціал рекомендованих методів у покращенні користувацького досвіду та ефективності е-систем з обмеженим асортиментом товарів.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Дмитренко А.О. Дослідження особливостей реалізації методів визначення рекомендацій в е-системах з обмеженим асортиментом товарів // The 13th International scientific and practical conference «Development trends and improvement of old methods» (December 12–15, 2023) Warsaw, Poland. International Science Group. 2023. Pages: 401-411. ISBN 979-8-89238-616-6. DOI: 10.46299/ISG.2023.2.13.

2. Клюка О. Дослідження ринку електронної комерції в Україні [Електронний ресурс]. – Режим доступу: <http://ain.ua/2013/04/11/120835>.

3. Маловичко С.В. Аналіз сучасних тенденцій та динаміки розвитку електронної торгівлі на підприємствах України / С.В. Маловичко // Проблеми економіки. – 2015. – № 2. – С. 71-77.

4. D. Jannach, M. Zanker, A. Felfernig, G. Friedrich. Recommender Systems. An Introduction. *Cambridge University Press 32 Avenue of the Americas, New York, NY 10013-2473*. 2011. URL: [https://assets.cambridge.org/97805214/93369/frontmatter/9780521493369\\_frontmatter.pdf](https://assets.cambridge.org/97805214/93369/frontmatter/9780521493369_frontmatter.pdf).

5. Pasquale Lops, Marco de Gemmis, Giovanni Semeraro. Content-based Recommender Systems: State of the Art and Trends. *Recommender Systems Handbook. Springer New York*, 2020. P. 73–105. ISBN eBook: 978-0-387-85820-3. DOI: <https://doi.org/10.1007/978-0-387-85820-3>.

6. Аналіз ринку квітів в Україні: які квіти купують частіше і на які свята. / Сайт «Аналітика» компанії «Про-Консалтінг». URL: <https://pro-consulting.ua/ua/pressroom/analiz-rynka-cvetov-v-ukraine-kakie-cvety-pokupayut-chashe-i-na-kakie-prazdniki>.

7. CRM система, її компоненти, класифікація [Електронний ресурс] – Режим доступу до ресурсу <http://websekretar.chizh.ua/?p=548>.

8. Francesco Ricci, Lior Rokach, Bracha Shapira *Recommender Systems Handbook Second Edition Springer Science+Business Media New York 2011, 2015*.

9. X. Su, T.M. Khoshgoftaar; A Survey of Collaborative Filtering Techniques; *Advances in Artificial Intelligence*, 2009.
10. P. Melville, V. Sindhvani; Recommender systems; *Encyclopedia of Machine Learning*, 2010.
11. Greg Linden, Brent Smith, Jeremy York. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. IEEE Computer Society, 2003.
12. Jannach, M. Zanker, A. Felfernig, G. Friedrich; Recommender Systems. An Introduction; Cambridge University Press 32 Avenue of the Americas, New York, NY 10013-2473, USA, 2011; 352 p.
13. Robin Burke. Knowledge-based recommender systems. *Encyclopedia of Library and Information Science*. Department of Information and Computer Science University, URL: <https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day6/burke-elis00.pdf> .
14. Noor Ifada, Triyani Fatchur Rahman, Mochammad Kautsar Sophan. Comparing Collaborative Filtering and Hybrid based Approaches for Movie Recommendation. *2020 6th Information Technology International Seminar* URL: [https://www.researchgate.net/publication/348673145\\_Comparing\\_Collaborative\\_Filtering\\_and\\_Hybrid\\_based\\_Approaches\\_for\\_Movie\\_Recommendation](https://www.researchgate.net/publication/348673145_Comparing_Collaborative_Filtering_and_Hybrid_based_Approaches_for_Movie_Recommendation).
15. Bei-Bei CUI. Design and Implementation of Movie Recommendation System Based on Knn Collaborative Filtering Algorithm. ITM Web of Conference, DOI: 10.1051/ 71204008, 2017.
16. Noor Ifada, Syafrurrizal Naridho, Mochammad Kautsar Sophan. Multi-criteria based Item Recommendation Methods. *Journal of Science and Technology* URL: <https://doi.org/10.21107/rekayasa.v12i2.5913>
17. Phongsavanh Phorasim, Lasheng Yu. Movies recommendation system using collaborative filtering and k-means. *International Journal of Advanced Computer Research*, Vol 7(29), 2017. ISSN (Print): 2249-7277 ISSN (Online): 2277-7970 URL: <http://dx.doi.org/10.19101/IJACR.2017.729004> .

18. Rahul Pradhan, Ashish Chandra Swami, Akash Saxena, Vikram Rajpoot. A Study on Movie Recommendations using Collaborative Filtering. Rahul Pradhan et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1119 012018 URL: <https://iopscience.iop.org/article/10.1088/1757-899X/1119/1/012018> .

19. Noor Ifadaa, Irvan Syachrudina, Mochammad Kautsar Sophana, Sri Wahyunib. Enhancing the Performance of Library Book Recommendation System by Employing the Probabilistic-Keyword Model on a Collaborative Filtering Approach. 4th International Conference on Computer Science and Computational Intelligence 2019 (ICCSCI)

20. Chong Chen, Liya Zhang, Huan Qiao, Shihong Wang, Yuchu Liu, Xiaohong Qiu. Book Recommendation Based on Book-Loan Logs. *International Conference on Asian Digital Libraries*, 2012. URL: [https://www.researchgate.net/publication/293099285\\_Book\\_Recommendation\\_Based\\_on\\_Book-Loan\\_Logs](https://www.researchgate.net/publication/293099285_Book_Recommendation_Based_on_Book-Loan_Logs)

21. Токмаков, Г. П. Бази даних та знань. Проектування баз даних за технологією «клієнт-сервер» та розробка клієнтських додатків: Навчальний посібник/Г.П. Токмаков.- Ульяновськ; УЛГТУ, 2005. – 143 с.

22. Магазин косметики та засобів гігієни «makeup.ua» [Електронне джерело] – Режим доступу до ресурсу: <https://makeup.com.ua/>.