

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) \_\_\_\_\_

Розробка та дослідження методів формування рекомендацій в системах  
інтернет комерції \_\_\_\_\_

(тема)

Виконав:

студент 2 курсу, групи ІТІМ-20-1

Бойко В.І.

(прізвище, ініціали)

Спеціальність 6.050101 Комп'ютерні  
науки

(код і повна назва спеціальності)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма Інформаційні технології  
проектування

(повна назва освітньої програми)

Керівник доцент Хряпкін О.В.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри

\_\_\_\_\_

(підпис)

проф. Гребеннік І.В.

(прізвище, ініціали)

2021 р.

Харківський національний університет радіоелектроніки

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

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_  
(підпис)  
« \_\_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_ р.

**ЗАВДАННЯ**  
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові \_\_\_\_\_ Бойко В.І. \_\_\_\_\_  
(прізвище, ім'я, по батькові)

1. Тема роботи \_\_\_\_\_ *Розробка та дослідження методів формування рекомендацій в системах інтернет комерції* \_\_\_\_\_  
затверджена наказом університету від \_\_\_\_\_ 08 \_\_\_\_\_ 11 \_\_\_\_\_ 2021 р. № \_\_\_\_\_ 1663 Ст
2. Термін подання студентом роботи до екзаменаційної комісії \_\_\_\_\_ 20 \_\_\_\_ р.
3. Вихідні дані до роботи: *Функція: Дослідження та розробка рекомендаційної системи на прикладі предметної галузі оренди та продажу нерухомості. Форма діалогу: веб-сервіс. Перелік використовуваних програмних засобів: ОС Microsoft Windows 10, середовище розробки NetBeans IDE, CASE-застосунки: «Allfusion ErWin Data Modeler», «Allfusion BpWin Data Modeler», draw.io; СУБД MySQL*
4. Перелік питань, що потрібно опрацювати в роботі: *4.1 Вступ, 4.2 Аналіз предметної галузі, 4.2.1 Інформація та інформаційна система, 4.2.2 Основні поняття Е-комерції, 4.2.3 Огляд існуючих систем електронної комерції, 4.2.4 Рекомендаційні системи в галузі електронної комерції, 4.3 Огляд методів та технологій, які застосовуються в предметній області, 4.3.1 Опис процедури формування рекомендацій, 4.3.2 Етапи рекомендаційного процесу, 4.3.3 Методи фільтрації рекомендацій, 4.3.3.1 Фільтрація на основі вмісту \_\_\_\_\_, 4.3.3.2 Спільна фільтрація, 4.3.3.3 Рекомендаційні системи засновані на знаннях, 4.3.3.4 Система рекомендацій на основі популярності, 4.3.4 Існуючі методи формування рекомендацій в досліджуваній галузі, 4.4 Постановка задачі, 4.5 Розробка інформаційної технології вирішення задачі, 4.6 Проектування системи, 4.7 Підбір параметрів алгоритму, 4.8 Розробка програмного забезпечення, 4.9 Оцінка алгоритму формування рекомендацій*
5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій: *5.1 Концептуальна діаграма системи (1 Аркуш А4). 5.2 Діаграма декомпозиції системи (1 Аркуш А4). 5.3 Діаграма декомпозиції функції «оформлення ріелторської*

діяльності» системи (1 Аркуш А4). 5.4 Діаграма варіантів використання системи (1 Аркуш А4). 5.5 Оцінка важливості характеристик для короткострокової оренди (1 Аркуш А4). 5.6 Оцінка важливості характеристик для довгострокової оренди (1 Аркуш А4). 5.7 Оцінка важливості характеристик для купівлі/продажу (1 Аркуш А4). 5.8 Важливість впливу реакції користувача на оголошення (1 Аркуш А4). 5.9 Логічне моделювання системи (1 Аркуш А4). 5.10 Алгоритм формування рекомендацій (1 Аркуш А4). 5.11 Порівняння роботи алгоритмів формування рекомендацій(1 Аркуш А4).

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1 )

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Основна частина	доцент Хряпкін О.В.		

### КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на дипломну роботу	08.11.2021	виконано
2	Аналіз завдання, літератури та аналогів з теми атестаційної роботи	08.11 — 10.11.2021	виконано
3	Огляд методів та технологій, що застосовуються в досліджуваній області	11.11-14.11.2021	виконано
4	Розробка інформаційної технології	15.11-20.11.2021	
5	Проектування	21.11-26.11.2021	виконано
6	Підбір параметрів алгоритму	27.11-30.11.2021	
7	Розробка програми	01.12-10.12.2021	виконано
8	Оцінка алгоритму формування рекомендації	10.12-13.12.2021	виконано
9	Оформлення пояснювальної записки	13.12-14.12.2021	виконано
10	Оформлення графічної частини та презентаційних матеріалів	15.12.2021	виконано
11	Представлення роботи на рецензування	15.12.2021	виконано
12	Представлення роботи в ДЕК	17.12.2021	виконано

Дата видачі завдання \_\_\_\_\_ 20\_\_ р.

Студент \_\_\_\_\_ (підпис) Бойко В.І. (прізвище, ініціали)

Керівник роботи \_\_\_\_\_ (підпис) доцент Хряпкін О.В. (посада, прізвище, ініціали)

## РЕФЕРАТ

Робота містить: 106 сторінок, 45 рисунка, 4 таблиці, 3 додатки, 39 джерел. Графічна частина дипломної роботи містить 12 плакатів.

### РЕКОМЕНДАЦІЙНІ СИСТЕМИ, CONTENT-BASED FILTERING, COLLABORATIVE FILTERING, KNOWLEDGE BASED RECOMMENDERS

Метою кваліфікаційної роботи є дослідження існуючих методів формування рекомендацій в системах Інтернет комерції на предмет можливості їх використання у складних сферах, що мають не велику частоту переглядів та створення системи рекомендацій для таких предметних областей.

Об'єктом дослідження є веб-сервіси, що спеціалізуються на наданні послуг кінцевому користувачеві з оренди та продажу нерухомості

Предметом дослідження є методи та алгоритми формування рекомендацій, що застосовуються для прогнозування уподобань користувачів та надання пропозицій, на основі аналізу їх уподобань.

Галузь застосування – підтримка системи управління взаємовідносинами з клієнтами агентства нерухомості.

Досягнення мети кваліфікаційної роботи базується на комплексному використанні загальновідомих методів фільтрації інформації, що стосуються створення рекомендаційних систем в галузі оренди та продажу нерухомості.

Результатом кваліфікаційної роботи є результати порівняння методів формування рекомендацій в системах інтернет комерції на прикладі системи для оренди та продажу нерухомості, алгоритм формування рекомендацій, а також інформаційна система, що використовує розроблений алгоритм.

## ABSTRACT

Thesis contains: 106 pages, 45 images, 4 tables, 3 applications, 39 sources. Graphic part of the thesis contains 12 posters.

### RECOMMENDATION SYSTEMS, CONTENT-BASED FILTERING, COLLABORATIVE FILTERING, KNOWLEDGE BASED RECOMMENDERS

The purpose of the qualification work is to study the existing methods of forming recommendations in Internet commerce systems for the possibility of their use in complex areas that do not have a high frequency of views and create a system of recommendations for such subject areas.

The object of the study are web services that specialize in providing services to the end user for renting and selling real estate

The subject of the research is the methods and algorithms of forming recommendations, which are used to predict the preferences of users and provide suggestions, based on the analysis of their preferences.

Field of application - support of the real estate agency's customer relationship management system.

Achieving the goal of qualification work is based on the integrated use of well-known methods of filtering information related to the creation of recommendation systems in the field of rental and sale of real estate.

The result of the qualification work is the results of comparing the methods of forming recommendations in e-commerce systems on the example of the system for renting and selling real estate, the algorithm of forming recommendations, as well as information system using the developed algorithm.

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ, ТЕРМІНІВ

IDEF0 (Function modeling) - методологія функціонального моделювання і графічна нотація, призначена для формалізації і опису бізнес-процесів.

БД – база даних;

СУБД – система управління базами даних;

MVC (Model-View-Controller) – архітектурний шаблон;

UML (Unified Modeling Language) — уніфікована мова моделювання.

## ЗМІСТ

Вступ.....	9
1 Аналіз предметної галузі .....	12
1.1 Інформація та інформаційна система .....	12
1.2 Основні поняття Е-комерції .....	12
1.3 Огляд існуючих систем електронної комерції .....	13
1.4 Рекомендаційні системи в галузі електронної комерції.....	18
2 Огляд методів та технологій, які застосовуються в предметній області.....	23
2.1 Опис процедури формування рекомендацій .....	23
2.2 Етапи рекомендаційного процесу.....	24
2.3 Методи фільтрації рекомендацій .....	26
2.3.1 Фільтрація на основі вмісту .....	27
2.3.2 Спільна фільтрація .....	28
2.3.3 Рекомендаційні системи засновані на знаннях .....	32
2.3.4 Система рекомендацій на основі популярності .....	34
2.4 Існуючі методи формування рекомендацій в досліджуваній галузі ....	34
3 Постановка задачі.....	36
3.1 Визначення класу досліджуваної системи.....	36
3.2 Характеристика предметної області.....	37
3.3 Формалізована постановка задачі.....	37
3.4 Вибір стратегії для надання рекомендацій .....	38
3.5 Вихідна інформація .....	39
3.6 Вхідна інформація .....	39
4 Розробка інформаційної технології вирішення задачі .....	41
4.1 Розробка стратегії надання рекомендацій .....	41
4.2 Математичний опис задачі .....	44
4.3 Оцінка точності роботи рекомендаційної системи.....	50
5 Проектування системи.....	51
5.1 Опис системних вимог.....	51
5.2 Опис функціональних вимог .....	51
5.3 Розробка діаграми варіантів використання системи .....	55
5.4 Розробка вимог до інтерфейсу користувача інформаційної системи ..	56
6 Підбір параметрів алгоритму .....	58

	8
7 Розробка програмного забезпечення.....	65
7.1    Опис розробленої бази даних.....	65
7.2    Опис алгоритмів інформаційної системи.....	66
7.3    Розробка інтерфейсу клієнтської частини .....	70
8 Оцінка алгоритму формування рекомендацій.....	74
Висновки .....	81
Список літератури .....	83
Додаток А Графічні матеріали.....	87
Додаток Б Текст програми .....	101
Додаток В Відомість атестаційної роботи.....	118

## ВСТУП

В епоху цифровізації люди використовують онлайн-платформи, щоб знайти потрібні їм предмети та послуги. Ці платформи зазвичай мають величезний каталог товарів, що ускладнює їх пошук користувачам, які шукають лише короткий список бажаних елементів із багатьох інших невідповідних елементів. Це збільшило попит на рекомендаційні системи більше, ніж будь-коли раніше.

Рекомендаційні системи — це системи фільтрації інформації, які вирішують проблему інформаційного перевантаження [1] шляхом фільтрації фрагмента важливої інформації з великої кількості динамічно генерованої інформації відповідно до уподобань, інтересів або спостережуваної поведінки користувача. Система рекомендацій має можливість передбачити, чи надасть перевагу користувач конкретному продукту чи ні, на основі профілю користувача. Користувачі взаємодіють з цими платформами різними способами. Зазвичай вони висловлюють свої потреби чітко встановивши певні критерії пошуку, такі як географічне розташування, ціна та кількість спалень. Крім того, вони можуть взаємодіяти з деякими властивостями на веб-сайті: натискання посилань на власність, створення закладок або запит на отримання додаткової інформації. Усі ці дії допомагають системі краще визначити уподобання користувача та надати більш відповідні рекомендації.

Системи рекомендацій вигідні як постачальникам послуг, так і користувачам. Вони знижують витрати часу та зусиль на пошук і вибір товарів у системах інтернет комерції [2]. В умовах електронної комерції рекомендаційні системи збільшують дохід, оскільки вони є ефективним засобом продажу більшої кількості продуктів.

Наразі рекомендаційні системи використовуються майже в усіх типах систем. Їх використовують кіно- та музичні платформи, онлайн-продавці, агентства бронювання, маркетинг агентства, ринок нерухомості [3]. Дана робота розглянута на ринку нерухомості.

Потреба в житлі є однією з основних потреб кожної людини. Люди купують та орендують нерухомість не часто протягом усього терміну їх експлуатації, що

робить вибір нерухомості особливо складною процедурою прийняття рішень. Під час пошуку бажаної властивості, користувачі враховують кілька критеріїв і факторів прийняття рішення. Деякі з цих критеріїв представляють ключові фактори прийняття користувачів, такі як географічне розташування, ціна та розмір власність [4]. Інші представляють незначні фактори, які можуть мати менший вплив на вибір користувачів, наприклад, специфічні об'єкти, близькість до шкіл, громадського транспорту та ринків. Наразі зростає інтерес до використання систем інтернет-комерцій для пошуку об'єктів нерухомості, оскільки користувачі можуть вказати свої побажання та знайти відповідні об'єкти, які найкраще відповідають обраним критеріям серед багатьох інших нерелевантних.

Таким чином, неможливо переоцінити необхідність використання ефективних і точних методів рекомендацій у системі, яка забезпечуватиме релевантні та надійні рекомендації для користувачів.

Метою кваліфікаційної роботи є дослідження існуючих методів формування рекомендацій в системах Інтернет комерції на предмет можливості їх використання у складних сферах, що мають не велику частоту переглядів та створення системи рекомендацій для таких предметних областей.

Об'єктом дослідження є веб-сервіси, що спеціалізуються на наданні послуг кінцевому користувачеві з оренди та продажу нерухомості

Предметом дослідження є методи та алгоритми формування рекомендацій, що застосовуються для прогнозування уподобань користувачів та надання пропозицій, на основі аналізу їх уподобань.

Методи дослідження:

- аналіз;
- дедукція;
- аналогія;
- індукція;
- емпіричні: експериментальна перевірка ефективності розробленого
- програмного забезпечення;
- синтез.

Основні існуючі методи формування рекомендаційних систем не можуть надавати релевантні рекомендації в досліджуваній предметній області оскільки потребують якісних профілів користувача та матриці «елемент - характеристика». Оскільки майже всі оголошення на веб-сайті динамічні, тобто зазвичай вони залишаються там протягом короткого періоду часу. А відвідувачі порталу зазвичай мають лише 1 «покупку» [5]. Людина зазвичай купує лише один будинок або орендує одну квартиру на тривалий період часу, а якщо на коротку – в різних місцях чи містах. Після первинної покупки показувати подібні списки марно. Процес прийняття рішення про покупку будинку чи довший у порівнянні з покупкою телефону чи книги на Amazon [6]. Це означає, що може бути багато взаємодій, які не обов'язково призводять до покупки продукту. У більшості випадків, особливо на порталах нерухомості, люди не входять у систему, тобто немає довідкової інформації про користувача. Виходячи із вищесказаного, існує необхідність у розробці стратегій та способів прогнозування уподобань користувачів, які могли б працювати із предметною галуззю з оренди та продажу нерухомості.

Тобто, актуальною є розробка рекомендаційної системи, яка може пропонувати користувачу деякий контент, який буде вважатися цікавими та доцільними для користувача.

Робота пройшла апробацію на Міжнародній науковій конференції «Наука, освіта та суспільство в XXI столітті: наукові ідеї та механізми реалізації» (14 грудня 2021 року в м. Полтава, Україна).

## 1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

### 1.1 Інформація та інформаційна система

Наразі інформація є основним активом компаній, незалежно від ринку, на якому вони працюють, профілю чи розміру. Успіх підприємства в конкурентній грі багато в чому залежить від його здатності здобувати, зберігати та використовувати інформацію. Основними факторами, які сприяли підвищенню важливості інформації для підприємств, є швидка глобалізація, розвиток інформаційних технологій та зміни в поведінці клієнтів. Те ж саме стосується підприємств, що працюють на ринку електронної комерції [7]. Незалежно від їхнього розміру чи сфери діяльності, вони обробляють велику кількість інформації – навіть більше, ніж у традиційній торгівлі, оскільки вони працюють на віртуальних ринках. Це призводить до зростання інтересу до інформації та її можливого використання для створення конкурентних переваг в електронній комерції. Полегшення доступу до джерел інформації означає, що доступна інформація є більш різноманітною як за кількістю, так і за якістю. Ці зміни призводять до ускладнення відносин між окремими компаніями електронної комерції, що, у свою чергу, породжує проблеми, які можна назвати патологіями інформації. Таким чином, виникає необхідність сприймати інформацію як актив, і інформаційна логістика починає відігравати все більш важливу роль у підприємствах електронної комерції.

Інформаційні системи - це загальний термін, що використовується для опису різних мереж, обладнання та програмного забезпечення, які люди та компанії використовують для аналізу даних та їх використання [8]. Крім інтерпретації даних, інформаційні системи також збирають, обробляють та розповсюджують їх.

### 1.2 Основні поняття Е-комерції

Термін електронна комерція відноситься до бізнес-моделі, яка дозволяє компаніям і приватним особам купувати та продавати товари та послуги через

Інтернет. Електронна може здійснюватися за допомогою комп'ютерів, планшетів, смартфонів та інших електронних пристроїв. Для транзакцій електронної комерції доступні майже всі можливі продукти та послуги, включаючи книги, музику, квитки на літак, продаж та оренда нерухомості та фінансові послуги, такі як інвестування в акції та онлайн-банкінг [9]. Наразі електронна комерція може замінити звичайні магазини, хоча деякі підприємства вирішують підтримувати і те, і інше.

Електронна комерція працює у всіх чотирьох наступних основних сегментах ринку:

- бізнес для бізнесу (B2B), що є прямим продажем товарів і послуг між підприємствами;
- бізнес для споживача (B2C), який передбачає продажі між підприємствами та їхніми клієнтами;
- споживач споживачу, що дозволяє особам продавати один одному, як правило, через сторонній сайт;
- споживач для бізнесу, який дозволяє особам продавати товари підприємствам.

### 1.3 Огляд існуючих систем електронної комерції

Існує безліч популярних систем електронної комерції. Найвідомішими в Україні є:

- Amazon.com — одним із найкращих прикладів процвітаючого бізнесу електронної комерції, який спочатку почався як платформа роздрібною торгівлі, однак з плином часу; вони перенесли свою діяльність в Інтернет і стали однією з найбільших платформ електронної комерції в світі (рисунок 1.1). Тут ви взаємодієте з продавцем (Amazon), обмінюєтеся даними у вигляді фотографій, тексту, адреси для доставки тощо, а потім робите. Спочатку Amazon.com розглядався як інтернет-магазин книг [6]. Проте з часом вони розширили свій ринок у різних сферах бізнесу, додавши такі продукти, як електронні гаджети,

програмне забезпечення, DVD, відеоігри, музичні компакт-диски, MP3, одяг, взуття, товари для здоров'я тощо.

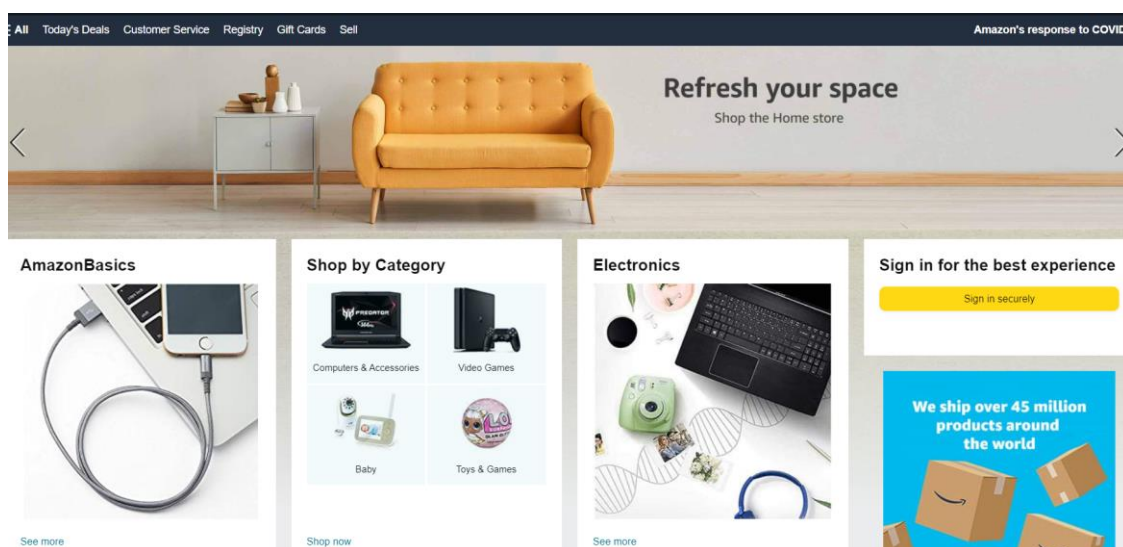


Рисунок 1.1 - Інтернет-магазин Амазон

– Rozetka.ua — український інтернет-магазин, що з'явився 2005 року. Станом на серпень 2020 року сайт посідає 7 місце серед найвідвідуваніших в Україні (рис 1.2).

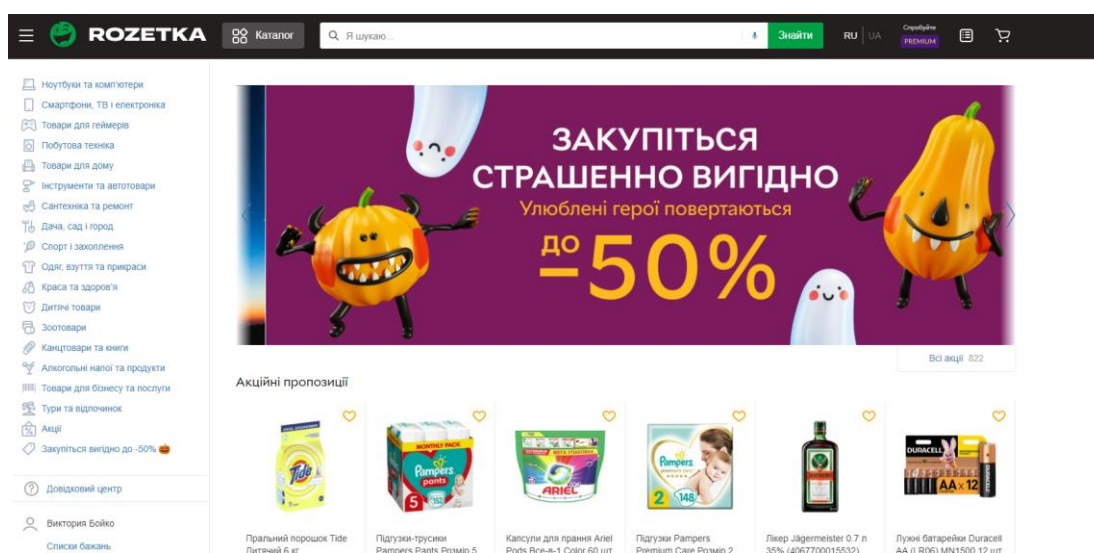


Рисунок 1.2 - Інтернет-магазин Розетка

Основні системи, що користуються популярністю в Україні та в світі для операцій з короткостроковою орендою є Booking та Airbnb. Системи, що об'єднує

всі 3 типи операцій (продаж, короткострокова та довгострокова оренда) - DomRia та OLX. Детальніше про них:

– Booking.com та Airbnb.com пропонують варіанти дозвілля, транспортні послуги та помешкання: від приватних помешкань до готелів і не тільки. Платформи мають великі бази даних (пропонують понад 28 мільйонів зареєстрованих помешкань). Структура сайту у них схожа: під час відкриття сайту основну увагу привертає форма фільтрування для пошуку необхідної нерухомості (куди вирушаємо, початкова та кінцева дати, кількість дорослих людей, кількість дітей та кількість номерів), далі – популярні категорії та підкатегорії для відпусток та вихідних [10]. На рисунку 1.3 показано як виглядає користувацький інтерфейс сайту Booking.

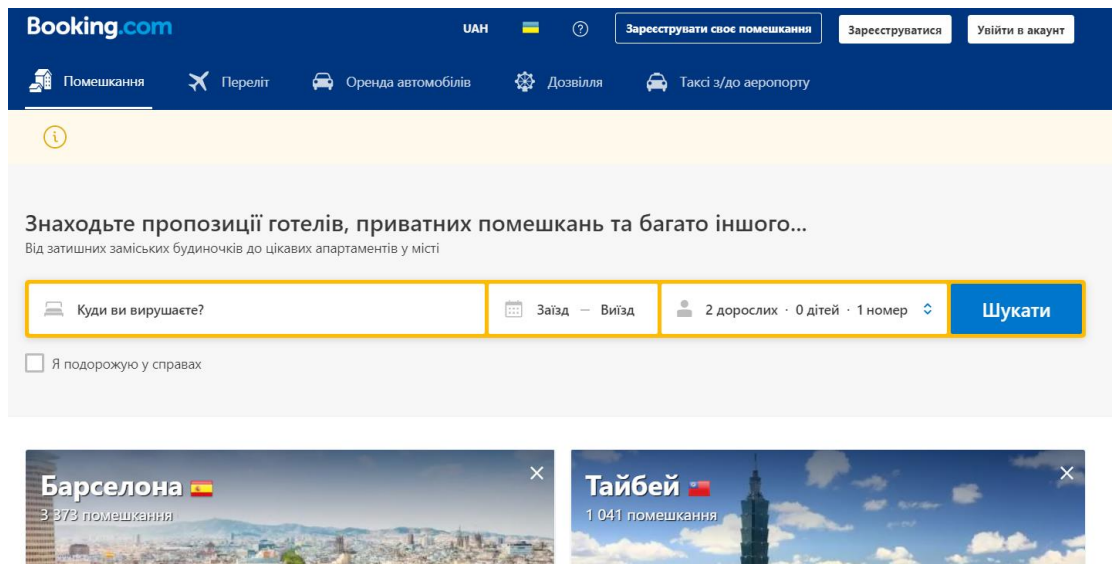


Рисунок 1.3 - Користувацький інтерфейс сайту Booking

Контент є якісним, постійно обновлюваним, охоплює велику кількість напрямків. Системи мають широкую функціональність: є можливість залишати коментарі та додавати відгуки про помешкання (але Booking.com не має можливості залишати відгуки про користувачі); зручна подача даних про інформацію та ціни, можливість зберегти нерухомість, що зацікавила у «зручності»; грамотне акцентування на важливій інформації, можливостях; доступна детальна фільтрація; займає багато часу; у Booking.com в особистому

кабінеті є список «Бажань», власних відгуків та історії бронювання; бронювання в цих системах займає багато часу [11]. На рисунку 1.4 показано як виглядає користувацький інтерфейс сайту Airbnb.

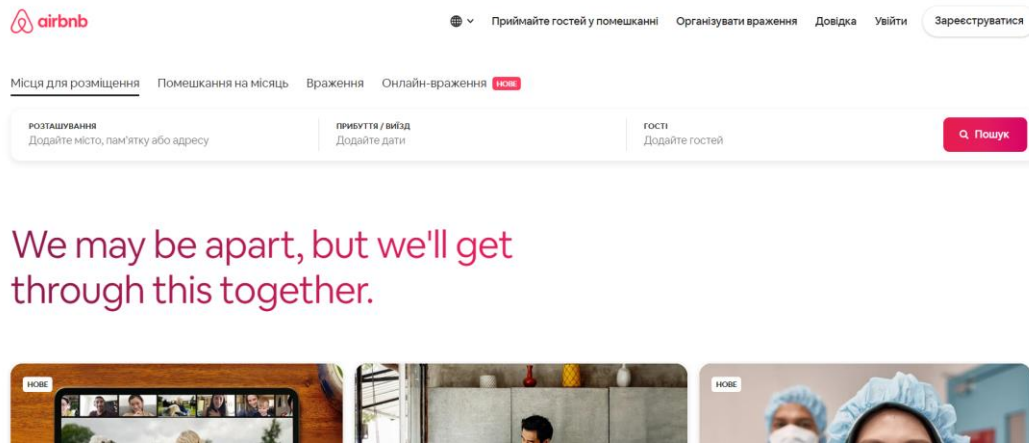


Рисунок 1.4 - Користувацький інтерфейс сайту Airbnb

– DomRia має у своєму функціоналі особистий кабінет та можливість додавати нерухомість у категорію «Обране», зображення приблизного місцезнаходження нерухомості на карті та можливість подати заявку на пошук нерухомості (рисунки 1.5). Фільтрування у DomRia є як по найрозповсюдженішим параметрам, так і за розширеними (більше 20 параметрів) [4]. Також DomRia дає можливість перегляду нерухомості у форматі 360, дає оцінку кожному власнику нерухомості (скільки часу працює з DomRia, скільки пропозицій має, скільки з них перевірили, чи перевірений банком, посилання на сторінку в соціальній мережі), можливість написати в чат до власника нерухомості, перевірити нерухомість або власника [12].

– Система «OLX» - чудовий приклад моделі електронної комерції C2C, це платформа онлайн-оголошень, яка об'єднує людей для продажу, покупки або обміну товарами та послугами. Одна з найпоширеніших платформ, проте має занадто простий фільтр: район та ціна (рис. 1.6). За такими критеріями користувачу буде необхідно самостійно продумати можливі маршрути (для роботи, університету, тощо) для кожної з пропозицій що задовільняють фільтр [13]. На

пошуки кращої пропозиції в даній системі користувач буде вимушений передивитись тисячі варіантів.

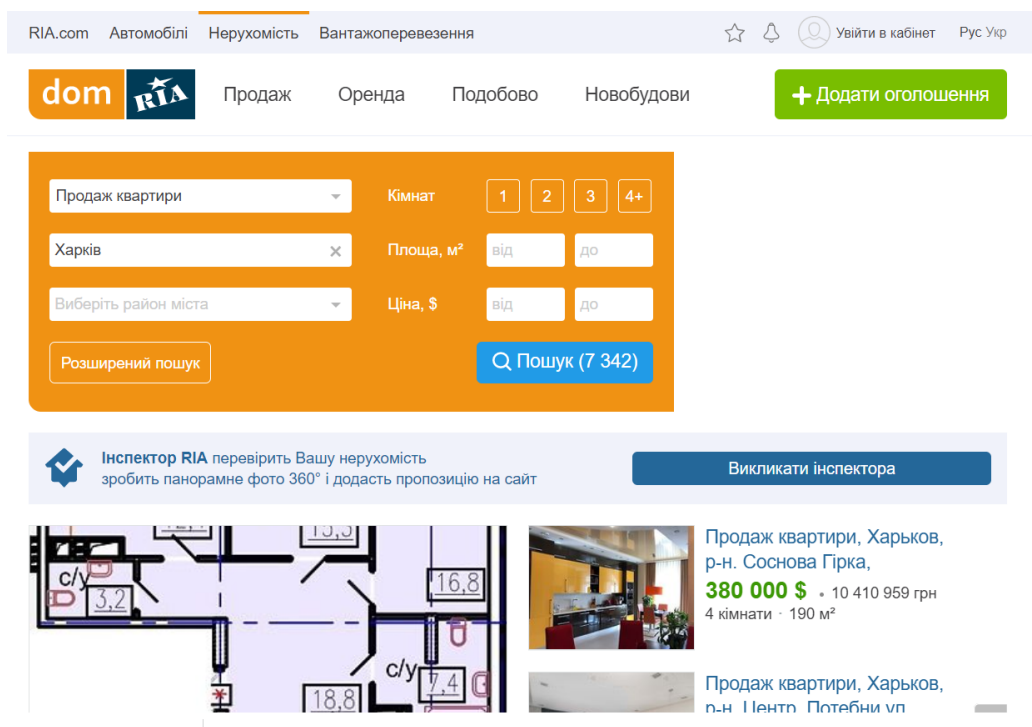


Рисунок 1.5 - Користувацький інтерфейс сайту Dom.ria

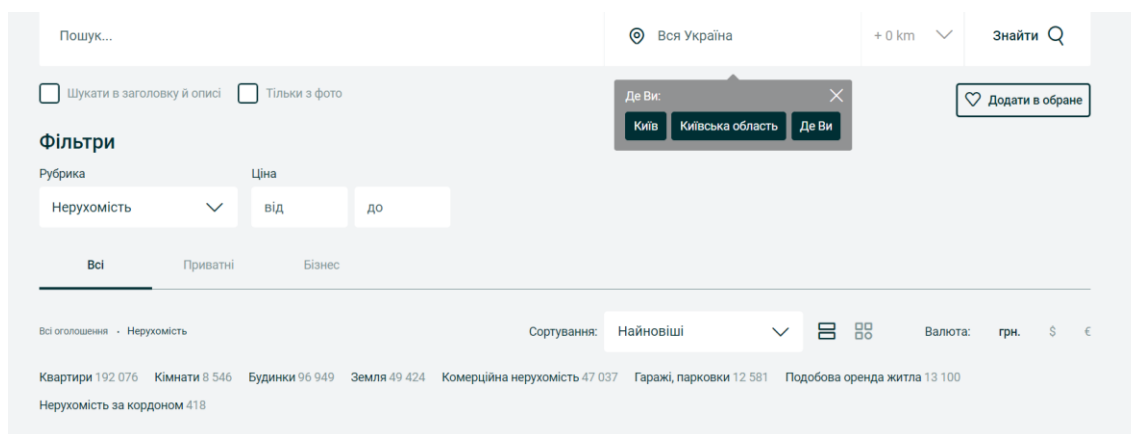


Рисунок 1.6 - Користувацький інтерфейс сайту ОЛХ

Рекомендації користувачам не доступні у всіх розглянутих системах.

#### 1.4 Рекомендаційні системи в галузі електронної комерції

Системи рекомендацій формують особливий тип техніки фільтрації інформації, яка намагається представити елементи інформації, які можуть зацікавити користувача. Як правило, система рекомендацій порівнює профіль користувача з деякими еталонними характеристиками і прагне передбачити «оцінку», яку користувач дасть елементу, який він ще не розглядав. Системи рекомендацій представляють уподобання користувачів з метою пропонування товарів для придбання або вивчення [14]. Вони стали фундаментальними застосуваннями в електронній комерції та доступі до інформації, надаючи пропозиції, які ефективно скорочують великі інформаційні простори, щоб користувачі були спрямовані на ті елементи, які найкраще відповідають їхнім потребам і вподобанням

Системи рекомендацій застосовують методи виявлення знань до проблеми створення рекомендацій щодо продуктів під час живої взаємодії з клієнтом. Сьогодні ці системи досягають широкого успіху в електронній комерції, особливо з появою Інтернету. Величезне зростання кількості клієнтів і продуктів ставить три ключові проблеми для рекомендаційних систем у сфері електронної комерції. Це: створення високоякісних рекомендацій, виконання багатьох рекомендацій в секунду для мільйонів клієнтів і продуктів, а також досягнення високого охоплення в умовах рідкості даних. Однією з успішних технологій системи рекомендацій є спільна фільтрація, яка працює шляхом узгодження вподобань клієнтів з іншими клієнтами під час створення рекомендацій.

При створенні профілю користувача розрізняють явні та неявні форми збору даних. Приклади явного збору даних включають таке: прохання користувача оцінити елемент за ковзною шкалою; просити користувача оцінити колекцію предметів від улюблених до найменш улюблених; представлення двох предметів користувачеві з проханням вибрати найкращий; просити користувача створити список предметів, які йому подобаються [15]. Приклади неявного збору даних включають наступне: спостереження за елементами, які користувач переглядає в

інтернет-магазині; аналіз часу перегляду товару/користувача; ведення обліку товарів, які користувач купує в Інтернеті; отримання списку елементів, які користувач слухав або дивився на своєму комп'ютері; аналізуючи соціальну мережу користувача та виявляючи схожі симпатії та антипатії. Система рекомендацій порівнює зібрані дані з аналогічними даними, зібраними від інших, і розраховує список рекомендованих елементів для користувача. Системи рекомендацій є корисною альтернативою алгоритмам пошуку, оскільки вони допомагають користувачам знаходити елементи, які вони, можливо, не знайшли самі.

На рисунку. 1.7 можна побачити принцип функціонування рекомендаційної системи. Користувач залишає інформацію явно або неявно, система фільтрує і зберігає її [15]. Збережені дані використовуються алгоритмами машинного навчання для тренування моделей, використовуючи які можна передбачити успішні рекомендації для користувачів.

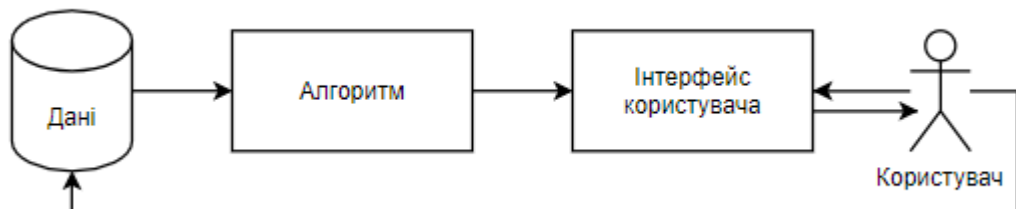


Рисунок 1.7 - Принцип роботи РС

Приклади популярних платформ, що використовують рекомендаційні системи:

- Netflix — «Фільми, які вам можуть сподобатися»;
- Google — «Візуально схожі зображення»;
- YouTube — «Відео, що може бути вам зацікавим» (рисунок 1.8) ;
- Facebook — «Люди, яких ви можете знати» (рисунок 1.9);
- Spotify — «Найкраща музика» (рисунок 1.10);
- LinkedIn — «Вакансії, які можуть вас зацікавити» (рисунок 1.11).

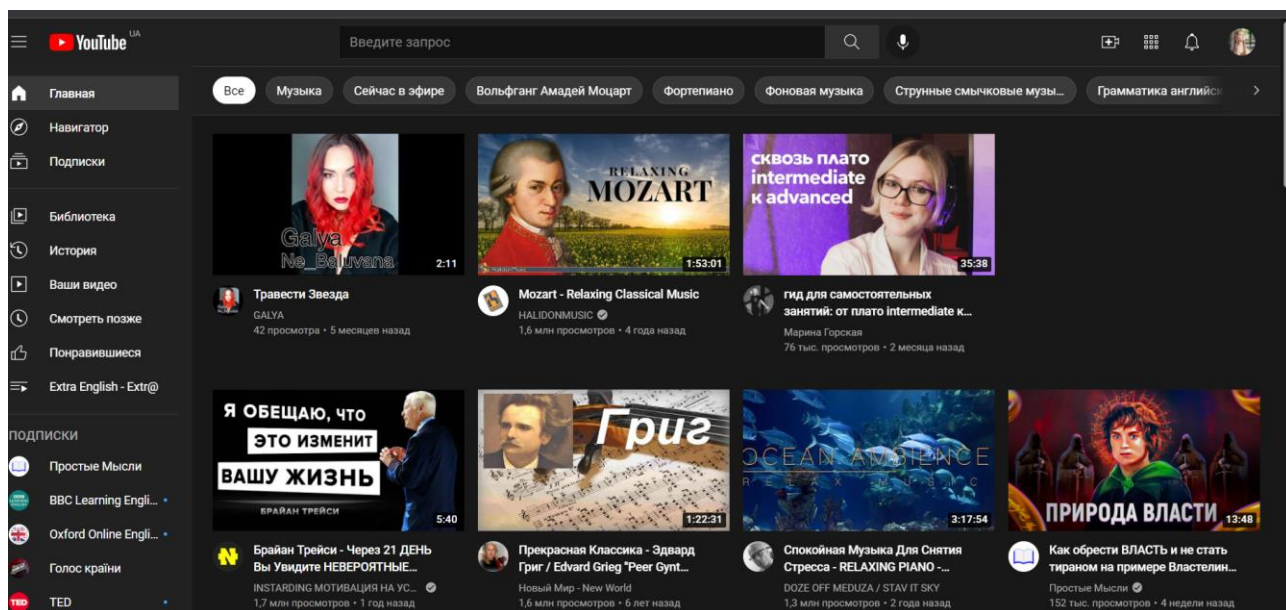


Рисунок 1.8 - Користувацький інтерфейс для рекомендації відео у YouTube

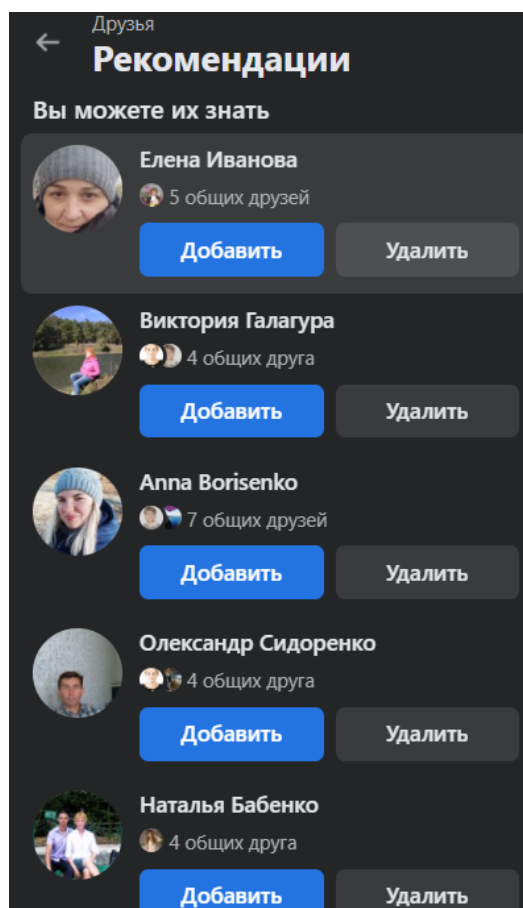


Рисунок 1.9 - Користувацький інтерфейс для рекомендації друзів у Facebook

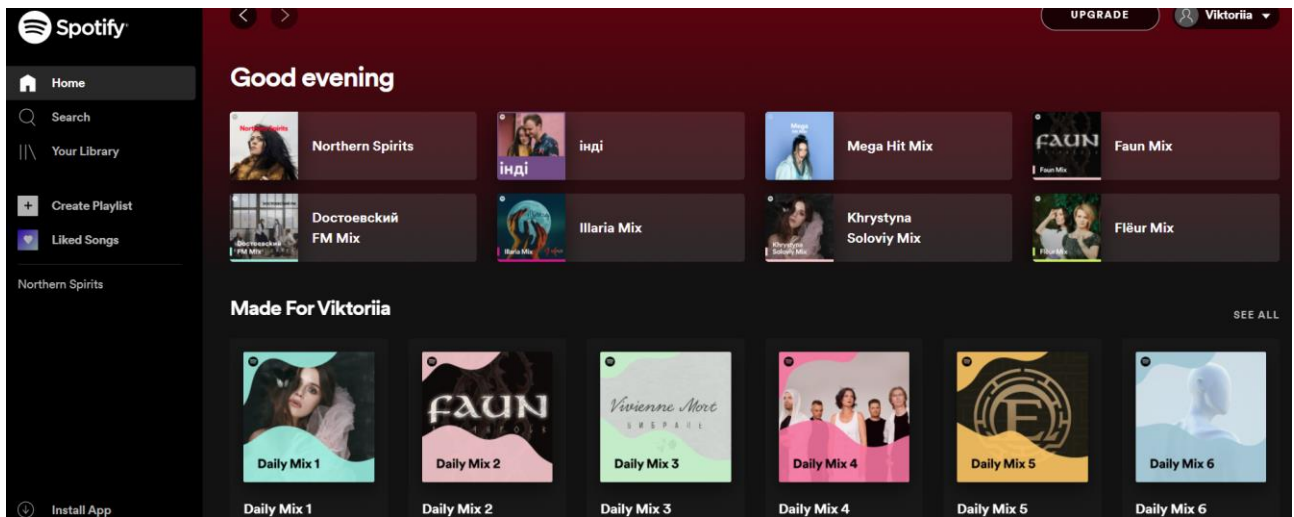


Рисунок 1.10 - Користувацький інтерфейс для рекомендації музики у Spotify

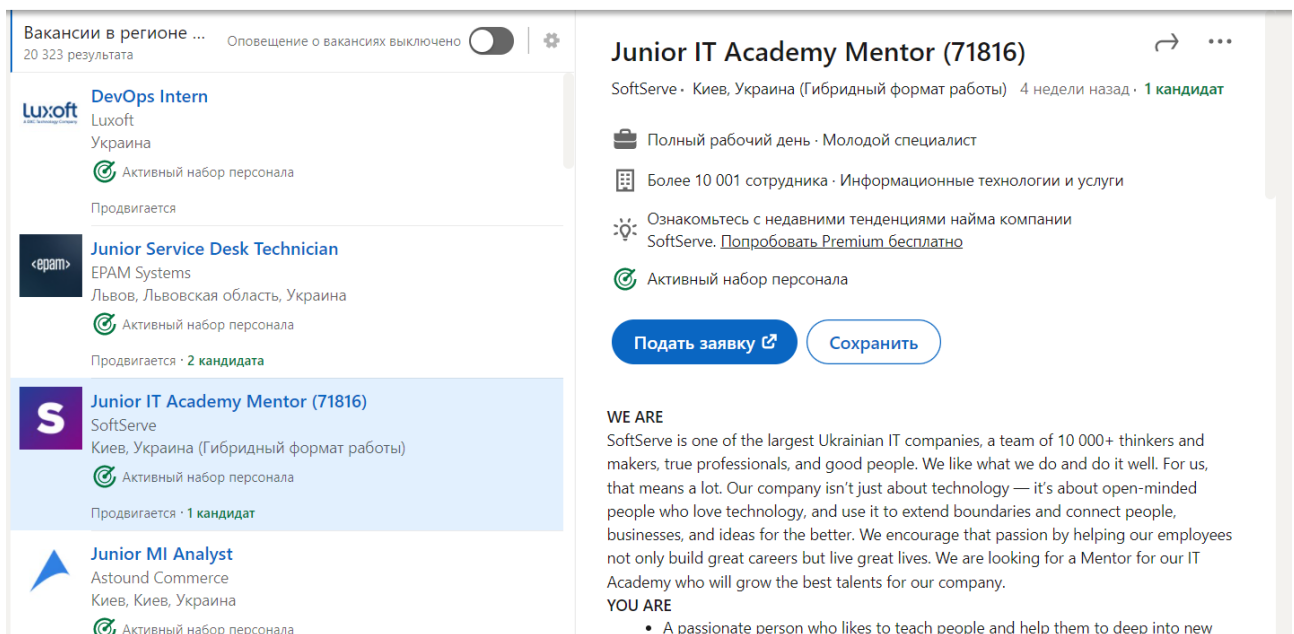


Рисунок 1.11 - Користувацький інтерфейс для рекомендації вакансій у LinkedIn

Зацікавлені сторони та їхні цілі складають одне ціле. Однак існують певні потенційні конфлікти між зацікавленими сторонами: наприклад вартість проти якості, капітальні витрати зараз чи в процесі експлуатації [14], естетика та комфорт проти вартості, навколишнє середовище проти потреб користувачів тощо (рисунок 1.12).

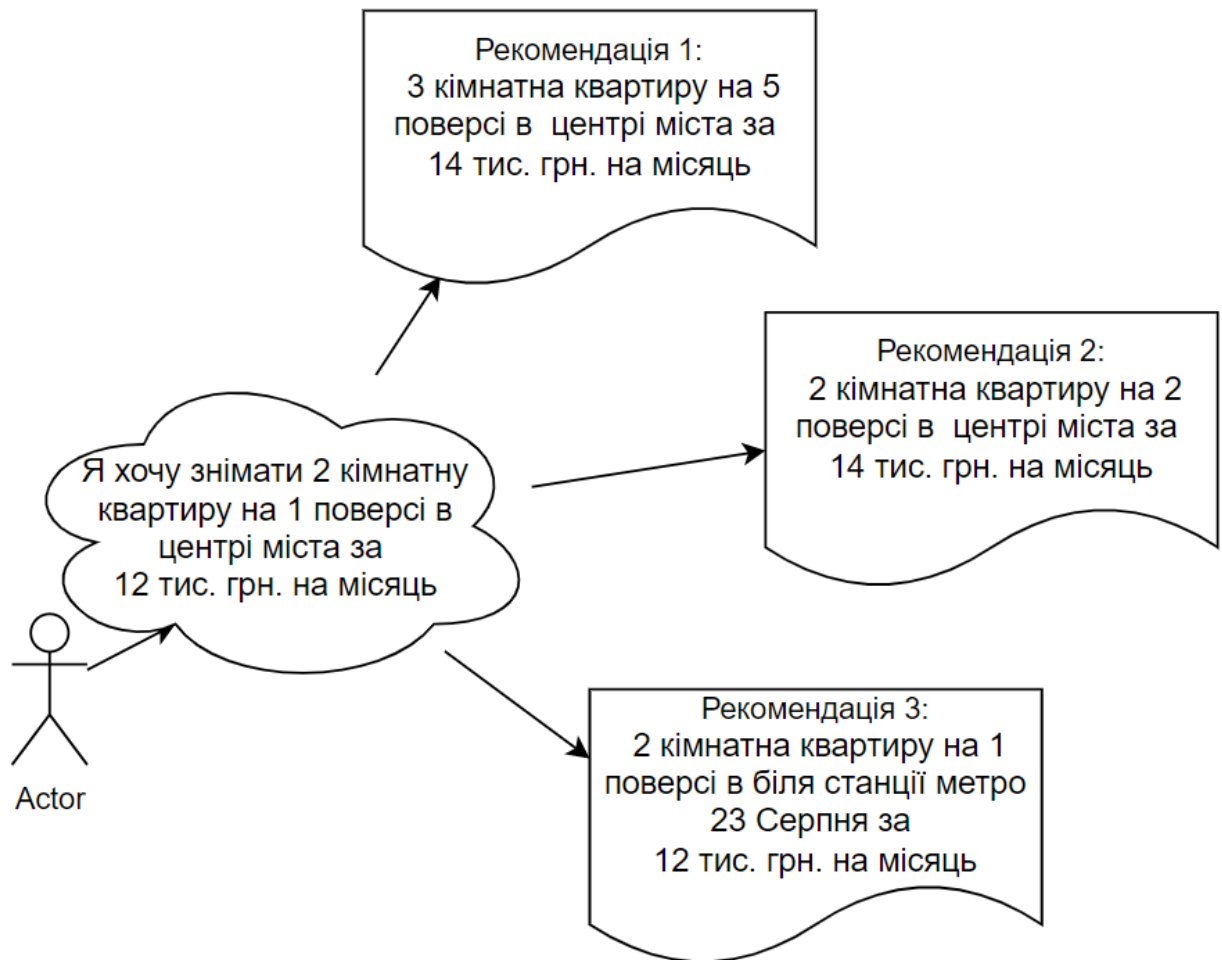


Рисунок 1.12 - Реальна взаємодія користувача із системою рекомендацій

Наведене вище зображення показує, що ідеальної рекомендації для користувача не буває. На зображенні користувач шукає квартиру 2-х кімнатну квартиру на 1 поверсі в центрі міста за 12 тис. грн. на місяць. Система рекомендувала користувачеві 3 квартири, що максимально задовільнять даний запит.

## 2 ОГЛЯД МЕТОДІВ ТА ТЕХНОЛОГІЙ, ЯКІ ЗАСТОСОВУЮТЬСЯ В ПРЕДМЕТНІЙ ОБЛАСТІ

### 2.1 Опис процедури формування рекомендацій

Для коректної роботи системи рекомендацій, потрібні початкові дані на яких будуть базуватися рекомендації. Перш за все це елементи всередині системи - оголошення (продукти або послуги) та персональні дані користувачів. Життєвий цикл роботи рекомендаційної системи починається з отримання даних про оголошення та користувачів, та закінчується наданням релевантних персональних рекомендацій користувачу оголошень на основі аналізу його вподобань [16]. Вибір методу формування рекомендацій в системах залежить від інформації, якою буде оперувати система. Зазвичай платформа надає велику кількість продуктів, які залишаються більш-менш незмінними протягом тривалого часу, інколи, деякі додаються на веб-сайт, а деякі видаляються. Більшість із них мають чітко визначені властивості. Відвідувачі платформи, зазвичай, здійснюють декілька покупок і взаємодій, особливо в довгостроковій перспективі. Наприклад, багато людей були клієнтами Amazon протягом багатьох років. Процес прийняття рішення про покупку телефону, книги, чи інших поширених речей досить короткий. Це означає, що за менший проміжок часу відбувається більше взаємодій, які призводять до покупки товару. Також, зазвичай, на подібних платформах у великої кількості людей вже є наявні облікові записи, що містять певні персональні дані.

На платформах, що оперують нерухомістю ситуація дещо складніша. Майже всі оголошення на веб-сайті динамічні, тобто зазвичай вони залишаються там протягом короткого періоду часу. За даними Zillow, середній час, необхідний для продажу будинку, становить від 65 до 93 днів [17]. А користувачі порталу зазвичай мають лише 1 «покупку». Людини не часто купують по декілька будинків та орендують одну нерухомість на досить тривалий період часу, а якщо на коротку (при подорожах чи відрядженнях) – в різних місцях чи містах. Тобто, після первинної покупки показувати подібні списки оголошень марно. Процес прийняття

рішення про покупку нерухомості доволі тривалий. Це означає, що користувач може зробити багато взаємодій з системою, які не обов'язково призводять до покупки/оренди нерухомості. Часто, люди не входять у систему, тобто вона не має первинної персональної інформації про користувача [14]. Даними, які використовує система є: переглянуті статті, зображення, кліки, прокрутки, час перебування на сторінці тощо.

## 2.2 Етапи рекомендаційного процесу

Рекомендаційний процес складається з трьох основних етапів:

- фаза збору інформації;
- фаза навчання;
- фаза прогнозу/рекомендації.

Рекомендаційна система збирає інформацію про користувачів для створення профілю користувача або моделі для прогнозування завдань, включаючи поведінку або вміст ресурсів, до яких користувач отримує доступ. Система повинна знати якнайбільше про користувача, щоб від початку давати розумні рекомендації. Рекомендаційні системи покладаються на різні типи введення даних: явний зворотний зв'язок, неявний зворотний та гібридний [18]. Профіль користувача – це набір особистої інформації, що включає когнітивні навички, інтелектуальні здібності, стилі навчання, інтереси, переваги та взаємодія із системою. Таким чином профіль користувача описує просту модель користувача. Успіх будь-якої рекомендаційної системи залежить від її здатності представляти поточні інтереси користувача.

При наданні явного зворотнього зв'язку система зазвичай пропонує користувачеві через системний інтерфейс надати рейтинги для елементів, щоб побудувати та покращити його модель. Точність рекомендації залежить від кількості оцінок, виставлених користувачем. Єдиний недолік цього методу полягає в тому, що він вимагає зусиль з боку користувачів, а користувачі не завжди готові надати достатньо інформації. Цей метод вважається найбільш якісним, оскільки не

передбачає отримання переваг з дій, а також забезпечує прозорість процесу рекомендацій, що призводить до дещо вищого сприйняття якості рекомендацій. і більше впевненості у рекомендаціях [19] .

При неявному зворотному зв'язку система автоматично визначає переваги користувача, відстежуючи різні дії користувачів, такі як історія покупок, історія навігації та час, проведений на деяких веб-сторінках, посиланнях, за якими слідував користувач, зміст електронної пошти та натискання кнопок серед іншого. Метод хоч і не вимагає зусиль від користувача, але менш точний. Також, наявні дослідження, що неявні дані про переваги можуть насправді бути більш об'єктивними, оскільки немає упередженості, що виникає через те, що користувачі реагують соціально бажаним чином [20], і немає проблем із самооцінкою або потреби в підтримці іміджу.

Сильні сторони як неявного, так і явного зворотного зв'язку можна поєднати в гібридній системі, щоб мінімізувати їх слабкі сторони та отримати найефективнішу систему. Цього можна досягти, використовуючи неявні дані як перевірку явної оцінки або дозволяючи користувачеві давати явний зворотний зв'язок тільки тоді, коли він вирішує виявити явний інтерес.

Фаза навчання. Рекомендаційна система застосовує алгоритм навчання для фільтрації оголошень на основі даних, зібраних на етапі збирання інформації [19].

Рекомендаційна система рекомендує або передбачає, які елементи краще. Це може бути зроблено або безпосередньо на основі набору даних, зібраних на етапі збору інформації, які може бути засновано на пам'яті або моделі, або на основі спостережуваних системою дій користувача. На рис. 2.1 показано етапи рекомендацій.

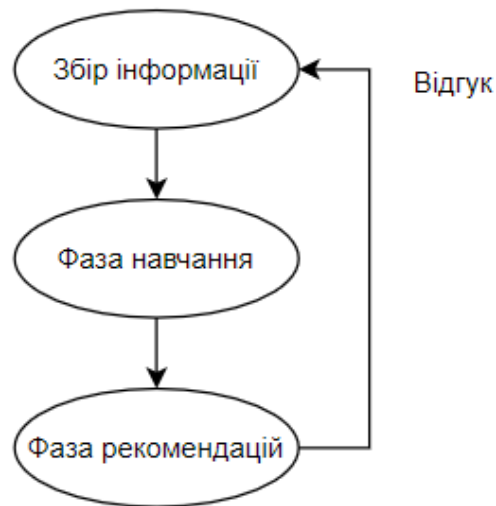


Рисунок 2.1 - Етапи рекомендаційного процесу

### 2.3 Методи фільтрації рекомендацій

Використання ефективних та точних методів рекомендацій дуже важливе для системи, яка надаватиме релевантні рекомендації своїм користувачам. Це пояснює важливість розуміння особливостей та можливостей різних технік рекомендацій. На рис. 2.2 показано структуру методів фільтрації рекомендацій.

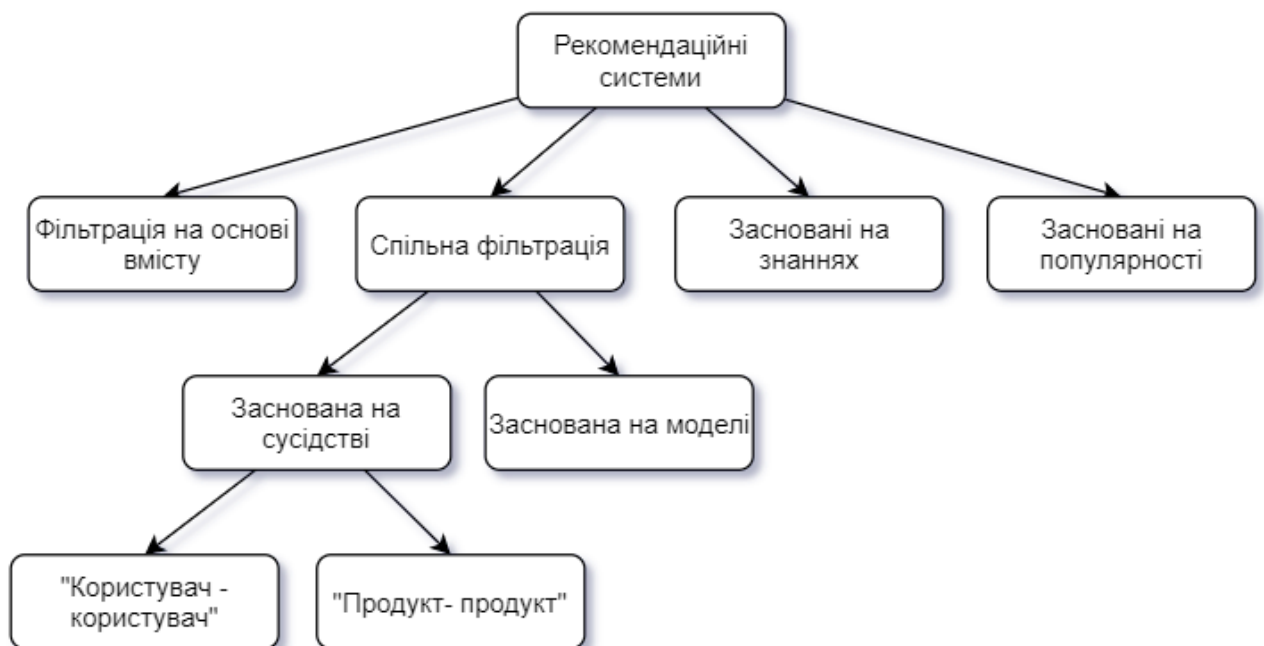


Рисунок 2.2 - Методи формування рекомендацій

Наразі існує декілька найбільш використовуваних методів формування рекомендаційних систем:

- фільтрація на основі вмісту;
- спільна фільтрація;
- засновані на знаннях;
- засновані на популярності
- гібридна фільтрація (використовує переваги декількох стратегій).

### 2.3.1 Фільтрація на основі вмісту

Фільтрація на основі вмісту - це алгоритм, який залежить від предметної області, де більше уваги приділяється аналізу атрибутів елементів до створення прогнозів. У методі фільтрації на основі вмісту рекомендації робляться на основі профілів користувачів з використанням функцій, вилучених із вмісту елементів, які користувач оцінював у минулому [21]. Користувачеві рекомендуються елементи, які в основному відносяться до елементів із позитивною оцінкою. Методика контентної фільтрації не вимагає профілів інших користувачів, оскільки вони не впливають на рекомендації. Крім того, якщо профіль користувача зміниться, метод може скоригувати свої рекомендації протягом короткого періоду часу.

Прикладом даного методу може бути користувач, що позитивно оцінив нерухомість, що знаходиться біля станції метро Наукова міста Харків, тоді йому будуть рекомендована нерухомість, що знаходиться поблизу. Кількість кімнат, рік будівництва, інфраструктура, наявність меблів, загальна площа також допомагають фільтрувати та ідентифікувати схожі оголошення. Візуалізацію цього підходу можна побачити на рисунку 2.3.

При такому підході зміст продукту оцінюється на основі уподобань користувача (профілю користувача), а значення атрибутів елемента - це неявні функції, які будуть використовуватися для створення профілю елемента. Потім оцінка предмета прогнозується з використанням обох профілів, і можуть бути зроблені відповідні рекомендації [22].

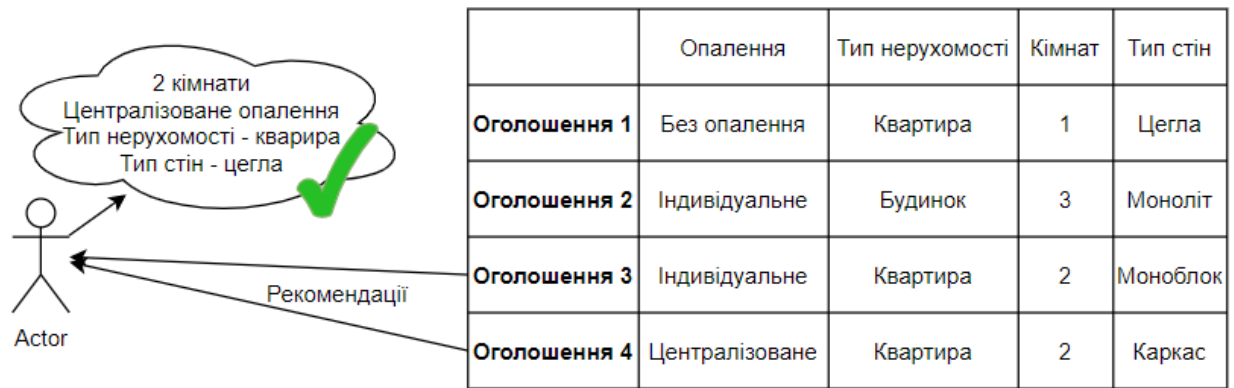


Рисунок 2.3 - Візуалізація підходу зі створенням профілю користувача і профілю елемента

Для формування рекомендації необхідно створити профіль користувача та профіль елементів. Для цього є дві групи вхідних даних:

- таблиця уподобань користувача, що демонструє відношення користувач-елемент;
- профіль користувача, що демонструє відношення елемент-атрибут

На основі цих даних створюється профіль користувача який демонструє відношення користувач-атрибут, тобто маємо набір значень атрибутів елементу деякого типу, яким користувач найбільше віддає перевагу [21]. Маючи профілі користувача та елементів, можна передбачити оцінку для інших елементів, відносно вподобань даного користувача.

Основним недоліком цього методу є необхідність мати глибокі знання та опис характеристик елементів у профілі.

### 2.3.2 Спільна фільтрація

Спільна фільтрація – це незалежний від предметної області метод прогнозування контенту, що працює шляхом створення бази даних (матриці користувач-елемент) переваг для елементів користувачам. Потім він зіставляє користувачів з відповідними інтересами та уподобаннями, обчислюючи схожість між їхніми профілями, щоб дати рекомендації [23]. Такі користувачі створюють

групу, яка називається сусідством. Користувач отримує рекомендації щодо тих оголошень, які раніше не оцінював, але які вже були позитивно оцінені користувачами в його сусідстві, тобто прогнози складаються індивідуально для кожного користувача, але використовувана інформація для цього зібрана від інших учасників (рис.2.4). Техніку спільної фільтрації можна розділити на дві категорії:

- memory based (заснований на сусідстві);
- model based (заснований на моделі).

	Оголошення 1	Оголошення 2	...	Оголошення j	Оголошення m
Користувач 1					
Користувач 2				Рекомендація	
...					
Користувач i					
Користувач p					

Рисунок 2.4 - Метод спільної фільтрації

### 2.3.2.1 Фільтрація заснована на сусідстві

Існує два основні різновиди цього методу: на основі користувачів, що займаються пошуком («користувач-користувач»), і на основі елементів, що утворюють ту чи іншу категорію («продукт-продукт»). Основні характеристики цих підходів полягають у тому, що вони використовують лише інформацію з матриці взаємодії «користувач-елемент» і не передбачають жодної моделі для вироблення нових рекомендацій.

Щоб дати користувачеві нову рекомендацію, метод "користувач-користувач" намагається приблизно ідентифікувати користувачів з найбільш схожим "профілем взаємодії" (найближчих сусідів), щоб запропонувати елементи, які є найбільш

популярними серед цих сусідів (і які є "новим" для нашого користувача) [24]. Цей метод називається "орієнтованим на користувача", оскільки він представляє користувачів на основі їх взаємодії з елементами та оцінює відстані між користувачами.

Кожен користувач може бути представлений своїм вектором взаємодій з різними елементами («його лінія» у матриці взаємодій). Також обчислюється «подібність» між еталонним користувачем, і всіма іншими користувачами. Ця міра подібності така, що двох користувачів з однаковими взаємодіями з одними й тими самими елементами слід вважати близькими. Після обчислення подібності для всіх користувачів ми можемо зберегти найближчих сусідів для нашого користувача, а потім запропонувати найпопулярніші серед них елементи (тільки дивлячись на елементи, з якими наш еталонний користувач ще не взаємодіяв).

При обчисленні подібності між користувачами необхідно уважно враховувати кількість «спільних взаємодій» (скільки елементів уже було розглянуто обома користувачами) [25]. Це необхідно, щоб уникнути того, щоб хтось, у кого є лише одна спільна взаємодія з нашим еталонним користувачем, міг мати 100% збіг і вважатися «найближчим», ніж той, хто має 100 спільних взаємодій та погоджується з ним тільки на 98% з них. Отже, два користувача вважаються схожими, якщо вони однаково взаємодіяли з багатьма спільними елементами (рис. 2.5).



Рисунок 2.5 - Колаборативна фільтрація «користувач-користувач»

Наприклад, якщо Користувачу 1 сподобалась Недухомість 1,2,3 та 4, а Користувачу 2 сподобалась Недухомість 1,2,3 та 5, то Користувачу 1 буде рекомендована Недухомість 5, а Користувачу 2 – Недухомість 4 відповідно.

Ідея методу елемент-елемент полягає в тому, щоб знайти елементи, схожі на ті, з якими еталонний користувач взаємодіяв «позитивно». Два елементи вважаються схожими, якщо більшість користувачів, що взаємодіяли з ними обома, зробили це однаково [24]. Цей метод називається «центрованим за елементами», оскільки він представляє елементи з урахуванням взаємодій із нею користувачів і оцінює відстані між цими елементами.

Припустимо, ми хочемо дати рекомендацію даному користувачеві. По-перше, ми розглядаємо елемент, який сподобався цьому користувачеві найбільше, і представляємо його (як і решту елементів) вектором взаємодії з кожним користувачем («його стовпець» в матриці взаємодії). Потім обчислюємо подібність між «кращим елементів» та іншими елементами. Після обчислення подібності ми можемо зберегти k-найближчих сусідів для вибраного «кращого елемента», які є новими для нашого користувача, і рекомендувати ці елементи.

Наприклад, якщо Користувачу 1 та Користувачу 2 сподобались Недухомість 1, 2, та 3 система робить висновок, що користувачам яким сподобались два з наведених елементів – може сподобатися і третій, тому Користувачеві 3, якому сподобалась Недухомість 1 та 2 буде запропоновано Недухомість 3 (рисунок 2.6).



Рисунок 2.6 - Колаборативна фільтрація «елемент-елемент»

### 2.3.2.2 Фільтрація заснована на моделі

У методах, заснованих на моделі, в контексті прогнозних моделей використовуються методи машинного навчання та аналізу даних. У випадках, коли модель параметризована, параметри цієї моделі вивчаються в контексті системи оптимізації. Деякі приклади таких методів, заснованих на моделях, включають дерева рішень, моделі на основі правил, байєсівські методи та моделі латентних факторів [26]. Багато з цих методів, наприклад моделі латентного фактора, мають високий рівень охоплення навіть для розріджених матриць рейтингів.

### 2.3.3 Рекомендаційні системи засновані на знаннях

Системи рекомендацій, заснованих на знаннях, особливо корисні в контексті товарів та послуг, які купуються не дуже часто. У таких випадках для процесу рекомендації може бути недоступним достатній рейтинг. Оскільки предмети купують не часто і з різними типами детальних опцій, важко отримати достатню кількість оцінок для конкретного елемента даного предмета. Ця проблема також зустрічається в контексті проблеми холодного запуску.

Процес рекомендацій виконується на основі подібності між вимогами користувача та описами елементів, або використання обмежень, що визначають вимоги користувачів. Процес полегшується при використанні баз знань, які містять дані про правила та функції подібності для використання під час процесу пошуку. Явна специфікація вимог призводить до більшого контролю користувачів над процесом рекомендацій. Системи, засновані на знаннях, унікальні тим, що дозволяють користувачам чітко вказувати, що вони хочуть.

Існує три основні ідеї які використовуються для створення рекомендаційних систем заснованих на знаннях:

– «діалог». Такі системи є діалоговими, тобто користувацькі уподобання і переваги виявляються в рамках циклу зворотного зв'язку [27]. Завдяки такому

діалогу уподобання не відомі точно на початку можуть бути виявлені в рамках сеансу рекомендацій (рис 2.7);

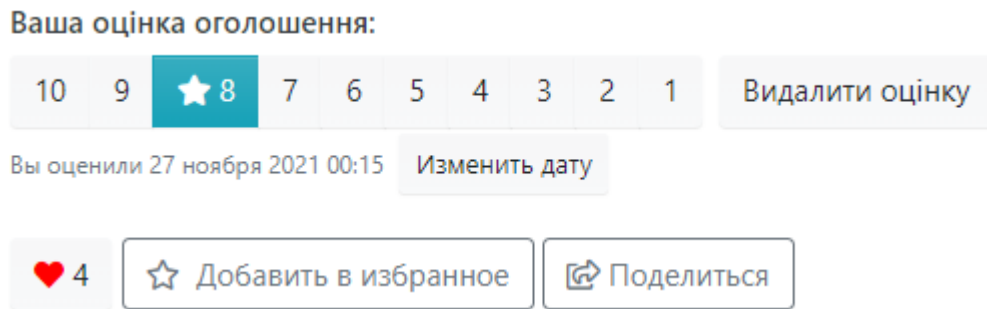


Рисунок 2.7 - Приклад зворотного зв'язку від користувача

– «обмеження». В таких системах зворотний зв'язок з користувачем реалізований у вигляді відповідей на запитання, які обмежують результуючу вибірку елементів. [28]. В такому разі рекомендаційна система є надбудовою над обмеженнями користувача, яка на основі встановлених правил робить пошук відповідних варіантів (рис 2.8);

– «критика». В таких системах зворотний зв'язок надається в термінах «критичних зауважень», тобто створюють запити на зміну, що стосуються рекомендованого елемента (подобається все окрім ціни). [27] Дані критичні зауваження використовується для навчання системи рекомендацій.

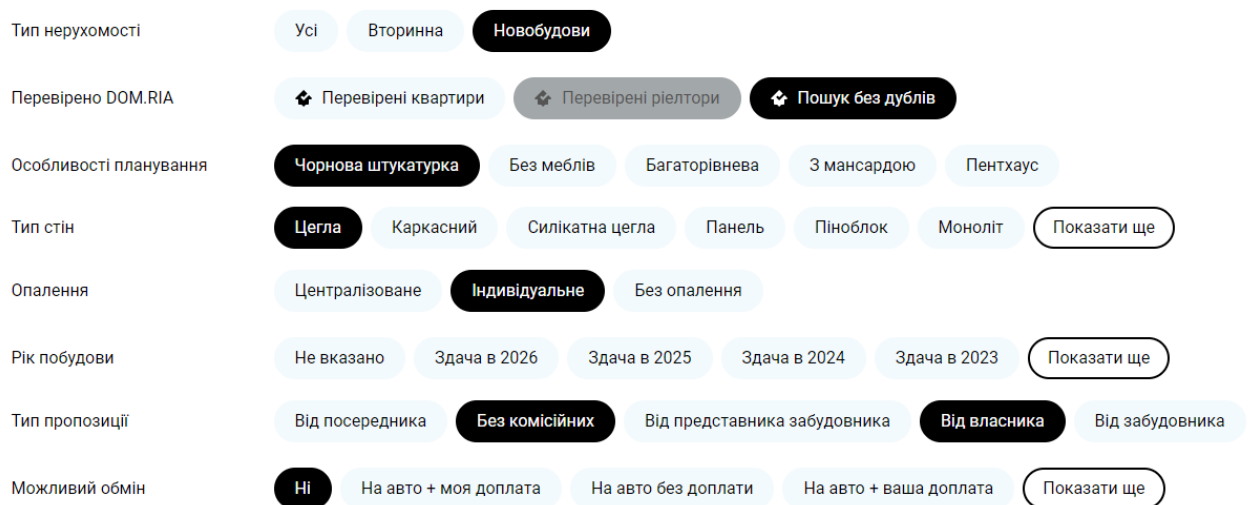


Рисунок 2.8 - Приклад обмежень користувача

### 2.3.4 Система рекомендацій на основі популярності

Це тип рекомендаційної системи, яка працює за принципом популярності або будь-чого, що є в тренді. Ці системи перевіряють оголошення, які найбільш популярними серед користувачів, і безпосередньо рекомендують їх.

Наприклад, якщо продукт часто купують більшість людей, система дізнається, що цей продукт є найпопулярнішим, тому для кожного нового користувача, система також рекомендуватиме цей продукт, і ймовірність того, що новий користувач також придбає це досить висока.

Головною перевагою такої системи є той факт, що вона не страждає від проблем з холодним стартом, оскільки немає потреби в історичних даних користувача [29]. Це ж є і головним недоліком – рекомендації не є персоналізованими.

### 2.4 Існуючі методи формування рекомендацій в досліджуваній галузі

В ході аналізу підходів спільної фільтрації був виявлений ряд її ключових недоліків:

- розрідженість даних – більшість рекомендаційних систем заснована на великій кількості даних (оголошень та їх характеристиках), а більшість користувачів не ставить оцінки оголошенням Як результат - матриця «предмет-користувач» - дуже велика та розріджена;

- масштабованість – кількість оголошень в системі досить велика, кожна має великий набір характеристик, а із збільшенням кількості користувачів в системі складність алгоритму колаборативної фільтрації стане занадто високою, а сам алгоритм досить складний для розрахунків;

- холодний старт – нові оголошення або користувачі представляють велику проблему для рекомендаційних систем. Такі оголошення, ще не мають оцінок користувачів, тому вони не будуть потрапляти до рекомендацій поки матриця «предмет-користувач» не буде наповнена певною кількістю значень. З цим же

пов'язана і проблема з наданням рекомендації для нового: він ще не має матриці переваг, а тому його не можливо віднести до жодної з груп [22, 24].

Стратегія на основі вмісту має ряд переваг відносно спільної фільтрації:

- незалежність користувача - для спільної фільтрації потрібна оцінка інших користувачів, щоб знайти схожість між ними, щоб надати рекомендацію. Замість цього метод, заснований на контенті, повинен тільки проаналізувати елементи і профіль користувача для рекомендації;

- відсутність холодного старту для нових оголошень - на відміну від спільної фільтрації, можна пропонувати нові елементи, перш ніж вони будуть оцінені значним числом користувачів, порівнюючи їх з іншими оголошеннями.

Цей підхід також має суттєві недоліки:

- обмежений аналіз оголошень - якщо оголошення не містять достатньо інформації для точного розрізнення елементів, рекомендація буде не зовсім остаточною;

- надспеціалізація - фільтрація на основі вмісту може не запропонувати нічого нового (що так необхідно сучасному користувачеві);

- холодний старт для нових користувачів: неможливо надати каректну рекомендацію, коли недостатньо інформації для створення релевантного профілю користувача.

Найбільшим недоліком стратегія заснованої на знаннях є вузьке місце набуття знань, викликане необхідністю явного визначення рекомендацій [19]. Для створення такої системи рекомендацій необхідна більша залученість та затрат по часу від користувача (а вони не дуже люблять це робити) та участь експертів предметної галузі, які побудують систему асоціативних правил, на яких будуть базуватися рекомендації. Такий спосіб знань є досить затратним з точки зору людських ресурсів, фінансів та часу. До того ж, при такому підході, кожна така рекомендаційна система буде унікальною і її не можливо буде застосувати до іншої предметної області без повторної побудови нових асоціативних правил.

Основний та суттєвий, в даній предметній області, недолік системи заснованої на популярності – рекомендації не є персоналізованими.

### 3 ПОСТАНОВКА ЗАДАЧІ

#### 3.1 Визначення класу досліджуваної системи

Завданням даної роботи є дослідження, аналізі існуючих методів створення рекомендаційних систем з ціллю виявлення найкращих, а також розробка методів формування рекомендацій в системах інтернет комерції на прикладі системи для оренди та продажу нерухомості.

Особливістю даної системи є те, що вона призначена для цільової категорії користувачів, що виконують наступні операції: орендування нерухомості, здавання нерухомості в оренду її покупку та продаж.

Специфічність обраної предметної сфери не дозволяє створювати навіть відносно постійні профілі користувачів. Існує дуже мала імовірність того, що користувач буде декілька разів шукати дуже схожу за параметрами нерухомість, а спрогнозувати його наступну активність неможливо, оскільки причин для пошуку оренди нерухомості може бути безліч – відрядження, відпустка, зміна постійного місця проживання. А вірогідність повторної купівлі нерухомості і зовсім дуже мала, особливо в однакових типах нерухомості. Системи, засновані на рейтингах, не працюють добре в цій області через малу кількість доступних рейтингів. Крім того, дуже часто виникає ситуація коли клієнти хочуть явно вказати свої уподобання (наприклад, вказати максимальну вартість чи наявність певної інфраструктури поблизу).

Також, враховуючи предметну область, надання рекомендацій ускладнюється тим, що нерухомість необхідно оцінювати не лише з кількісних характеристик, а й з урахуванням якісних. Наприклад, аналіз обслуговування нерухомості зазвичай виконується з урахуванням продуктивності експлуатації, естетичної цінності або суспільного іміджу, комфорту (шум, колір, якість повітря, тепловий комфорт, умови роботи), гнучкості та вартості (дизайн, будівництво, непрямі витрати, експлуатаційні та експлуатаційні витрати, витрати на реконструкцію, сплачені відсотки за кредит).

### 3.2 Характеристика предметної області

В даний час існує безліч систем, які допомагають здійснювати послуги між продавцем і покупцем або орендарем та орендодавцем на ринку нерухомості. Зазвичай, вони надають такі послуги:

- перегляд та пошук нерухомості
- купівля нерухомості;
- продаж нерухомості;
- здавання нерухомості в оренду;
- короткострокова оренда нерухомості;
- довгострокова оренда нерухомості.

Такі сервіси зазвичай складаються з клієнтського додатку, що представлений веб-інтерфейсом, та відповідає за отримання даних з сервера для подальшої візуалізації у зручному для користувача вигляді, а також відправки даних на сервер, який приймає запити клієнта, обробляє і зберігає інформацію.

Досліджуваний сервіс буде спеціалізуватися на продажі та оренді нерухомості та буде виконувати усі описані стандартні можливості подібних систем, а також надавати рекомендації.

### 3.3 Формалізована постановка задачі

Нехай  $C$  - сукупність усіх користувачів,  $S$  є набором усіх можливих елементів, які можуть бути рекомендовані (нерухомість), а  $u$  буде деякою функцією, що визначає корисність елемента  $s$  для користувача  $c$ .

$$u: C \times S \rightarrow R,$$

де  $R$  - повністю упорядкована множина (наприклад, позитивні цілі числа в деякому діапазоні). Тоді для кожного користувача  $c \in C$  ми намагаємось обрати

такий елемент  $s' \in S$ , що максимізує функцію корисності для цього користувача. Тобто:

$$\forall c \in C, s' = \operatorname{argmax}_{s \in S} u(c, s)$$

Функція корисності  $u$  позначає наскільки сильно певний елемент може зацікавити користувача та може мати довільну форму, виходячи з алгоритму

### 3.4 Вибір стратегії для надання рекомендацій

Аналіз основних стратегій на основі вмісту та спільної фільтрації для створення рекомендаційних систем показав, що вони не можуть бути застосовані напряму для вирішення поставленої задачі оскільки:

- активність користувачів на таких сервісах буде тимчасовою, а матриці «елемент-користувач» будуть дуже розрідженими, адже користувач буде оцінювати лише певну обмежену групу товарів (наприклад трьох кімнатну квартиру у певному районі з обмеженою максимальною вартістю оренди) [25]. Тому проблема розрідженості даних стає критичною;

- специфічність обраної предметної сфери не дозволяє створювати навіть відносно постійні профілі користувачів. Існує дуже мала імовірність того, що користувач буде декілька разів шукати дуже схожий за параметрами елемент, а спрогнозувати його наступну активність неможливо. Отже основне припущення, що допускає спільна фільтрація у рамках такої системи не дійсне для кожного конкретного користувача [20]. Тобто групи користувачів, що формуються, будуть непостійними, і оцінки такої групи буде складно використовувати для прогнозування переваг нового користувача;

- проблема «холодного старту» не може існувати у рамках задачі, що розглядається, адже тут спостерігається постійна «текучість» активних елементів.

А рекомендаційні системи засновані на знаннях, потребують великих затрат та вони не є універсальними через необхідність явного збору знань для кожної

предметної області, та побудові асоціативних правил взаємодії з залученням експертів предметної області.

Тобто існує необхідність у розробці стратегій та способів прогнозування уподобань користувачів, які могли б працювати із предметною галуззю з оренди та продажу нерухомості.

Деякі аспекти розглянутих методів можна використати та об'єднати між собою, для створення системи рекомендацій що здатна працювати з досліджуваною предметною областю та мінімізує вузьке місце у способі отримання знань.

### 3.5 Вихідна інформація

Результатом роботи рекомендаційної системи та вихідними даними є результати порівняння методів формування рекомендацій в системах інтернет комерції на прикладі системи для оренди та продажу нерухомості, а також список оголошень, запропонованих розробленою системою рекомендацій, які певним чином відображаються у зовнішньому користувацькому інтерфейсі у веб додатку.

### 3.6 Вхідна інформація

Вхідною інформацією є:

- оголошення;
- вектор значень характеристик, що описують об'єкт;
- шкали в яких вимірюються характеристики;
- особисті дані клієнта;
- обмеження користувача;
- зворотній зв'язок користувача;
- експертні дані, з урахуванням яких буде формуватися рекомендаційна система.

Опис вхідних даних представлено в табл. 3.1.

Таблиця 3.1 Перелік і опис вхідних документів

Назва документу	Періодичність введення даних	Джерело даних
Оголошення	Постійно	Введення даних оператором або користувачем (власником нерухомості) з клавіатури
Вектор значень характеристик, що описують об'єкт	Постійно	Введення даних оператором або користувачем (власником нерухомості) з клавіатури
Шкали в яких вимірюються характеристики	Одноразово	Введення даних оператором або користувачем (власником нерухомості) з клавіатури
Особисті дані клієнта	Зазвичай одноразово, з можливістю повторного введення	Введення даних користувачем з клавіатури
	Постійно	Неявний збір системою
Обмеження користувача	Періодично	Введення даних користувачем
Зворотній зв'язок користувача	Періодично	Введення даних користувачем
Експертні дані	Одноразово	Введення даних програмістом в програмний код

## 4 РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ВИРІШЕННЯ ЗАДАЧІ

### 4.1 Розробка стратегії надання рекомендацій

Деякі аспекти розглянутих в попередніх розділах методів можна використати та об'єднати між собою, для створення системи рекомендацій що здатна працювати в досліджуваній предметній області та мінімізує вузьке місце у способі отримання знань.

По-перше, при потраплянні на платформу нового користувача, про якого в нашій системі немає жодних даних, а якісь рекомендації необхідно надати – доцільно застосувати метод рекомендацій на основі популярності. При типі пошуку короткострокової оренди даний метод може бути досить ефективним, а в той же час для продажу та довгострокової оренди – ефективність буде мінімальною. Не зважаючи на це, застосування даного методу має сенс для початкового набору рекомендацій. В подальшому, система має зводити до мінімуму вплив цього методу на рекомендації (для продажу та довгострокової оренди) при достатній кількості зібраних даних про користувача.

По-друге, слід застосувати ідею «обмежень», що зменшить простір елементів, що рекомендуються, відповідно до основних обмежень у пошуковому запиті. Тобто, не має сенсу пропонувати користувачеві квартири у Відні, якщо він задав обмеженням місто Харків, навіть якщо всі інші характеристики співпадають ідеально. При цьому необхідно виділити строгі характеристики, зміна яких є критичною (наприклад місто), важливі характеристики (ціна, дата заїзду може дуже незначно варіюватися, наявність ремонту/меблів та ін.) та слабкі характеристики (наприклад рік побудови нерухомості, матеріал стін) які, зазвичай, мають менше значення для користувача. Тобто, необхідно отримати вектор вагових коефіцієнтів, що описують важливість характеристик. При цьому початковий вектор базується на основі експертного бачення та може бути скоректованим кожним користувачем окремо (наприклад, якщо для більшості користувачів колір даху не має критичного значення, а Користувач 1 хоче тільки дах рожевого кольору – він може вказати, що

ця характеристика є важливою для нього, або для користувача, що приїде на своїй машині наявність пакувального місця є дуже важливим, а для користувача, що машини не має – це зайва опція за яку він не бажає платити). Такий вектор вагових коефіцієнтів є важливим, оскільки виконання фільтрації за всіма обмеженнями користувача може призвести до появи пустого списку оголошень (хочу найкраще за дешеву ціну), що є неприпустимим.

По-третє, можна частково застосувати метод фільтрації заснований на контенті, але в цьому разі основна увага буде приділятися опису товарів зменшуючи роль профілю користувача [19]. Стратегія полягає у наступному: користувачу пропонуються об'єкти, які максимально схожі на вподобаний ним об'єкт, при цьому «вподобаним» вважається об'єкт, що переглядається користувачем у даний момент. Тобто в основу такої рекомендації лягає так званий «метод ідеальної точки», ідеальною точкою при цьому і є об'єкт, що вподобав користувач. Такий підхід дозволить знаходити елементи, які не зовсім відповідають пошуковому запиту (фільтру) користувача, що є першочерговою задачею. Для порівняння значень атрибутів елементів та знаходження «подібності» елементів необхідно використовувати метрики подібності. При цьому для кожного атрибута необхідно знати дві речі: його відносну важливість (засновану на експертній оцінці), та вигляд функції яка описує подібність двох значень цього атрибуту.

По-четверте, маючи певні дані з методу заснованому на контенті можемо застосувати метод спільної фільтрації на основі сусідства «користувач-користувач». Метод заснований на контенті добре рекомендує схожі оголошення, в той час як метод спільної фільтрації дозволить урізноманітнити список рекомендацій, що є важливим фактор в сучасному світі, а також врахувати такі характеристики як престижність, що явно не вказується в оголошенні.

По-п'яте, можна використати ідею «діалогу» із методу експертних систем, для організації зворотного зв'язку від користувача [21].

По-шосте, оскільки користувачі зазвичай купують лише одну однотипну нерухомість або орендують одну квартиру на тривалий період часу, а якщо на

коротку – в різних місцях чи містах, з плином часу збільшується вірогідність того, що вподобання користувача зміняться (врахування попереднього досвіду, збільшення сім'ї – збільшення необхідної площі нерухомості, вплив зовнішніх факторів). Для врахування цього рекомендації формуються методом «ковзаючого вікна». У процесі роботи алгоритму «вікно» послідовно зміщується на одиницю спостереження, при цьому «забуваючи» попередні спостереження. Такий підхід реалізується на основі «діалогу», відслідковуючи уподобання користувача та враховуючі їх зміну шляхом «забування» старих елементів.

Процес створення рекомендацій наведено на рисунку 4.1.



Рисунок 4.1 - Процес створення рекомендацій

Аналіз реакції користувача проводиться за наступними критеріями, від найвпливовіших до найменш впливових: нерухомість додана до обраного  $\Rightarrow$  час перегляду оголошення  $\Rightarrow$  явна оцінка оголошення користувачем  $\Rightarrow$  відкриття сторінки детального перегляду оголошення.

Описана стратегія має наступні ключові переваги, відносно вирішуваної проблеми:

- зменшення важливості профілю користувача;
- вирішення проблема холодного старту, адже елементи не залежать від розміру відношення «користувач-елемент»;
- уподобання користувача виявляються у реальному часі;
- уподобання користувача забуваються з плином часу;
- новому користувачеві відразу надаються певні рекомендації.

#### 4.2 Математичний опис задачі

Запропонована стратегія передбачає, для початку необхідно визначитися з найпопулярнішими оголошеннями:

$$q_i = \text{countView}(p_i)$$

де  $q_i$  – популярність оголошення,  $p$  – конкретне оголошення,  $i$  – індекс оголошення,  $\text{countView}$  – функція, що визначає кількість переглядів конкретного оголошення, тоді:

$$\text{Popular} = \text{sort}(Q)$$

де

$$Q = \{q_1, q_2, q_i, \dots, q_n\}$$

Кожне оголошення характеризується вектором значень деяких характеристик. Характеристики можуть мати як якісні так і кількісні характеристики. Особливістю характеристик, вимірних в якісних шкалах, є те, що вони в принципі не містять інформацію про домінування значень показників  $i$ ,

отже, в початковому вигляді не можуть бути використані для ранжирування альтернатив. Якісні шкали можуть бути використані в цих цілях тільки в тому випадку, якщо є додаткова інформація про домінування значень ознаки, а не просто про угруповання об'єктів. Для цих цілей використано експертні оцінки. На відміну від кількісних оцінок, які відповідають, як правило, об'єктивним вимірам показників, експертні оцінки, характеризують суб'єктивні думки експертів і часто проводяться в бальних шкалах.

Для базової оцінки елементів характеристики мають бути однотипними для, не залежати від їхнього змісту і відображати уявлення про перевагу різних значень оцінки (щоб не виникало проблем з порівнянням, наприклад, ціни, що вимірюється в тисячах, та кількістю кімнат, що зазвичай варіюється от 1 до 4). Для цього вона повинна відповідати наступним вимогам:

- мати єдиний інтервал змін  $[0; 1]$ ;
- бути безрозмірною і інваріантною до виду екстремуму часткового критерію (мінімум або максимум):

$$k_j[s_j(x)] = \left( \frac{s_j(x) - s_{min_j}}{s_{max_j} - s_{min_j}} \right)$$

де  $s_j(x)$  - значення  $j$ -ї характеристики оголошення;

$s_{max_j}$  та  $s_{min_j}$  - відповідно найкраще та найгірше значення характеристики.

Тепер кожне оголошення можливо представити вектором критеріїв:

$$K = \{k_1, k_2, k_j, \dots, k_m\}$$

де  $k_j$  – кількісна оцінка схожості  $j$ -тої характеристики елемента з  $j$ -тою характеристикою «еталонного» елемента.

Відносна важливість кожної з характеристик виражена вектором вагових коефіцієнтів:

$$W = \{w_1, w_2, w_i, \dots, w_m\}$$

де  $w_i$  – відносна важливість  $i$ -тої характеристики елемента;

$$\sum w_j = 1$$

Для оцінки схожості деякого елемента з «еталонним» використовується метод зваженої суми критеріїв [20]. Оцінка схожості елемента розраховується за формулою:

$$Score_i = \sum w_j \times k_j \quad (4.1)$$

Оцінка близькості елементів відбувається окремо за кожною характеристикою. Для оцінки на основі групи елементів, оцінки (4.1) нового елемента, що розраховуються відносно кожного елемента у межах групи переглянутих, підсумовуються між собою для отримання кінцевого результату:

$$Score_i^* = \sum_{i=0}^N Score_i$$

де  $N$  – кількість переглянутих елементів.

В цьому випадку подібний розрахунок можна розглядати як голосування переглянутих сторінок за непереглянуті, та сторінка яка має найбільші значення голосів (4.1) відносно елементів у групі і отримує найвищий бал.

Бальна оцінки, на яку система виставила нерухомості для певного користувача відбувається за рахунок денормалізації оцінки за формулою:

$$y_i = \text{round}((rate_{max} - rate_{min}) \times Score^* + rate_{min}) \quad (4.2)$$

Де  $rate_{max}$  та  $rate_{min}$  - відповідно найкраща та найгірша оцінка,  $\text{round}$  – функція, що округляє значення до цілого числа за правила математики.

Для розрахунку методу заснованому спільній фільтрації є матриця рейтингів розміру  $n \times m$ , яку ми назвемо  $R$ , матриця користувачів позначена як  $U$ , а матриця

елементів як Р. Користувач і представлений вектором  $c_i$ , і це  $t$  користувачів, тому  $i = 1, \dots, t$ . Елемент представлений вектором  $c_i$ , і це  $n$  елементів, тому  $i = 1, n$ . Тепер ми хочемо спрогнозувати рейтинг  $r_{ti}$ , якщо цільовий користувач  $i$  не дивився/не оцінював елемент  $i$ . Процес полягає в тому, щоб обчислити подібність між цільовим користувачем  $t$  та всіма іншими користувачами, вибрати  $X$  найбільш схожих користувачів і взяти середньозважене значення оцінок цих  $X$  користувачів з подібністю як ваги. Рейтинг  $r_{ti}$  визначається як:

$$r_{ti} = \frac{\sum_k \text{Similarities}(c_k, c_t) * r_{ki}}{\text{number\_ratings}}$$

Зважаючи на те, що у людей різні шкали оцінок, і деякі люди зазвичай дають більш високі оцінки, ніж інші. Цього зсуву можна уникнути, віднімаючи середній бал кожного користувача по всіх пунктах при обчисленні середньозваженого значення і додавши його для цільового користувача:

$$r_{ti} = \bar{r} + \frac{\sum_k \text{Similarities}(c_k, c_t) * (r_{ki} - \bar{r}_k)}{\text{number\_ratings}}$$

Для розрахунку методу заснованому на контенті застосуємо Байєсовський класифікатор, орієнтований на елемент. Для кожного елемента ми хочемо навчити класифікатор Байєса, який приймає користувачькі характеристики в якості вхідних даних і виводить або «подобається», або «не подобається» [16]. Отже, щоб розв'язати завдання класифікації, необхідно розрахувати співвідношення між ймовірністю того, що користувачу сподобається аналізований елемент, і ймовірністю, що не сподобається:

$$\frac{p_{item}(\text{like} | \text{user\_features})}{p_{item}(\text{dislike} | \text{user\_features})}$$

Це відношення умовних ймовірностей, яке визначає наше правило класифікації (з простим порогом), може бути виражене за формулою Байєса:

$$p_{item}(like|user\_feature) = \frac{p_{item}(user\_features|like) \times p_{item}(like)}{p_{item}(user\_features)}$$

$$p_{item}(dislike|user\_feature) = \frac{p_{item}(user\_features|dislike) \times p_{item}(dislike)}{p_{item}(user\_features)}$$

$$\frac{p_{item}(like|user\_features)}{p_{item}(dislike|user\_features)} = \frac{p_{item}(user\_features|like) \times p_{item}(like)}{p_{item}(user\_features|dislike) \times p_{item}(dislike)}$$

Де  $p_{item}(like)$  та  $p_{item}(dislike)=1- p_{item}(like)$  є апіорними значеннями, обчисленими на основі даних, тоді як передбачається, що ймовірності  $p_{item}(user\_feature|like)$  та  $p_{item}(user\_feature|dislike)$  передбачається, наслідують гауссівський розподіл з параметрами, які також визначені на основі даних.

Для характеристик, що представлені у інтервальній оцінці схожості характеристики елемента аналогічній характеристиці еталонного елемента буде розраховуватися за допомогою функцій затухання. Вона оцінює певну величину за допомогою функції, що затухає в залежності від збільшення відстані між значенням цієї величини та певним заданим джерелом (центром мас). Така оцінка схожа на запит, щодо входження деякого значення в заданий діапазон, з подальшим призначенням кількісної оцінки цьому значенню в залежності від близькості до центра діапазону. Для оцінки характеристик необхідно для кожної з них визначити значення наступних параметрів (рис. 4.2):

- *origin* – значення еталонного оголошення;
- *offset* – відстань, на якій значення характеристики буде вважатися рівним еталонному значенню;
- *scale* – відстань від центру (+ зміщення) яка визначає діапазон прийнятної (допустимої) схожості;
- *decay* – значення схожості в яке оцінюється значення характеристики на відстані масштабу від обраного центру [21].

Параметри *Scale* та *Decay* разом встановлюють швидкість спаду залежності як видно з рисунку 4.2.

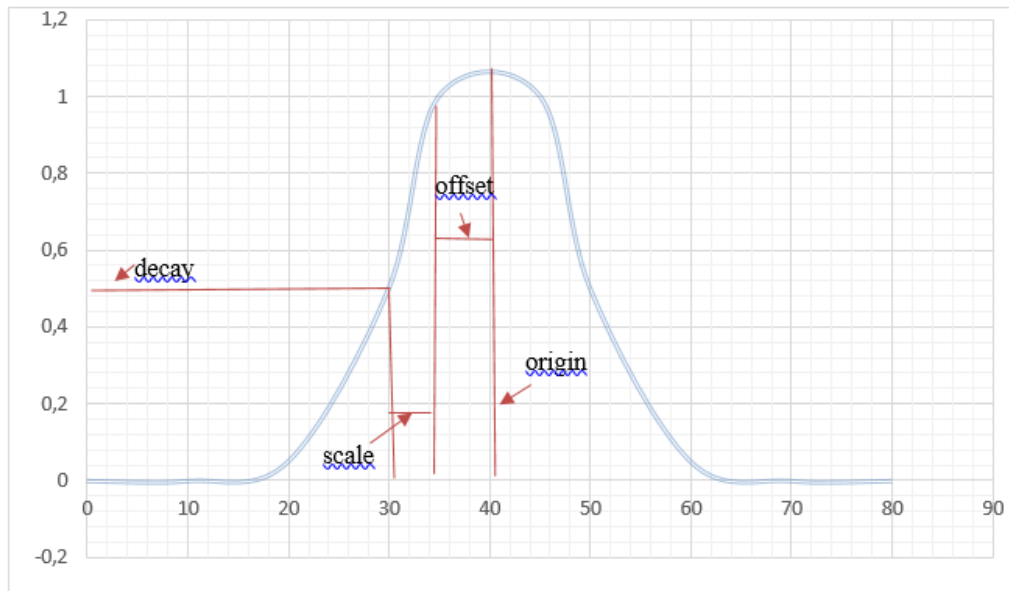


Рисунок 4.2 - Інтерпретація параметрів затухання

Тоді затухання для певної характеристики деякого елемента можна визначити наступним чином:

$$k_j = \left( -\frac{(|value_j - origin| - offset)^2}{2 \times \left(-\frac{scale^2}{2 \times in(decey)}\right)} \right)$$

де  $K_j$  – кількісна оцінка схожості  $j$ -тої характеристики елемента з  $j$ -тою характеристикою «еталонного» елемента, приймає значення в діапазоні  $[0;1]$ ;  $value_j$  – значення  $j$ -тої характеристики елемента;  $origin$  – значення  $j$ -тої характеристики еталонного елемента;  $offset$  – відстань, на якій значення характеристики буде вважатися рівним еталонній;  $scale$  – відстань від центру (+ зміщення) яка визначає діапазон прийнятної (допустимої) схожості;  $decaу$  – значення схожості в яке оцінюється значення характеристики на відстані  $scale$  від  $origin$ .

Для характеристик, що представлені за допомогою номінальних шкал можливо встановити лише подібність або відмінність елементів. Тому подібність значень між ними підраховується наступним чином:

- $K_j=1$ , якщо значення відповідних характеристик однакові;
- $K_j= 0$ , якщо значення відповідних характеристик різні.

Якщо значення якоїсь характеристики відсутнє, то її оцінка схожості з аналогічною характеристикою еталонного елемента розраховується як:  $K_j = 0$ . Тобто значення функції схожості в такому разі буде зменшуватися.

#### 4.3 Оцінка точності роботи рекомендаційної системи

Найбільшим показником якості роботи рекомендаційної системи є точність прогнозування вподобань користувача. Для перевірки точності прогнозування вподобань порівнюють вектор прогнозованих оцінок користувача, та вектор фактичних оцінок користувача. Для порівняння даних векторів може бути використані різноманітні метрики [30]. Найпопулярнішими з них є: середня абсолютна помилка (mean absolute error - MAE) та середньоквадратична помилка (root mean square error - RMSE).

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{countRate}$$

де, *countRate* – кількість оцінок в тестовій вибірці,

$y_i$  - оцінка визначена системою (4.2),

$x_i$  - оцінка визначена користувачем (не відома на момент прогнозування).

Основна відмінність RMSE полягає в тому, що він більше карає нас, коли наші рейтингові прогнози помиляються, і менше карає нас, коли ми досить близькі:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{countRate}}$$

Для більшої наочності дані нормалізуються, тому отриманий результат варіюється в межах 0-1.

Як для MAE, так і для RMSE чим нижчі показники – тим більш точна наша система, тобто нам потрібні найнижчий показник, а не найвищий.

## 5 ПРОЕКТУВАННЯ СИСТЕМИ

### 5.1 Опис системних вимог

Інформаційна система призначена для автоматизації роботи агентства нерухомості. Виходячи з цілей створення системи, вона має відповідати наступним системним вимогам:

- наявність наступних компонентів програми: клієнтський додаток, що підключений до сервера додатків, який в свою чергу підключений до сервера бази даних []. Інформаційна система передбачає публікацію веб-додатку в мережі Інтернет та забезпечення доступу користувачів до інформації;

- клієнтська частина системи реалізована у вигляді web-сайту
- забезпечення багатокористувацького режиму роботи,

Система має забезпечувати надійність, а саме:

- забезпечувати захист даних від несанкціонованого доступу;
- надавати повідомлення про помилку при некоректних діях користувача зі вказуванням місця, де сталася ця помилка, та шляхів її виправлення;
- наявність контролю на коректність всіх введених даних;
- контролювати небезпечні операції (видалення оголошення, підтвердження покупки і т.д.) [31].

### 5.2 Опис функціональних вимог

Для визначення функціональних вимог до інформаційної системи було розроблено її функціональну модель із застосуванням стандартів IDEF0 [32]. На рис. 5.1 представлена контекстна діаграма системи. Ціль: виявлення переліку функцій, що потребують автоматизації в агентстві нерухомості. Точка зору: клієнт системи. На рис. 5.2 представлена декомпозиція концептуальної діаграми. Вона дозволяє уточнити основну функцію системи «Облік ріелторської діяльності».

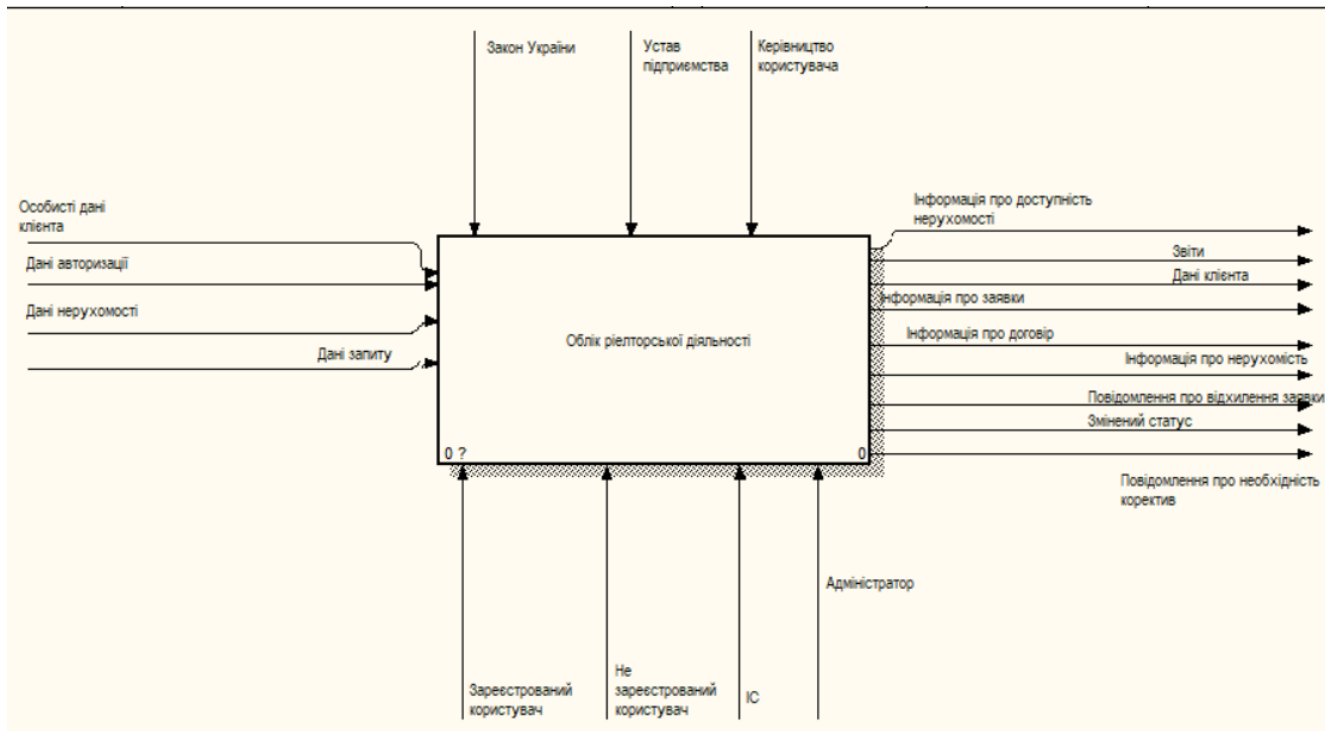


Рисунок 5.1 – Контекстна діаграма інформаційної системи

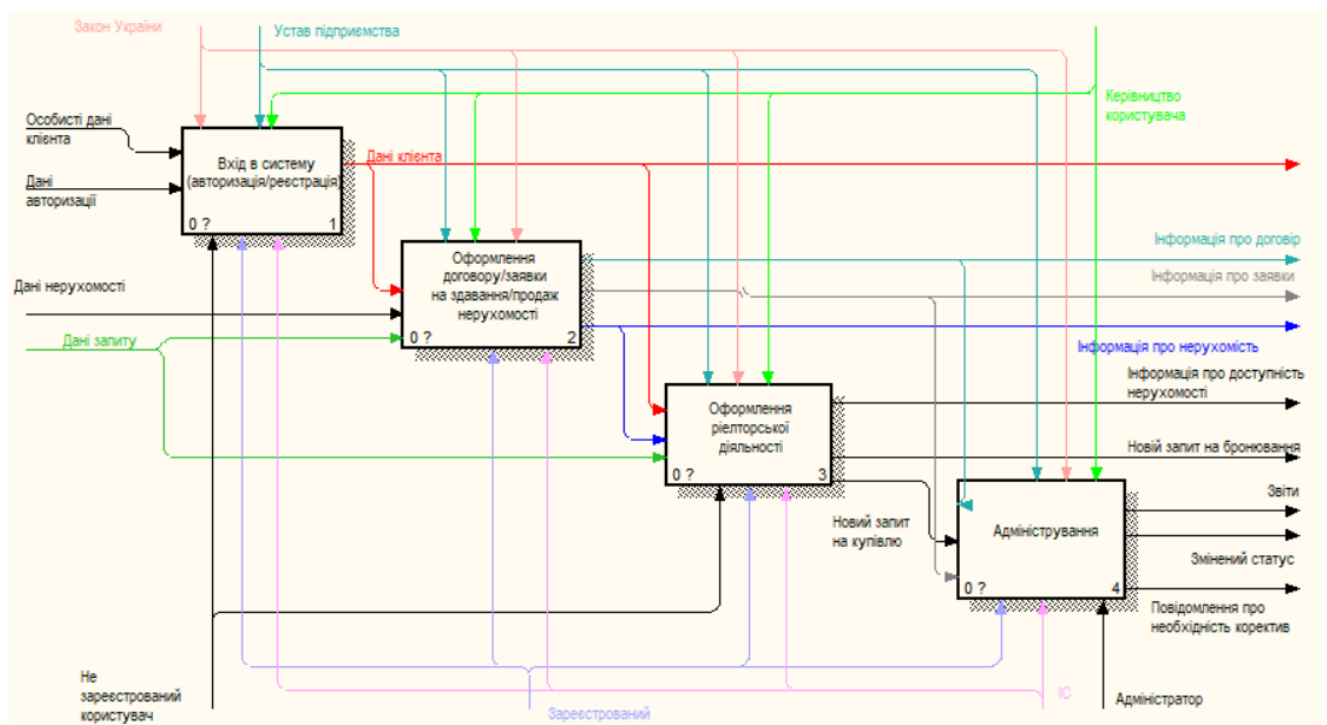


Рисунок 5.2 – Діаграма декомпозиції, що демонструє функції системи

В рамках даної роботи основна увага буде приділятися функції «Оформлення ріелторської діяльності» (рис. 5.3). Ця функція необхідна для підтримки дій бронювання та покупки нерухомості. Тобто, для початку роботи користувач має

отримати змогу переглянути існуючий каталог нерухомості та рекомендовані йому позиції, фільтрувати каталоги за деякими характеристиками ( місце призначення, ціна, дата публікації оголошення, кількість кімнат), переглядати детальну інформацію обраної нерухомості, коментарі на цю нерухомість від інших користувачів (для нерухомості із типом короткострокова оренда). Ця інформація може бути доступною для будь-якого (zareєстрованого або незареєстрованного) користувача. Zareєстровані користувачі також можуть при перегляді детальної інформації, додати нерухомість до каталогу «Обране» - спеціальний каталог нерухомості який знаходиться у особистому кабінеті користувача (для швидкого повторного доступу) [33]. Переглядати такий каталог може тільки той користувач, що його створює. Для zareєстрованого клієнта, якщо він обирає нерухомість для орендування, є можливість забронювати нерухомість, для довгострокової оренди, чи короткострокової оренди на певні дати (власник нерухомості може підтвердити бронювання чи відхилити його вказав причину). Якщо користувач обирає нерухомість для покупки він може зробити запит на купівлю нерухомості [32].

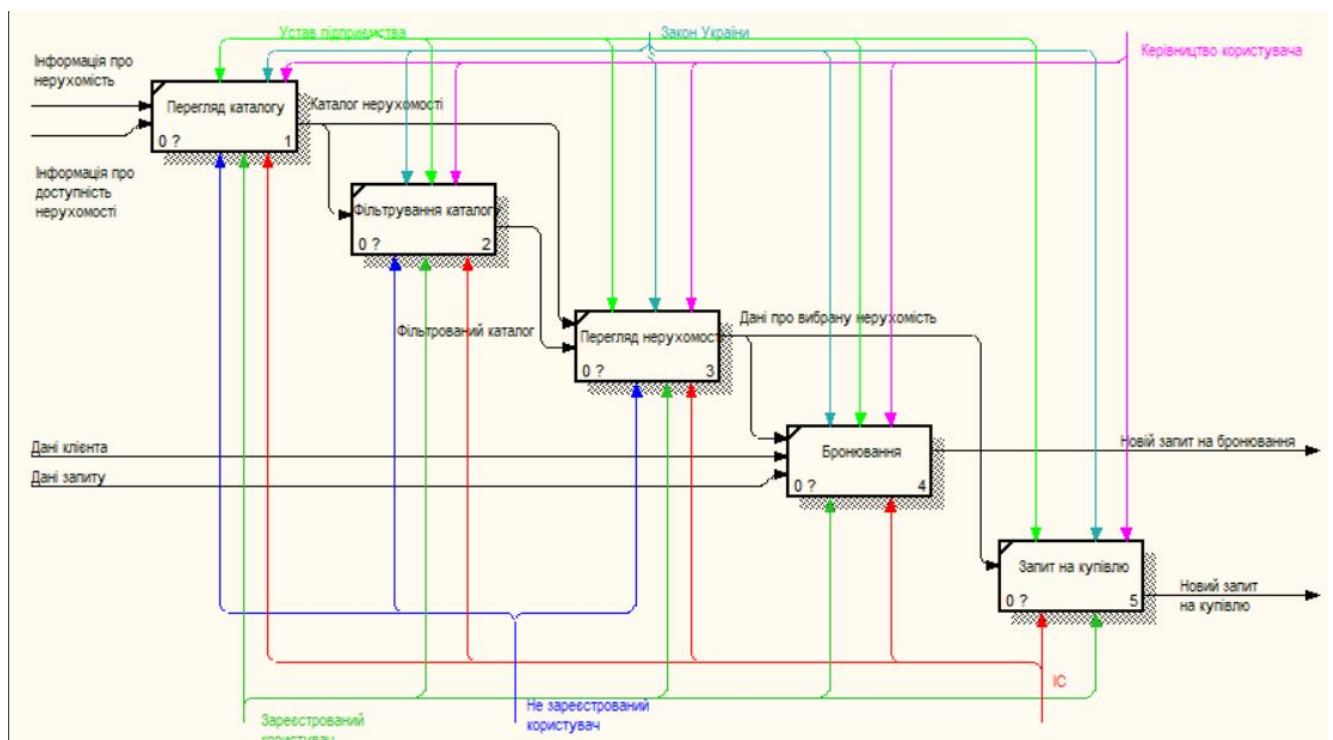


Рисунок 5.3 – Діаграма декомпозиції функції «Оформлення ріелторської діяльності» системи

Функціональне моделювання дозволило уточнити функціональні вимоги до розроблюваної системи:

- а) система має надавати наступні типи користувачів:
  - зареєстрований користувач;
  - незареєстрований користувач;
  - адміністратор.
- б) система має забезпечити наступні функції для незареєстрованого користувача:
  - перегляд каталогу нерухомості короткострокової оренди, довгострокової оренди, продажу, а також рекомендації за всіма типами операцій;
  - фільтрування каталогів;
  - реєстрація;
  - авторизація.
- в) система має забезпечити наступні функції для зареєстрованого користувача:
  - функції доступні для незареєстрованого користувача;
  - вихід із системи;
  - перегляд особистого кабінету;
  - створення договору на нерухомість та додавання оголошення про оренду / продажу;
  - зміна персональних даних (логін, пароль тощо);
  - перегляд і редагування нерухомості користувача;
  - видалення нерухомості користувача;
  - перегляд орендованої користувачем нерухомості;
  - перегляд архіву орендованої користувачем нерухомості;
  - додавання нерухомості в «Обране»;
  - видалення нерухомості з «Обране»;
  - створення шаблону фільтрації;
  - редагування шаблону фільтрації;
  - видалення шаблону фільтрації;
  - перегляд і зміна статусу власної нерухомості;
  - створення заявки на послугу агентства.
- г) система має забезпечити наступні функції для адміністратора:
  - функції доступні для зареєстрованого користувача;
  - складання договору купівлі чи продажу;

- контроль за дотриманням правил агентства нерухомості;
- обробка заявок та договорів;
- блокування користувачів [34].

### 5.3 Розробка діаграми варіантів використання системи

Для проектування системи було використано UML діаграму варіантів використання. Вона описує взаємини і залежності між групами варіантів використання і дійовими особами, які беруть участь в процесі, тобто вказує на те, що система повинна робити, не вказуючи як саме [35]. На рисунку 5.4 схематично зображений функціонал інформаційної системи для агентства нерухомості.

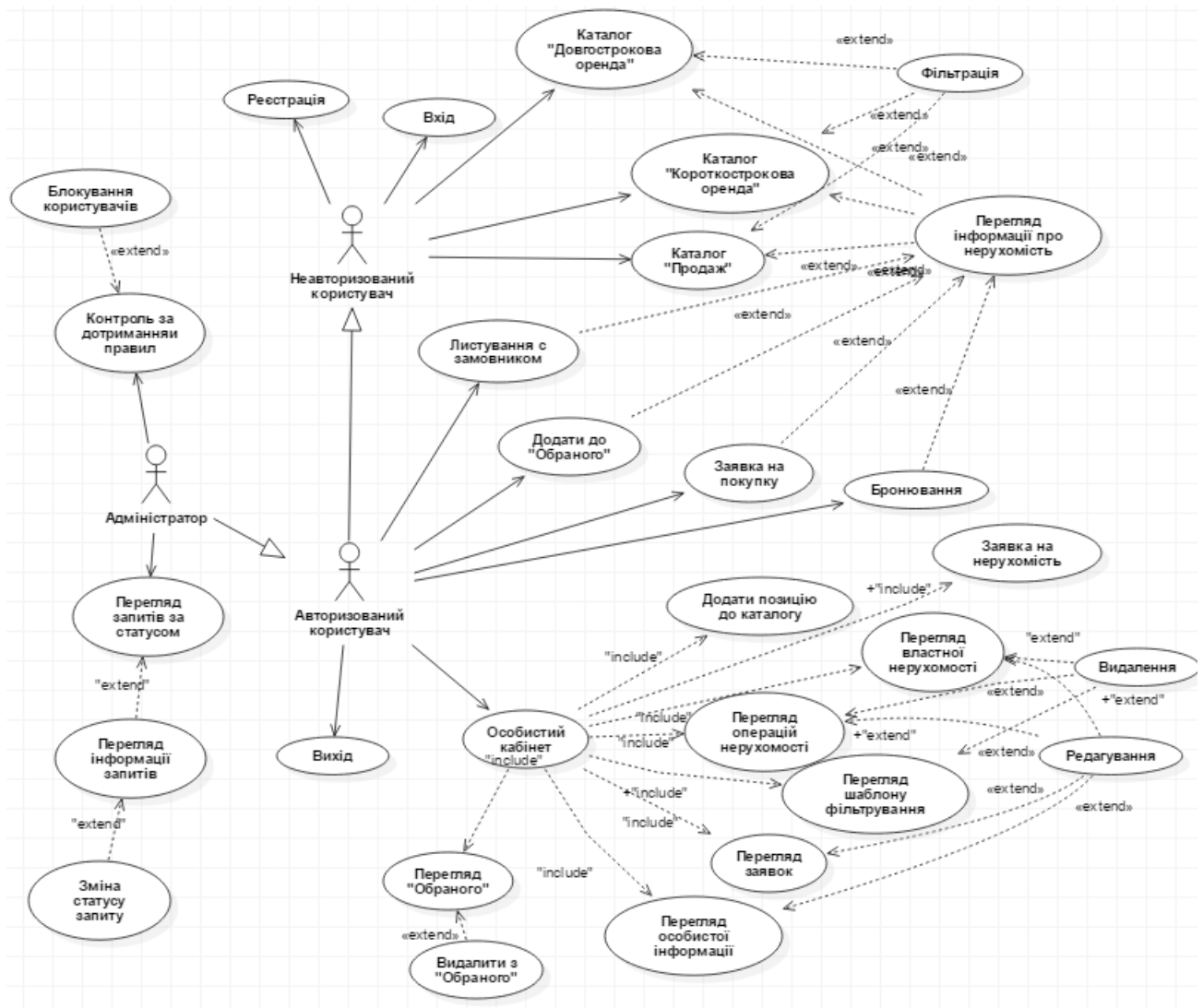


Рисунок 5.4 - Діаграми варіантів використання

На діаграми прецедентів представлені наступні актори:

- неавторизований користувач ;
- авторизований користувач;
- адміністратор.

Для неавторизованого користувача можливі такі варіанти використання:

- реєстрація в системі;
- авторизація в системі;
- перегляд каталогу («Короткострокова оренда», «Довгострокова оренда», «Продаж»). Кожен з каталогів може бути відфільтрованим за певними значеннями. Будь-яке оголошення нерухомості з каталогів може бути відкрите з відображенням детальної інформації.

Для авторизованого користувача можливі такі варіанти використання:

- вихід із системи;
- бронювання нерухомості;
- листування із замовником;
- додавання нерухомості до обраного;
- залишити заявку на покупку нерухомості;
- перегляд особистого кабінету, що включає в себе: можливість додати свою нерухомість до каталогів, створювати заявку на купівлю або орендування нерухомості, переглядати каталог «Обране» (та його редагування), переглядати власні оголошення нерухомості, бронювати нерухомість, переглядати власні заявки, та особисту інформацію (та редагувати власні дані) [35].

Для адміністратора можливі такі варіанти використання:

- контроль за дотриманням правил;
- перегляд запитів за статусом та їх редагування.

#### 5.4 Розробка вимог до інтерфейсу користувача інформаційної системи

Виходячи з аналізу предметної галузі, інформації про користувачів системи та сучасних тенденцій, інтерфейс користувача має бути розробленим в мінімалістичному стилі. У веб-дизайні мінімалізм характеризується відмовою від зайвих деталей, а також використанням малої кількості тексту, зображень та декоративних елементів. Кожен елемент спрямований на конкретне цільове завдання. В кольоровій палітрі, зазвичай, використовують тільки 2–3 кольори,

поєднання яких базуються на принципах гармонії кольору [36]. Для системи, що розроблюється рекомендовано обрати форми об'єктів – прямокутні, оскільки вони асоціюються з балансом, стабільністю, професіоналізмом, силою та ефективністю. Також система має відповідати евристичним стандартам:

- виводити користувачеві зрозумілі повідомлення при реєстрації , авторизації та бронюванні нерухомості для розуміння користувачем де він знаходить, та що відбувається в системі;
- система має «спілкуватися» з користувачем зрозумілою для нього мовою (українська, російська або англійська), а відображений текст, має бути інтуїтивно зрозумілі та не потребувати додаткового пояснення;
- система має надавати контроль та свободу користувачеві, тобто має бути передбачена можливість скасування бронювання, заявки на покупку тощо, якщо це дозволено бізнес-логікою конкретного випадку;
- всі компоненти системи мають бути реалізовані в одному стилі;
- послідовність операції має бути логічною та зрозумілою;
- передбачати можливі помилки в діях користувача, та не давати їм трапитися (уточнювати вірність його намірів при видаленні даних);
- дані введені раніше (особисті, дані фільтрації) мають дублюватися при переході з одного вікна в інше;
- підтримка використання «горячих» клавіш;
- допомога користувачеві у розпізнанні та виправленні помилок [16].

Система має надавати можливість обрати один із типів користувачів (незарєєстрований користувач, зарєєстрований користувач та адміністратор) та реалізувати основні бізнес-процеси системи, при цьому Зарєєстрованому користувачеві доступні всі можливості незарєєстрованого користувача, а Адміністратору доступні всі можливості зарєєстрованого користувача.

## 6 ПІДБІР ПАРАМЕТРІВ АЛГОРИТМУ

Підбір параметрів відбувається шляхом аналізу законів функціонування предметної галузі, даних, наданих спеціалістами в галузі – ріелторами, що виступають експертами в предметній галузі з одного боку, а також думки користувачів системи, що в дані ситуації виступають в ролі експертів з іншого боку. Експертні дані від користувачів системи зібрані шляхом соціального опитування проведеного в інтернеті.

Виходячи із законів функціонування предметної галузі для роботи системи є два параметри які користувач має обов'язково вказати: тип нерухомості (короткострокова, довгострокова оренда та продаж) та місто, в якому необхідно шукати нерухомість. Ці параметри є ключовими, тому для кожного користувача система формує окремі рекомендації за всіма типами операцій з нерухомістю та за кожним містом. Наприклад, якщо користувач на протязі тижня шукає 2-х кімнатну квартиру поруч з містом роботи (станція метро Наукова, місто Харків) для купівлі і отримує відповідні рекомендації, а сьогодні дізнався про відрядження та ввів у пошук нові параметри: короткострокова оренда та місто Київ. Ні одна з рекомендацій по місту Харків не буде актуальною в цьому випадку, так само як і купляти нерухомість в Києві, щоб пожити там декілька днів він навряд чи буде. Більше того, є мала ймовірність, що в такій квартирі його будуть хвилювати такі питання як: яким чином нагрівається вода (за допомогою бойлеру, газової установки чи централізованої подачі гарячої води), в якому році чи з якого матеріалу побудований будинок.

Тільки за відсутністю даних у користувача за певним містом, система може використати дані із цього ж типу операцій, але з іншого міста, при цьому надаючи дані тільки по шуканому місту. За відсутності інформації у користувача за певним типом операції взагалі, для довгострокової оренди можуть бути використані дані з покупки нерухомості і навпаки. За рахунок різних критеріїв вибору цих типів з короткостроковою оренду запозичувати у них дані для короткострокової оренди

сенсу немає. У разі повної відсутності даних, що система може використати для формування рекомендації – застосовується метод на основі популярності.

На думку ріелторів, кількості елементів, із яких формується «ковзаюче вікно» повинно залежати від часу, що пройшов від моменту попередньої рекомендації. Так, якщо з моменту попередньої рекомендації пройшло менше 7 діб кількість елементів дорівнює – 20, якщо більше 7 але менше 15 днів – 10 елементів, якщо більше 15 – 3 елементи. Це пов'язано з тим, що з плином часу побажання користувача часто змінюються з набуттям нового досвіду та впливом зовнішніх факторів, але враховуючи велику кількість параметрів – якомога більша вибірка надає найбільш релевантні дані.

Для визначення ступеня важливості критеріїв було проведено соціальне опитування користувачів, в якому їх просили оцінити кожен характеристику від 1 до 10 балів (рис. 6.1 – 6.3). З отриманих даних обрано середнє значення за кожною характеристикою. У таблиця 6.1 – 6.2 наведено характеристики нерухомості за типами, їх важливість, а також значення параметрів для розрахунку оцінки близькості, обраних відповідно до рекомендацій ріелторів.



Рисунок 6.1 – Оцінка важливості характеристик для короткострокової оренди

Таблиця 6.1 – Характеристики та їх відносна важливість для операцій з короткостроковою орендою.

№	Характеристика	Середня оцінка експертів		Назви параметрів		
				Offset	Scale	Decay
1	Ціна	10	0,05	4%	15%	0,5
2	Тип помешкання (все помешкання, окрема кімната, спальне місце)	9	0,045	0	1	0,5
3	Кількість спальних місць	8	0,04	0	1	0,5
4	Рейтинг орендодавця	9	0,045	0	1	0,5
5	Можливість самостійного прибуття	6	0,03	0	1	0,5
6	Мова господаря	1	0,005	0	1	0,5
7	Умови проживання (з тваринами чи)	6	0,03	0	1	0,5
8	Поверх	3	0,015	0	2	0,5
9	Наявність пакувального місця	7	0,035	0	1	0,5
10	Наявність телебачення	6	0,03	0	1	0,5
11	Наявність інтернету	9	0,045	0	1	0,5
12	Наявність кондиціонера	8	0,04	0	1	0,5
13	Близькість до центру міста	8	0,04	0	0,2	0,5
14	Близькість до залізничного чи автомобільного вокзалу	6	0,03	0	0,2	0,5
15	Близькість до головного туристичного місця в місті	6	0,03	0	0,2	0,5
16	Наявність додаткових технічних засобів (пральна/ посудомийна машина, праска)	7	0,035	0	1	0,5
17	Наявність передплати	5	0,025	0	1	0,5
18	Наявність кухні	4	0,02	0	1	0,5
19	Особливості планування	1	0,005	0	1	0,5
20	Додаткові зручності (наявність басейну, каміну)	1	0,005	0	1	0,5

Таблиця 6.2 – Характеристики та їх відносна важливість для операцій з довгостроковою орендою та купівлею.

№	Характеристика	Оцінка (довгострокова оренда)		Оцінка (продаж)		Назви параметрів		
						Offset	Scale	Decay
1	Ціна	10	0,05	4%	15%	4%	15%	0,5
2	Тип помешкання (все помешкання, окрема кімната)	9	0,045	-	-	0	1	0,5
3	Рейтинг орендодавця	8	0,04	-	-	0	1	0,5
4	Поверх	4	0,02	4	0,3	1	2	0,5
5	Наявність пакувального місця	7	0,035	7	0,053	0	1	0,5
6	Близькість до центру міста	8	0,04	8	0,061	0	0,2	0,5
7	Наявність додаткових технічних засобів (праска, пральна/посудомийна машина)	9	0,045	-	-	0	1	0,5
8	Особливості планування	7	0,035	7	0,053	0	1	0,5
9	Умови проживання (з тваринами чи дітьми)	8	0,04	-	-	0	1	0,5
10	Тип стін	2	0,01	2	0,015	0	3	0,5
11	Опалення (відсутнє, централізоване, індивідуальне)	9	0,045	9	0,069	0	1	0,5
12	Наявність комунальних	10	0,05	-	-	0	1	0,5
13	Тип пропозиції (посередник, власник забудовник)	7	0,035	-	-	0	1	0,5
14	Період здачі	8	0,04	-	-	2 діб	5 діб	0,5

Продовження таблиці 6.2

№	Характеристика	Оцінка (довгострокова оренда)		Оцінка (продаж)		Назви параметрів		
		Offset	Scale	Decay				
15	Кількість поверхів	2	0,01	2	0,015	0	2	0,5
16	Адміністративний район;	9	0,045	9	0,069	0	1	0,5
17	Кількість кімнат	8	0,04	8	0,061	0	1	0,5
18	Загальна площа	6	0,03	6	0,046	6%	18%	0,5
19	Житлова площа	7	0,035	7	0,053	5%	12%	0,5
20	Близькість до метро	8	0,04	8	0,061	0	0,2	0,5



Рисунок 6.2 – Оцінка важливості характеристик для довгострокової оренди

Важливість реакції користувача, згідно із законами функціонування предметної галузі обрана наступна:

- нерухомість додана до обраного - 0,45;
- явна оцінка оголошення користувачем - 0,2;

- відкриття сторінки детального перегляду оголошення - 0,1;
- час перегляду оголошення- 0,25 (рис. 6.4).



Рисунок 6.3 – Оцінка важливості характеристик для купівлі



Рисунок 6.4 – Важливість впливу реакції користувача на оголошення

Оскільки основним методом рекомендаційної системи є метод фільтрації заснований на контенті, а допоміжним метод спільної фільтрації на основі сусідства «користувач-користувач», було прийнято рішення використовувати ці методи незалежно, а отриманні оголошення поєднувати у наступному співвідношенні: 5 оголошень рекомендовані за допомогою методу фільтрації заснованому на контенті до 1 оголошення рекомендованого за допомогою методу спільної фільтрації. Тобто, спочатку буде надано перші 5 оголошень рекомендовані за допомогою методу фільтрації заснованому на контенті, потім одне оголошення з методу спільної фільтрації (рекомендовано 6 оголошень), далі 6-10 оголошення з методу фільтрації заснованому на контенті (рекомендовано 11 оголошень), після нього – друге оголошення з методу спільної фільтрації (рекомендовано 12 оголошень), і так далі.

## 7 РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

### 7.1 Опис розробленої бази даних

При проектуванні системи були пред'явлені вимоги до бази даних системи. Моделювання даних виконано за допомогою CASE-застосунку «Allfusion ErWin Data Modeler», що призначений для проектування та документації баз даних [33]. Основні сутності системи наведені на рисунку 7.1.

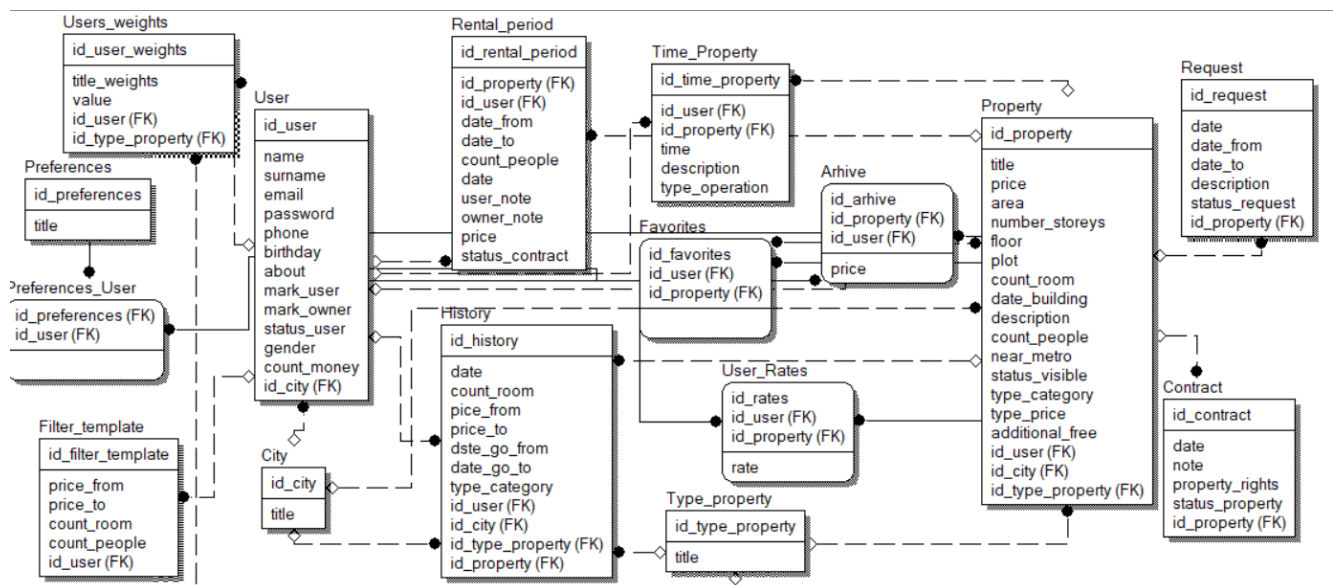


Рисунок 7.1 – Логічна модель даних системи

У відповідності до пред'явлених вимог були визначені наступні сутності:

- user – містить інформацію про клієнтів;
- property: містить інформацію про нерухомість;
- contract: містить інформацію про договори на операції з нерухомістю;
- request: містить інформацію про заявки на пошук;
- rental\_period: містить інформацію про бронювання нерухомості;
- arhive: містить дані про всі операції клієнта;
- filter\_template: містить дані про шаблони фільтрування;
- favorites: містить список нерухомості, що користувач додав до Обраного;
- history: містить дані про всі дії користувачів;
- preferences: містить побажань орендодавців, якщо вони не бажають здавати нерухомість людям з певними характеристиками;

- time-property: містить інформацію про час затрачений користувачем на перегляд інформації;
- user\_weights: містить інформацію про вагові коефіцієнти встановлені користувачем;
- user\_rates: містить дані про оцінку нерухомості користувачем.

## 7.2 Опис алгоритмів інформаційної системи

Алгоритм роботи інформаційної системи в цілому наведено на діаграма діяльності (рис. 7.2), що містить інформацію не тільки про послідовність виконання дій, але і про те, хто має їх виконувати. Розглянуто декілька варіантів розвитку подій починаючи від потрапляння користувача на головну сторінку сайту, до завершення бронювання – оплати, а також зони відповідальності користувача та системи за допомогою «доріжок».

При вході користувача на сайт, система виводить на екран головну сторінку з рекомендаціями для перегляду. Користувач може переглядати інформацію, та вводити дані фільтрації. Система виконує запит фільтрації та виводить на екран оновлені дані. Користувач натискає на конкретне оголошення нерухомості, на що система відповідає виведенням на екран детальної інформації про неї. Користувач натискає кнопку «Забронювати». Система спочатку перевіряє, чи авторизований користувач: якщо авторизований – виводить форму бронювання (маркер 1), якщо не авторизований – система відображає форму авторизації, з попередженням, що бронювання доступне тільки авторизованим користувачам. Користувач має ввести у відповідних полях логін і пароль (або зареєструватися). Система перевіряє коректність введених даних [31]. У разі помилки, система повідомляє про це користувача. (Маркер 1) Якщо бронювання не можливе (на дані дати вже є підтвержене бронювання) система має повідомляє про це користувача, та перенаправляє його до головної сторінки сайту. Якщо бронювання можливе - система формує акт оренди нерухомості та пропонує користувачеві його підтвердити, корегувати або відхилити. Користувач може обрати будь-яку з цих дій.

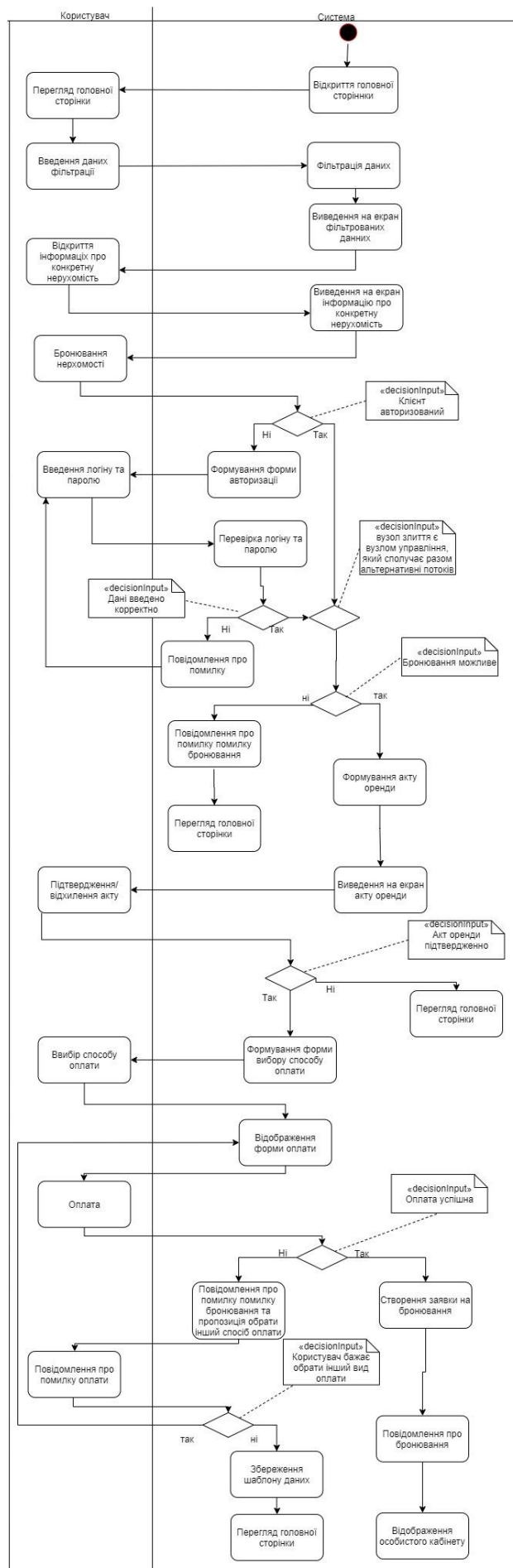


Рисунок 7.2 – Діаграма діяльності бронювання нерухомості

Якщо користувач відхилив акт оренди нерухомості, система повертає його на головну сторінку сайту не створивши заявку на бронювання. Якщо користувач підтверджує акт оренди (або коригує та підтверджує), система пропонує користувачеві обрати спосіб оплати [36]. Якщо оплата не відбулася з будь-яких причин, система сповіщає про це користувача та пропонує виконати дію повторно. Якщо користувач з цим погодиться, система повертає його на сторінку вибору способу оплати, якщо користувач не бажає повторювати спробу система зберігає шаблон ведених даних та перенаправляє користувача на головну сторінку системи. Якщо оплата пройшла успішно система завершує процес бронювання, тобто створює заявку на бронювання, та сповіщає про це користувача – перенаправляє користувача на сторінку особистого кабінету, з відображенням таблиці заброньованої нерухомості.

На рисунку 7.3 наведено алгоритм формування рекомендацій. При завантаженні будь-якої сторінки системи користувачем – алгоритм створення рекомендацій оновлює свої дані для надання найбільш релевантних рекомендацій. Для початку система перевіряє чи є в неї дані про обмеження користувача (місто, в якому знаходиться нерухомість, дата заїзду, площа нерухомості). Якщо такі дані наявні – система одразу відфільтрує нерелевантні дані для прискорення роботи алгоритму. Наступною важливою інформацією є наявність даних про вподобання користувача. Найкраща ситуація для цього – користувач зареєструвався та авторизувався в системі. В цьому випадку є велика ймовірність наявного профілю користувача, історії операцій, історії переглядів, вподобань та користувальницьких вагових характеристиках користувача. Тому спочатку перевіряється чи є даний користувач авторизованим. Якщо користувач не авторизований – система перевіряє чи є якісь наявні дані про користувача в системі (дані з сесії або кукі). Якщо про користувача немає ніяких даних, і персоналізовані рекомендації система надати не зможе – рекомендації формуються на основі популярності оголошень. Якщо користувач не авторизований, але система має якісь дані про вподобання користувача, на їх основі система розраховує «близькість» кожного оголошення за кожною з характеристик та створює відповідну матрицю «близькості» оголошень.

Також формується матриця «схожих» користувачів. На основі даних з цих матриць формуються списки рекомендованих оголошень, які надаються користувачеві в певному співвідношенні (5 оголошень з першої матриці та 1 оголошення з другої матриці).

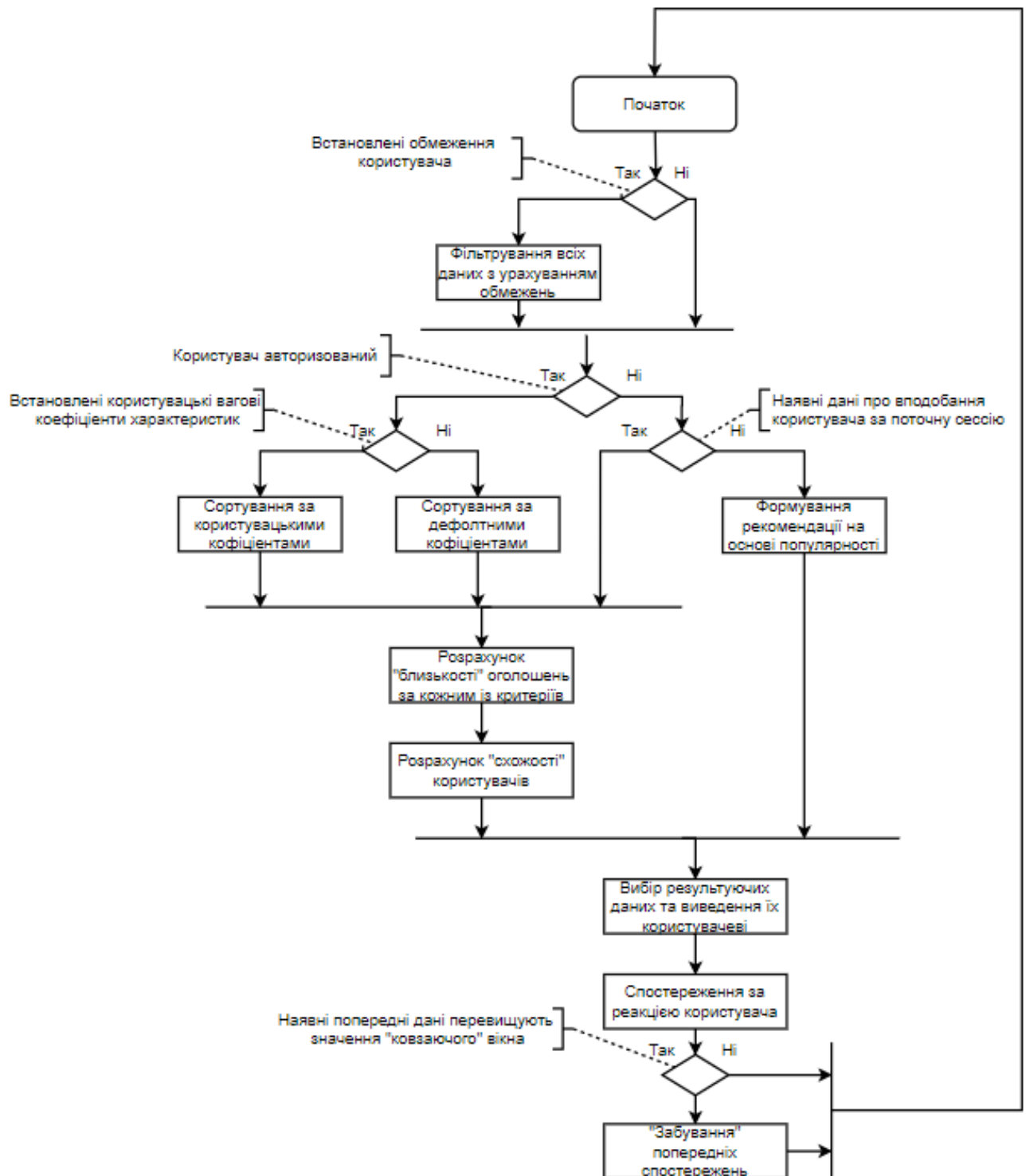


Рисунок 7.3 – Схема формування рекомендацій

Далі система спостерігає за реакцією користувача на запропоновані рекомендації та зберігає ці дані (в базу даних для авторизованих користувачів в дані куки та сесії для неавторизованих користувачів). Якщо збережені дані перевищують значення «ковзаючого» вікна система «забуває» попередні значення, тобто, вони залишаються в базі даних але не впливають на подальшу роботу алгоритму. Подібний алгоритм виконується кожного разу при будь-якій взаємодії користувача з системою.

### 7.3 Розробка інтерфейсу клієнтської частини

Інтерфейс користувача створено за допомогою мови розмітки гіпертексту HTML та мови стилю сторінок CSS, з використанням мови програмування JavaScript, бібліотеки JQuery та фреймворку Bootstrap. Команди PHP включаються в HTML-сторінки за допомогою спеціальних тегів, які і виконують потрібні дії на сервері за допомогою PHP-машини [26].

При запуску головної сторінки сайту відображається сторінка з основними розділами системи: Довгострокова оренда (що містить дані про нерухомість, що здається в оренду помісячно), Короткострокова оренда (що містить дані про нерухомість, що здається в оренду подовово), Продаж (що містить дані про нерухомість, що може бути куплена) (рис. 7.4). Маніпулятором миша можливо, натиснувши по назві одного із розділів, обрати необхідний розділ.

Натиснувши на один із розділів, наприклад, розділ Короткострокова оренда – відкривається сторінка, що відображає каталог нерухомості, що містить дані про нерухомість, що може бути орендована подовово (рис. 7.4). Обравши на один із розділів, наприклад, розділ Короткострокова оренда – відкривається сторінка, що відображає каталог нерухомості, що містить дані про нерухомість, яка може бути орендована подовово (рис. 7.5).

На сторінці Короткострокової оренди можливі наступні функції:

- перегляд каталогу нерухомості;
- фільтрація каталогу;
- додавання до «Обране»;
- перехід до сторінки бронювання певної нерухомості;
- перехід до сторінки детальної інформації певної нерухомості.

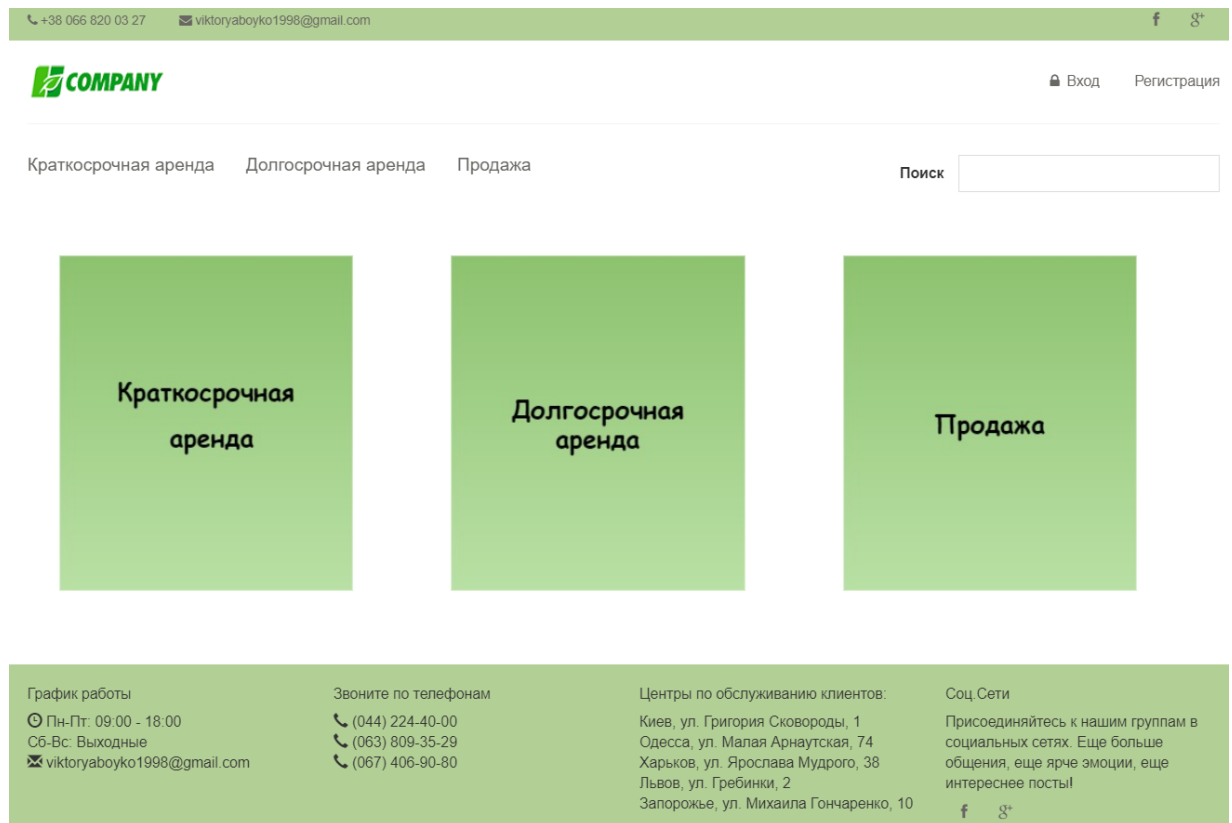


Рисунок 7.4 – Схема формування рекомендацій

Для того, щоб переглянути більше позицій нерухомості необхідно прокрутити вікно сторінки донизу, та натиснути на стрілочку «>>» або натиснути на бажану цифру, що позначає номер сторінки.

Для фільтрування даних необхідно натиснути на поле для введення «Дата» та ввести її з клавіатури, або за допомогою стрілочок, які з'являється в полі для введення «Дата», при наведенні на неї курсору миші. При натисненні на парні стрілочки (вверх та вниз) можна по черзі налаштувати дату, місяць та рік. При натисненні на одинарну стрілочку з'являється звичний кожному користувачеві календар, на якому можна обрати дату, натиснувши на відповідний номер (день місяця) певного місяця – поле для вводу заповниться автоматично. Обраним чином потрібно заповнити два поля – дату початку бронювання та дату кінця бронювання. В поле пункт проживання необхідно вручну ввести місто, в якому необхідно шукати нерухомість. В пункті кількість людей, вибрати кількість людей, яким необхідна нерухомість для цього потрібно один раз натиснути ЛКМ (лівою клавішею миші) в будь-якому місці поля Кількість людей, що призведе до появи повного списку варіантів. Для вибору необхідного – по ньому потрібно ще раз натиснути ЛКМ.

Краткосрочная аренда Долгосрочная аренда Продажа

Поиск

## ПОДБОР

Дата

Пункт назначения

Количество людей

## КАТАЛОГ АРЕНДЫ







 <p><b>96100 грн</b></p> <p>р-н. Поділля вул. Зодчих м. Вінниця</p> <p><input type="button" value="Забронировать"/></p>	 <p><b>90600 грн</b></p> <p>р-н. Поділля вул. Академіка Ющенка м. Вінниця</p> <p><input type="button" value="Забронировать"/></p>	 <p><b>54700 грн</b></p> <p>р-н. Павлове Поле 23 августа м. Харків</p> <p><input type="button" value="Забронировать"/></p>
 <p><b>74000 грн</b></p> <p>р-н. Центр вул. Дарвіна м. Харків</p> <p><input type="button" value="Забронировать"/></p>	 <p><b>75900 грн</b></p> <p>р-н. Нові Дома вул. Ньютона м. Харків</p> <p><input type="button" value="Забронировать"/></p>	 <p><b>79500 грн</b></p> <p>р-н. Центр просп. Московський м. Харків</p> <p><input type="button" value="Забронировать"/></p>

Рисунок 7.5 – Відображення в браузері сторінки каталогу


Після введення даних фільтрації (будь-яке поле може бути як і введене, так і не введене) необхідно натиснути на кнопку «Фільтрувати». Це призведе до того, що буде відображатися тільки та нерухомість, що відповідає введеним даним фільтрації.

Для додавання нерухомості до «Обране» необхідно бути авторизованим та натиснути на сердечко, що знаходиться під інформацією про нерухомість – під її назвою та поруч із кнопкою «Забронювати». Якщо нерухомість додана до «Обране» - сердечко виділене зеленим кольором. Якщо додати до «Обране» намагається користувач, що не авторизований – при натисненні на кнопку він буде перенаправлений на сторінку авторизації з повідомлення про необхідність авторизації для такої дії.

Для перегляду більш детальної інформації про певну нерухомість необхідно натиснути ЛКМ на будь-які дані про цю нерухомість (її зображення, ціна або назва).

Сторінка детальної інформації про нерухомість представлена на рисунку 7.6. На цій же сторінці наявні рекомендаційні оголошення для подальшого перегляду.

Для початку бронювання нерухомості необхідно натиснути на кнопку «Забронювати», що знаходиться під інформацією про нерухомість – під її назвою. Це призведе до перенаправлення користувача на сторінку детальної інформації про нерухомість для підтвердження її бронювання (рис. 7.6).



**Et Limited**

Код товару: 1


**6617 грн**

Общая площадь: 2  
 Участок у здания: 1  
 Этаж: 3  
 Этажность здания: 6  
 Количество комнат: 4  
 Дата постройки: 2019-05-12  
 Вид категории: Дом  
 Город: Харьков  
 Близость метро: 10 мин

#### Описание недвижимости


aliquet libero. Integer in magna. Phasellus dolor elit, pellentesque a, facilisis non, bibendum sed, est. Nunc laoreet lectus quis massa. Mauris vestibulum, neque sed dictum eleifend, nunc risus varius orci, in consequat enim diam vel arcu. Curabitur ut odio vel est tempor bibendum. Donec felis orci, adipiscing non, luctus sit amet, faucibus ut, nulla. Cras eu tellus eu augue porttitor interdum. Sed auctor odio a purus. Duis elementum, dui quis accumsan convallis, ante lectus convallis est, vitae sodales nisi magna sed dui. Fusce aliquam, enim nec tempus scelerisque, lorem ipsum sodales purus, in molestie tortor nibh sit amet orci. Ut sagittis lobortis mauris. Suspendisse aliquet molestie tellus. Aenean egestas hendrerit neque. In ornare sagittis felis. Donec tempor, est ac mattis semper, dui lectus rutrum urna, nec luctus felis purus ac tellus. Suspendisse sed dolor. Fusce mi lorem, vehicula et, rutrum eu, ultrices sit amet, risus. Donec nibh enim, gravida sit amet, dapibus id, blandit at, nisi. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Proin vel nisi. Quisque fringilla euismod enim. Etiam gravida molestie arcu. Sed eu nibh vulputate mauris sagittis placerat. Cras dictum ultricies ligula. Nullam enim. Sed nulla ante, iaculis nec, eleifend non, dapibus rutrum, justo. Praesent luctus. Curabitur egestas nunc sed libero. Proin sed turpis nec mauris blandit mattis. Cras eget nisi dictum augue malesuada malesuada. Integer id magna et ipsum cursus vestibulum. Mauris magna. Duis dignissim tempor arcu. Vestibulum ut eros non enim commodo hendrerit. Donec porttitor tellus non magna. Nam ligula elit, pretium et, rutrum non, hendrerit id, ante. Nunc mauris sapien, cursus in, hendrerit consectetur, cursus et, magna. Praesent interdum ligula eu enim. Etiam imperdiet dictum magna. Ut tincidunt orci quis lectus. Nullam suscipit, est ac facilisis facilisis, magna tellus faucibus leo, in lobortis tellus justo sit amet nulla. Donec non justo. Proin non massa non

#### РЕКОМЕНДАЦИИ




**74000 грн**

р-н. Центр вул. Дарвіна м. Харків



**75900 грн**

р-н. Нові Дома вул. Ньютона м. Харків



**79500 грн**

р-н. Центр просп. Московський м. Харків

Рисунок 7.6 – Відображення детальної інформації про нерухомість

## 8 ОЦІНКА АЛГОРИТМУ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ

Оцінка алгоритму формування рекомендаційної системи здійснювалась за допомогою збору даних у реальному часу від користувачів в мережі Інтернет на трьох платформах які використовували контентний, колаборативний та гібридний алгоритми формування рекомендацій відповідно. Результати оцінювання ефективності розраховувані за допомогою середньоквадратичного відхилення (RMSE) для трьох типів операцій з нерухомістю (короткострокова оренда, довгострокова оренда та купівля/продаж) та трьох типів алгоритмів ( на основі вмісту, на основі спільної фільтрації та на основі розробленого гібридного методу). Результати експериментів, наведені у таблиці 8.1 та на рисунку 8.1.

Таблиця 8.1 – Порівняльна характеристика алгоритмів формування рекомендацій.

Алгоритм	Тип операції	Показник RMSE
На основі вмісту	Короткострокова оренда	0,4
	Довгострокова оренда	0,49
	Купівля/продаж	0,47
На основі спільної фільтрації	Короткострокова оренда	0,36
	Довгострокова оренда	0,51
	Купівля/продаж	0,52
Гібридний	Короткострокова оренда	0,29
	Довгострокова оренда	0,33
	Купівля/продаж	0,35

Згідно з результатами експериментів – розроблений алгоритм дає найбільш точні результати за показником RMSE по всіх типах операцій (0,29; 0,33; 0,35), що є досить точним результатом.

Варто відзначити, що для короткострокової оренди значення також є не великими по всім алгоритмам, що пояснюється тим, що для даного типу операції активність користувачів більш повторювана (раз на рік Користувач орендує квартиру в Львові для відпочинку), більшість користувачів залишають відгуки за

своїми враженнями, за рахунок чого матриці «елемент-користувач» менш розріджені, а профіль користувача має бодай мінімальну інформацію. Також, оголошення, в порівнянні з іншими типами більш статичні – нерухомість може бути орендована на одну дату одним користувачем, потім іншим користувачем на наступну дату і т.д. В разі купівлі обраної нерухомості, з нею більше не можуть взаємодіяти інші користувачі, а в разі довгострокової оренди, повторна взаємодія може відбутися через досить тривалий час (від декількох місяців до років).

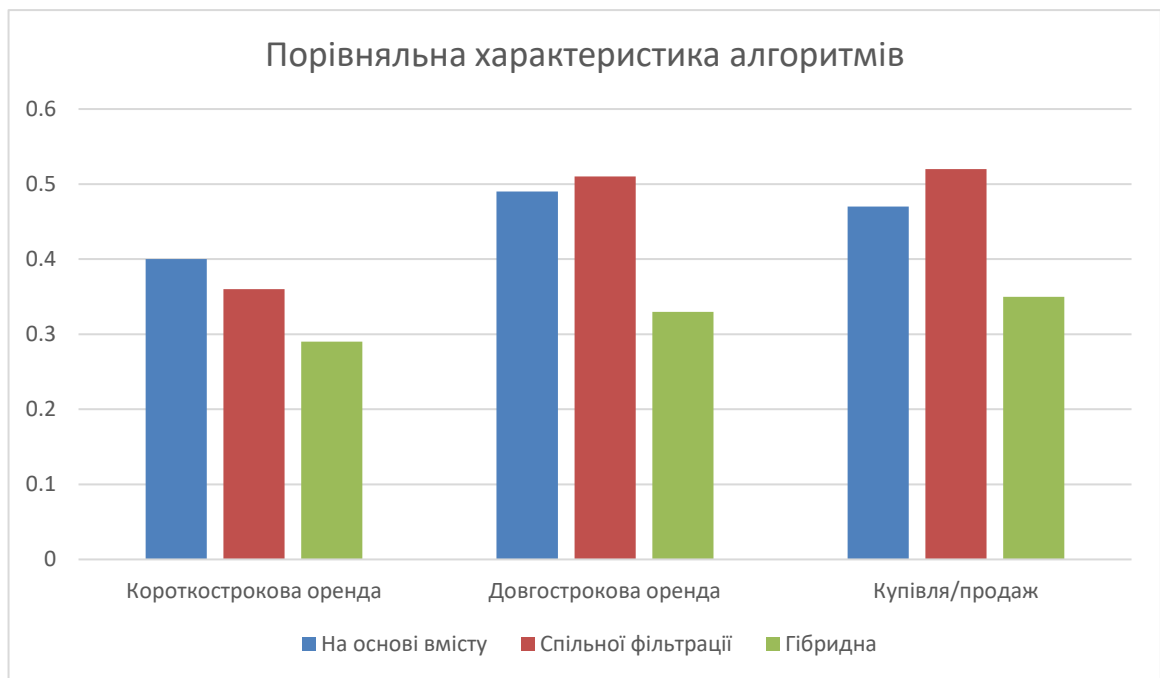


Рисунок 8.1 – Порівняння роботи алгоритмів формування рекомендацій

Оскільки розроблений алгоритм дає найбільш точні результати, розглянемо його більш детально. Результати впливу того, що користувач авторизований чи ні наведено на рисунках 8.2. В рамках дослідження враховувалися перші 10 оцінок у кожного користувача. Аналіз наступних оцінок не має сенсу включати до даного досліджу оскільки при отриманні системою цих даних про користувача – він вже не буде вважатися зовсім невідомим користувачем, а його подальші оцінки не значно відрізняються від авторизованого користувача. З наведених даних бачимо, що для короткострокової оренди те, що користувач був авторизованим вплинуло більше на кількість оцінок 4 та 5 (бажані оцінки), та мало вплинуло на кількість оцінок 1,

2 та 3 (не бажані оцінки). Для довгострокової оренди та купівлі/продажу наявність профіля користувача вплинула доволі сильно за всіма оцінками.

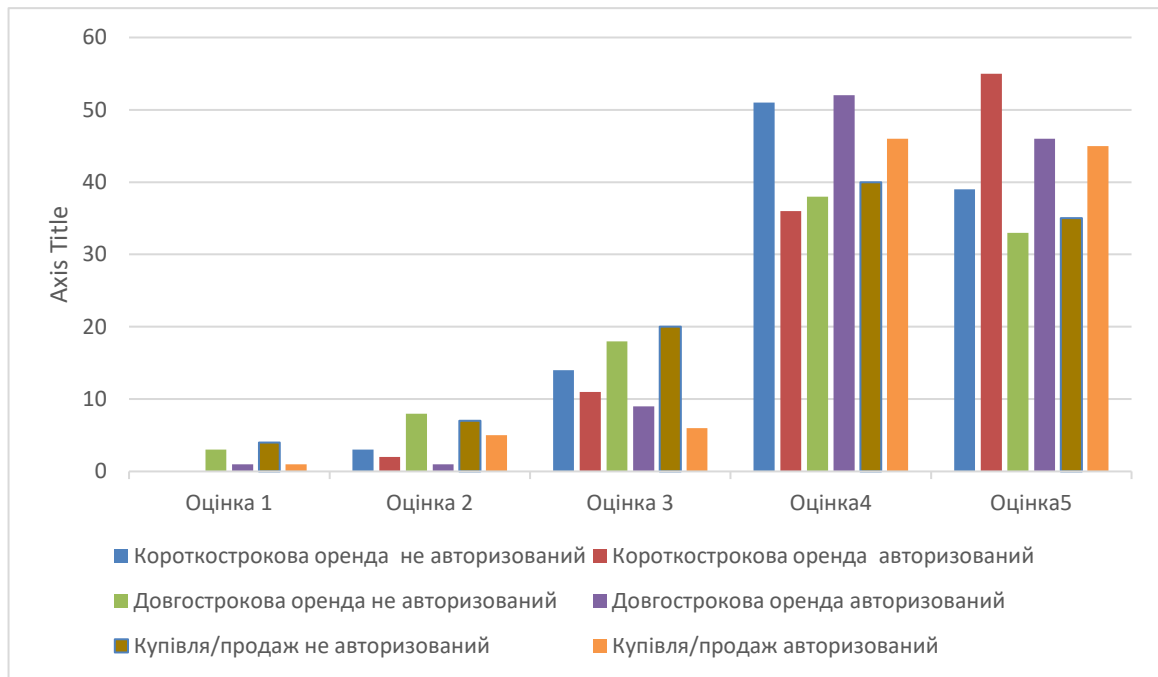


Рисунок 8.2 – Порівняння роботи алгоритмів формування рекомендацій для авторизованого та неавторизованого користувача

На рисунку 8.3 показано, як оцінки зроблені при перших переглядах вплинули на результати алгоритму через пів року. Для порівняння використовувалися оцінки від однієї і тієї ж людини. З графіка можемо бачити, що результати дуже схожі. Тобто, обрана система досить гарно враховує зміну побажань користувачів з плином часу.

Також було перевірено, як впливає на результати рекомендацій встановлення власних вагових коефіцієнтів кожним користувачем (рис. 8.4). В такому випадку результат майже ідеальний. Чи може це слугувати приводом для того, щоб змінити значення вагових коефіцієнтів? Однозначно ні. По-перше їх значення також було отримано від експертів-користувачів які були підтверджені експертами-ріелторами. По-друге, така ситуація виникає через те, що всі люди дуже різні та мають різні уподобання (для людини, що має автомобіль наявність пакувального місця – зайва трата грошей, і пріоритет на цю характеристику буде нульовий, а для

людини, що має автомобіль – дана характеристика є дуже важливою), а як ми бачимо, вагові коефіцієнти все одно дуже позитивно впливають на якість рекомендацій.

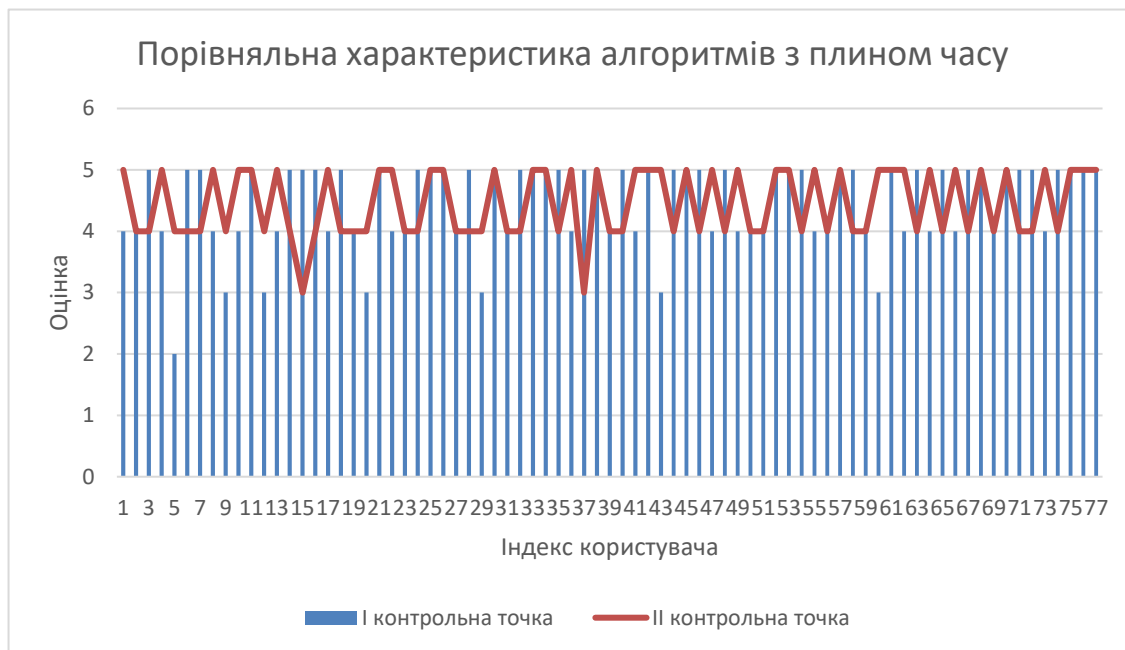


Рисунок 8.3 – Порівняння роботи алгоритмів роботи алгоритму з плином часу

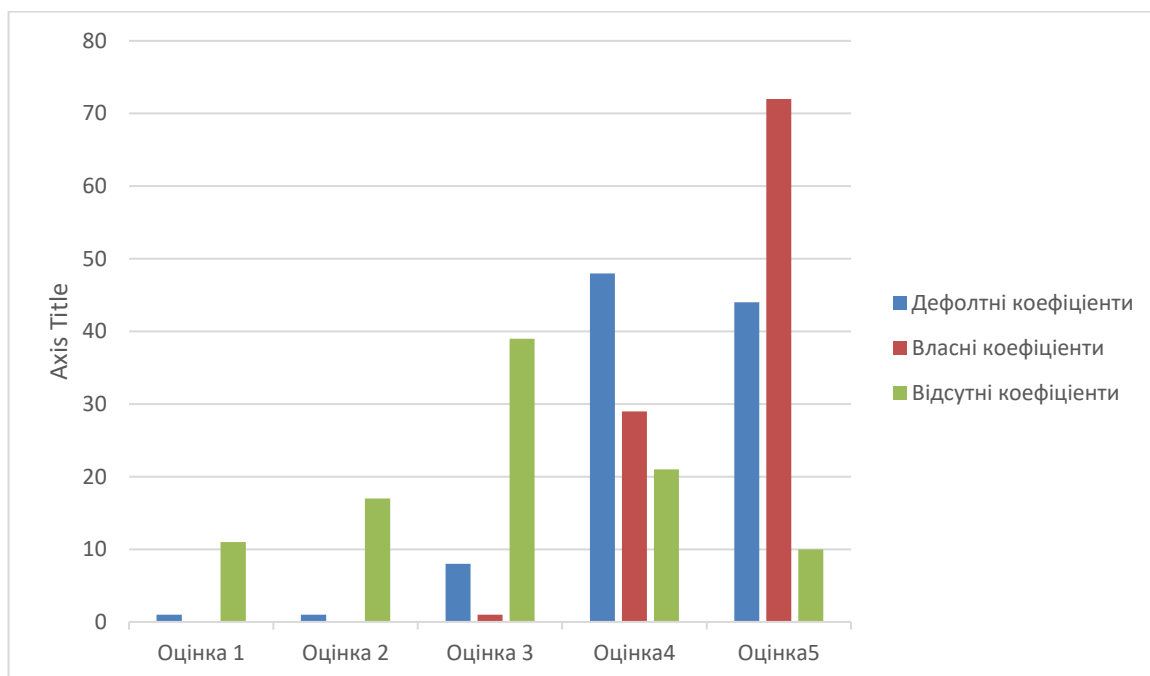


Рисунок 8.4 – Порівняння роботи алгоритмів роботи з різними ваговими коефіцієнтами

Результат без внесених даних про вагові коефіцієнти дуже поганий та надто хвилеподібний – оголошення в системі мають велику кількість характеристик, і система не розуміє при яких співпаданнях значень варто рекомендувати оголошення, а при яких воно навпаки не сподобається користувачеві. Тобто, внесені дані про важливість коефіцієнтів на основі думки експертів є ефективним для роботи рекомендаційної системи, та ця ефективність може бути покращення для кожного користувача окремо. Саме тому, важливо постійно рекомендувати користувачам встановлювати власні коефіцієнти важливості, а також систематично оновлювати їх.

Виходячи з того, що в реальних умовах користувачі залишають зворотній відгук не за кожним із оголошень, а в нашому дослідженні користувачі оцінювали 100% запропонованих оголошень, було проведено дослідження про те як впливає зворотній зв'язок користувача на результати рекомендаційної системи (рис. 8.5). Для цього в одній із тестових систем – система «бачила» оцінки користувачів та використовувала їх для подальших рекомендацій, а в другій тестовій системі – користувачі все так же оцінювали запропоновані оголошення, але ці дані не були використані рекомендаційним алгоритмом, а були зібрані тільки для статистичної оцінки. Як результат бачимо, що без зворотного зв'язку від користувача система надає менш релевантні рекомендації, але вони все ще досить на високому рівні. Це можливо завдяки збору неявних даних про вподобання користувача (кількість часу, що користувач переглядав оголошення, відкриття детальної інформації та інше).

На рисунку 8.6 наведено результати порівняння розробленого алгоритму, з подібним до досліджуваного, але без частини, що включає спільну фільтрацію.

Виходячи тільки із кількісних показників можна було б зробити висновок, що додавання аспектів із методу спільної фільтрації погіршує надані рекомендації – збільшилося кількість оцінок 3, 2 та 1, що доволі сильно впливає на алгоритм RMSE. При цьому можна помітити, що кількість вищої оцінки не зменшилася. В ході даного дослідження після оцінювання запропонованих рекомендацій користувачеві пропонувалося надати суб'єктивну думку про різноманітність чи подібність представлених оголошень та відношення до цього користувача.

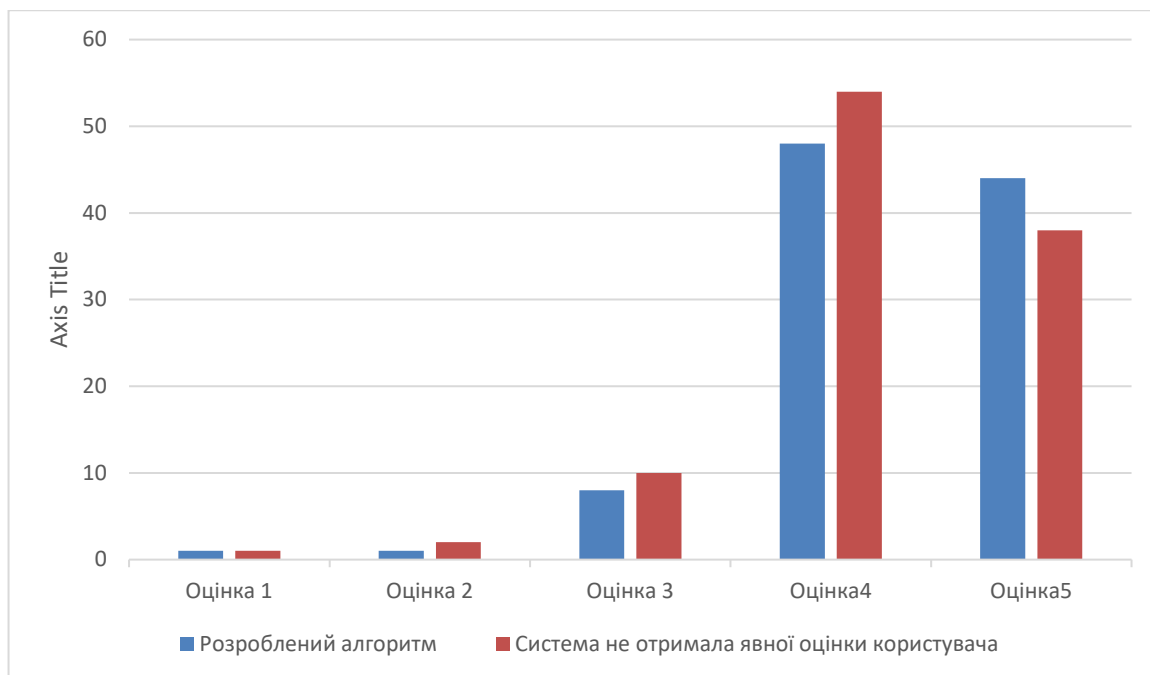


Рисунок 8.5 – Порівняння роботи алгоритмів роботи з веденням явної оцінки користувача та без оцінок

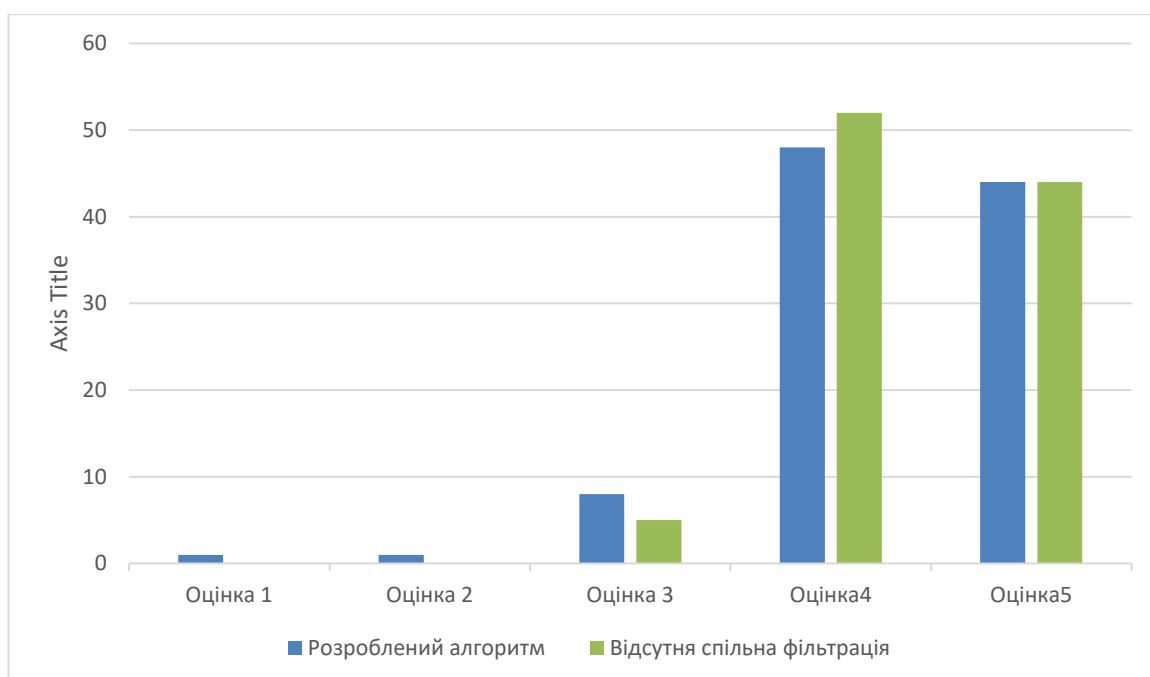


Рисунок 8.6 – Порівняння роботи алгоритмів роботи з веденням явної оцінки користувача та без оцінок

Виявилось, що незважаючи на збільшення кількості низьких оцінок користувача, загальна його задоволеність – збільшилася. Це пояснюється тим, що алгоритм спільної фільтрації вносить додаткові оголошення, які відповідають

«сусіднім» користувачам, що привносять ефект неочікуваності рекомендацій. Неочікуваність представляє собою деякий сюрприз у списку рекомендацій, несхожість на історію дій користувача. Збільшення різноманітності, неочікуваності та новизни рекомендацій може знизити точність прогнозування та точність ранжування, в той же час може підвищитися покриття та різноманітність каталогу оголошень [37, 38].

Практичні дослідження підтвердили ефективність впровадження розробленого алгоритму формування рекомендацій, що можуть слугувати основою для прийняття реальних рішень при роботі з нерухомістю. Застосування гібридного підходу дає найкращий результат, бо використовує переваги традиційних підходів та компенсує недоліки кожного з них окремо. Таким чином розроблений метод є гнучким та досить точним за RMSE метрикою, а також відгуками користувачів при роботі з тестовими системами. Реальні його дані можуть бути покращені за допомогою активної допомоги користувача в цьому.

## ВИСНОВКИ

У результаті виконання кваліфікаційної роботи виконано огляд існуючої проблеми її актуальності, а також методи вирішення в рамках формування рекомендацій в сфері оренди та продажу нерухомості.

Під час аналізу предметної області було визначено сферу застосування інформаційних систем взагалі, систем електронної комерції, а також місце рекомендаційних систем в сучасному світі.

Для визначення алгоритму формування рекомендацій було проведено огляд методів та технологій формування рекомендацій: фільтрації на основі вмісту, спільна фільтрація, системи засновані на знаннях, а також системи засновані на популярності. В результаті дослідження було визначено, що у кожного з представлених методів є свої позитивні та негативні сторони, і для побудови алгоритму, що максимально відповідає досліджуваній предметній області є доцільним створити гібридний алгоритм, який об'єднає позитивні якості усіх розглянутих методів. Розроблений алгоритм опирається на особливості предметної області: велика «текучість» оголошень, активність користувача в системі буде тимчасовою, не велика кількість явного зворотного зв'язку, не постійні профілі користувачів.

За допомогою діаграми IDEF0, а також діаграми використання було описано основні функціональні вимоги до системи, а також взаємини і залежності між групами варіантів використання і дійовими особами. Також розроблено системні вимоги та вимоги до інтерфейсу користувача. Підбір параметрів алгоритму відбувався шляхом аналізу законів функціонування предметної галузі, а також даних експертів (ріелторів, користувачів системи). Експертні дані від користувачів системи зібрані шляхом соціального опитування проведеного в інтернеті. На основі розроблених вимог, а також підібраних параметрів алгоритму розроблено інформаційну систему.

Розроблений алгоритм було оцінено за допомогою порівняння результатів тестових експериментів, в результаті яких було отримано суб'єктивні оцінки користувачів, а також оцінку RMSE на досить хорошому рівні. Порівняння

отриманих результатів підтвердило, що застосування саме гібридного підходу дає найбільш точний результат в обраній предметній галузі.

Виконана робота відповідає поставленій задачі, а всі вимоги, які мали бути висунуті до першої версії продукту виконані у повному обсязі.

За темою кваліфікаційної роботи опубліковано тези до Міжнародної науково-практичної конференції «Наука, освіта та суспільство в XXI столітті: наукові ідеї та механізми реалізації» (14 грудня 2021 року в м. Полтава, Україна) [39].

## СПИСОК ЛІТЕРАТУРИ

1. Федоровський, О.М., Логачова, В.К. Архітектура рекомендаційної системи, що працює на основі неявних оцінок користувача - RCDL'2011, Воронеж, Росія, 2011
2. Konstan J.A., Riedl J. Recommender systems: from algorithms to user experience *User Model User-Adapt Interact*, 22 (2012), pp. 101-123
3. Pu P, Chen L, Hu R. A user-centric evaluation framework for recommender systems. In: *Proceedings of the fifth ACM conference on Recommender Systems (RecSys'11)*, ACM, New York, NY, USA; 2011. p. 57–164.
4. Кучеренко В. Р., Заєць М. А., Захарченко О. В., Сментина Н. В., Улибіна В. О., Оцінка та управління нерухомістю: навчальний посібник – Одеса: Видавництво ТОВ «Лерадрук», 2013. – 272 с.
5. Введення у рекомендаційні системи. Огляд основних алгоритмів рекомендацій. URL - <https://towardsdatascience.com/introduction-to-recommender-systems-6c6b6cf15ada> (дата звернення: 15.09.2021)
6. Linden G., Smith B. and York J., “Amazon.com recommendations: item-to-item collaborative filtering” *IEEE Internet Computing*, vol. 7, no. 1, pp. 76–80, 2003.
7. Плєскач В.Л., Затонацька Т.Г. Інформаційні системи і технології на підприємствах: підручник - К., 2011. -718 с., тв. пал., (ст. 10 пр.).
8. Плахотнікова, М. А. Інформаційні технології в менеджменті: підручник і практикум для прикладного бакалаврату - 2-ге вид. - Москва: Видавництво Юрайт, 2015. - 326 с.
9. Береза А.М., Козак І.А., Шевченко Ф.А. Електронна комерція: Навч. посібник– К.: КНЕУ, 2002. – 326 с
10. Пошук нерухомості по всьому світу URL: <https://www.airbnb.com.ua/> (дата звернення: 05.03.2021).
11. Портал нерухомості – інформаційно-аналітичний ресурс URL: <https://www.booking.com/index.uk.html> (дата звернення: 05.03.2021).

12. Недухомість України URL: <http://www.dom.ria.com> (дата звернення: 05.03.2021).
13. Сервіс об'яв в Україні URL: <https://www.olx.ua/> (дата звернення: 05.03.2021).
14. Adomavicius, G.; Tuzhilin, A.. Toward the Next Generation of Recommender Systems: A survey of the state-of-the-art and possible extensions, *IEEE Transactions on Knowledge and Data Engineering* 17(6): 2005 734–749.
15. Burke, R. D. Hybrid Recommender Systems, Survey and Experiment, *User Modeling and User-Adapted Interaction* 12(4): 2002 331–370.
16. Herlocker, J. L.; Konstan, J. A.; Terveen, L. G.; Riedl, J.. Evaluating Collaborative Filtering Recommender Systems, *ACM Transactions on Information Systems (TOIS)* 22(1): 2004 5–53.
17. Real Estate. URL: <https://www.zillow.com/research/hidden-costs-selling-home-19105/> / (дата звернення: 05.05.2021).
18. Park D H, Kim H K, Choi I Y, et al. A literature review and classification of recommender systems research[J]. *Expert Systems with Applications*, 39(11): 10059–10072, 2012
19. Jia, B.; Huang, X.; Jiao, S. Application of semantic similarity calculation based on knowledge graph for personalized study recommendation service. *Kuram Uygulamada Egit. Bilim.* 2018, 18, 2958–2966.
20. Manouselis, N.; Drachsler, H.; Vuorikari, R.; Hummel, H.; Koper, R., *Recommender Systems in Technology Enhanced Learning*. In *Recommender Systems Handbook*; Ricci, F., Rokach, L., Shapira, B., Kantor, P.B., Eds.; Springer: Boston, MA, USA, 2011; pp. 387–415.
21. Zhu, B.; Hurtado, R.; Bobadilla, J.; Ortega, F. An efficient recommender system method based on the numerical relevances and the non-numerical structures of the ratings. *IEEE Access* 2018, 6, 49935–49954.
22. Wang, H.; Wang, Z.; Hu, S.; Xu, X.; Chen, S.; Tu, Z. DUSKG: A fine-grained knowledge graph for effective personalized service recommendation. *Future Gener. Comput. Syst.* 2019, 100, 600–617.

23. Valdiviezo-Diaz, P.; Ortega, F.; Cobos, E.; Lara-Cabrera, R. A Collaborative Filtering Approach Based on Naïve Bayes Classifier. *IEEE Access* 2019, 7, 108581–108592.
24. Wasid, M.; Ali, R. An improved recommender system based on multi-criteria clustering approach. *Procedia Comput. Sci.* 2018, 131, 93–101.
25. Bobadilla, J.; Ortega, F.; Hernando, A.; Gutiérrez, A. Recommender systems survey. *Knowl. Based Syst.* 2013, 46, 109–132.
26. Fayyaz, Z.; Ebrahimian, M.; Nawara, D.; Ibrahim, A.; Kashef, R. Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities. *Appl. Sci.* 2020, 10, 7748.
27. Lu, J.; Wu, D.; Mao, M.; Wang, W.; Zhang, G. Recommender system application developments: A survey. *Decis. Support Syst.* 2015, 74, 12–32. [CrossRef]
28. Colombo-Mendoza, L.O.; Valencia-García, R.; Rodríguez-González, A.; Alor-Hernández, G.; Samper-Zapater, J.J. RecomMetz: A context-aware knowledge-based mobile recommender system for movie showtimes. *Expert Syst. Appl.* 2015, 42, 1202–1222.
29. Singh, M.; Rishi, O. Event driven recommendation system for E-commerce using knowledge based collaborative filtering technique. *Scalable Comput.* 2020, 21, 369–378.
30. Root Mean Square Error (RMSE). URL: <https://c3.ai/glossary/data-science/root-mean-square-error-rmse/> (дата звернення: 03.10.2021).
31. Гагаріна, Л. Г. Розробка і експлуатація автоматизованих інформаційних систем: навч. посібник для студ. / Л. Г. Гагаріної. - М.: Форум, 2010. - 384 с.
32. Маклаков С.В. ВРwin и Erwin : CASE-средства разработки информационных систем – М.: Диалог-МИФИ, 1999. – 256 с.
33. Проектування інформаційних систем з СА ERwin Modeling Suite 7.3: навчальний посібник / В. І. Горбаченко, Г. Ф. Убієнних, Г. В. Бобришева - Пенза: Вид-во ПГУ, 2012. - 154 с.

34. Оцінка та управління нерухомістю: навчальний посібник / [В. Р. Кучеренко, М. А. Заєць, О. В. Захарченко, Н. В. Сментина, В. О. Улибіна]. – Одеса: Видавництво ТОВ «Лерадрук», 2013. – 272 с.

35. Мацяшек Л. Аналіз вимог і проектування систем. Розробка інформаційних систем з використанням UML / Пер. з англ. - М.: Видавничий дім «Вільямс», 2002. - 432 с.

36. Джеф Раскін Інтерфейс. Нові напрямки в проектуванні комп'ютерних систем Видавництво «Символ-Плюс», 2007. – 272 с.

37. Ngaffo, A.; Ayeb,W.; Choukair, Z. A Bayesian Inference Based Hybrid Recommender System. IEEE Access 2020, 8, 101682–101701.

38. Grebennik I., Khriapkin O., Ovezgeldyyev A., Pisklakova V., Urniaieva I. (2019) The Concept of a Regional Information-Analytical System for Emergency Situations. In: Murayama Y., Veleev D., Zlateva P. (eds) Information Technology in Disaster Risk Reduction. ITDRR 2017. IFIP Advances in Information and Communication Technology, vol 516. Springer, Cham Scopus