

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Системотехніки
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження та адаптація методів побудови рекомендацій для нових користувачів при проектуванні e-commerce - систем
(тема)

Виконав:

студент 2 курсу, групи ІТІМ-22-1

Удовідченко Я.О.

(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки

(код і повна назва спеціальності)

Тип програми освітньо-професійна

Освітня програма Інформаційні

технології проектування

(повна назва освітньої програми)

Керівник доц. Панкратов О.В.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри

(підпис)

Гребеннік І.В.

(прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

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

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

« ____ » _____ 20 ____ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові _____ Удовідченко Яну Олександрович _____
(прізвище, ім'я, по батькові)

1. Тема роботи: Дослідження та адаптація методів побудови рекомендацій для нових користувачів при проектуванні e-commerce – систем

затверджена наказом по університету від 20.11 2023 р. № 1373Ст

2. Термін подання студентом роботи до екзаменаційної комісії: 18.01.2024 р

3. Вихідні дані до роботи: розробити гібридний метод побудови рекомендацій, що забезпечить формування рекомендацій користувачеві в системі електронної комерції у випадку неповноти або відсутності даних про вибір користувача.

4. Перелік питань, що потрібно опрацювати в роботі: Вступ. Дослідження властивостей систем e-commerce. Порівняльний аналіз платформ e-commerce. Аналіз структури системи e-commerce. Дослідження процесу побудови рекомендацій. Розробка гібридного методу побудови рекомендацій. Проведення експерименту. Висновки.

Кваліфікаційна робота не містить відомостей заборонених до відкритого опублікування.

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

Попередній захист проведено.

Керівник кваліфікаційної роботи

доц . Панкратов О.В.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій: результати порівняльного аналізу платформ e-commerce; узагальнена структура системи e-commerce; функції підсистем back office системи електронної комерції; процес продажів з використанням рекомендацій в системах e-commerce; методи побудови рекомендацій; альтернативні підходи до оцінки схожості при побудові рекомендацій; проблеми побудови рекомендацій, що потребують використання гібридних підходів; стратегії гібридизації підходів до побудови рекомендацій; гібридні підходи до побудови рекомендацій; гібридний метод побудови рекомендацій; реалізація гібридного методу; результати експериментальної перевірки.

6. Консультанти розділів роботи

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підп	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на виконання роботи	16.10.2023	виконано
2	Дослідження властивостей систем e-commerce	17-23.11.2023	виконано
3	Порівняльний аналіз платформ e-commerce	24-31.11.2023	виконано
4	Аналіз структури системи e-commerce	01-17.11.2023	виконано
5	Дослідження процесу побудови рекомендацій	18-30.11.2023	виконано
6	Розробка гібридного методу побудови рекомендацій	01-15.12.2023	виконано
7	Проведення експерименту	10-15.12.2023	виконано
8	Збор та аналіз даних експерименту	15-20.12.2023	виконано
9	Оформлення пояснювальної записки	20-30.12.2023	виконано
10	Представлення на рецензування	03.01.2024	виконано

Дата видачі завдання 16.10.2023 р.

Студент _____
(підпис)

Керівник роботи _____
(підпис)

доц. Панкратов О.В.
(посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи магістра містить: 82 с., 11 табл., 14 рис., 2 додатки, 37 джерел інформації.

ІНФОРМАЦІЙНА СИСТЕМА, РЕКОМЕНДАЦІЙНА СИСТЕМА,
ЕЛЕКТРОННА КОМЕРЦІЯ, КОЛАБОРАТИВНА ФІЛЬТРАЦІЯ, ГІБРИДНА
СТРАТЕГІЯ

Об'єктом дослідження в рамках магістерської атестаційної роботи є процес побудови рекомендацій для нових користувачів систем e-commerce.

Предметом дослідження являються гібридні методи побудови рекомендацій для нових користувачів.

Метою даної роботи є дослідження методів гібридних методів побудови рекомендацій у процесах продажів в системі e-commerce.

Актуальність роботи пов'язана з тим, що існуючі гібридні методи колаборативної фільтрації враховують окремо характеристики споживача або характеристики предметної області. Проте на практиці необхідно формувати рекомендації з врахуванням змін інтересів користувачів різних демографічних груп.

В роботі удосконалено гібридний метод побудови рекомендацій, що використовує каскадну стратегію, шляхом додаткового відбору актуальних даних про продажі.

Сфера застосування методу – рекомендаційна підсистема системи e-commerce.

ABSTRACT

Explanatory note to the qualification work of the masters contains: 82 pages, 11 tables, 14 figures, 2 appendices, 37 sources of information.

INFORMATION SYSTEM, RECOMMENDER SYSTEM, E-COMMERCE, COLLABORATIVE FILTERING, HYBRID STRATEGY

The object of research within the master's thesis is the process of building recommendations for new users of e-commerce systems.

The subject of research is hybrid methods of building recommendations for new users.

The purpose of this work is to study the methods of hybrid methods of building recommendations in sales processes in the e-commerce system.

The relevance of the work is related to the fact that existing hybrid methods of collaborative filtering take into account the characteristics of the consumer or the characteristics of the subject area separately. However, in practice, it is necessary to formulate recommendations taking into account the changing interests of users of different demographic groups.

This paper improves a hybrid recommendation method that uses a cascade strategy by additional selection of relevant sales data.

The scope of the method is the recommendation subsystem of an e-commerce system.

ЗМІСТ

Перелік умовних позначень.....	9
Вступ.....	10
1 Дослідження методів побудови рекомендацій в системах e-commerce .	13
1.1 Дослідження властивостей систем e-commerce.....	13
1.2 Порівняльний аналіз платформ e-commerce	17
1.3 Аналіз структури системи e-commerce.....	22
1.4 Дослідження процесу побудови рекомендацій в системах e-commerce	28
2 Огляд методів та технологій, які застосовуються в предметній області	32
2.1 Опис профілю та оцінка користувача при побудові рекомендацій..	32
2.2 Методи побудови рекомендацій	35
2.3 Стратегії гібридизації підходів до побудови рекомендацій в системах e-commerce	39
2.4 Постановка задачі дослідження	43
3. Розробка гібридного методу побудови рекомендацій	45
3.1 Вибір гібридного підходу до побудови рекомендацій в системах e-commerce	45
3.2 Удосконалений гібридний метод побудови рекомендацій з урахуванням демографічних характеристик та контексту вибору споживачів.....	49
4. Проведення експериментальної перевірки гібридного методу	55
4.1 Визначення показника оцінювання результатів роботи методу	55
4.2 Реалізація гібридного методу	56
4.3 Результати експериментальної перевірки.....	61
Висновки	62
Перелік джерел посилання	64
Додаток А Графічний матеріал кваліфікаційної роботи	69
Додаток Б Відомість кваліфікаційної роботи	81

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ІС – інформаційна система;

ПЗ – програмне забезпечення.

КФ – колаборативна фільтрація;

AUC – Area under ROC curve;

CMS – Content Management System;

CRM – Customer Relationship Management;

ROC – Receiver operating characteristic

ВСТУП

У сучасному інформаційному середовищі активний розвиток технологій відкриває нові перспективи для персоналізації товарів та послуг для кінцевих споживачів, зокрема, в рамках системи e-commerce. Основними параметрами персоналізації в електронній комерції є географічне положення, історія переглядів товарів і покупок.

Аналіз цих даних дозволяє власникам компаній створювати індивідуальні пропозиції, спрямовані на максимальне задоволення потреб кожного конкретного клієнта.

Рекомендаційні підсистеми, які використовуються для персоналізації продажів у системах e-commerce, використовують аналіз профілів користувачів. Ці підсистеми надають індивідуальні рекомендації, враховуючи попередні покупки та перегляди товарів.

Ефективність рекомендаційних систем проявляється в різних сферах, таких як e-commerce, веб-пошук, в контексті формування рекомендацій для програм цифрового телебачення і т.п.

Для створення привабливого переліку товарів та послуг, призначеного для конкретного споживача, рекомендаційні системи використовують методи колаборативної фільтрації, фільтрації на основі змісту, методу на основі знань та демографічний підхід. Сучасні тенденції побудови рекомендацій також включають використання семантичного та контекстно-залежного підходів.

Колаборативна фільтрація (КФ) є одним з основних та широко застосовуваних методів фільтрації, де система рекомендує елементи, що отримали високі рейтинги від інших користувачів із подібними смаками та інтересами.

У системах, що використовують колаборативну фільтрацію, профіль користувача формується на основі оцінок, які користувачі присвоюють різним елементам. Ці рейтинги можуть бути явними, отриманими в результаті відповідей користувачів, або неявними, внаслідок аналізу їх взаємодії з

рекомендаційною системою, що використовує колаборативну фільтрацію з неявним зворотнім зв'язком.

Навіть при успішному використанні колаборативної фільтрації, існують виклики, такі як проблема побудови рекомендацій для нових користувачів рекомендаційної системи, спам-атаки, неузгодженість інтересів споживачів та масштабованість.

Проблема побудови рекомендацій для нових користувачів стосується складності надання рекомендацій користувачам, які тільки-но приєдналися до системи. Для вирішення цього використовують початкові рейтинги, але це може призводити до упередженості. Ця проблема також впливає на нові товари, та послуги які не отримують рекомендацій, доки їх не оцінили достатньо користувачів.

Спам-атаки відображаються у введенні системи рекомендацій в оману недобросовісними користувачами. Неузгодженість інтересів споживачів або проблема "сірих овець " виникає, коли думки деяких користувачів не узгоджуються з будь-якою групою. Такі користувачі не отримують значущих вигод від використання рекомендаційної системи.

Проблема масштабованості виникає через необхідність пошуку рекомендацій у всій базі даних рейтингів споживачів, що обмежує можливість персоналізації рекомендацій. Для подолання цих викликів використовують гібридні методи, які поєднують колаборативну фільтрацію з аналізом профілю користувача.

Зазначені методи, однак, не вирішують проблему побудови рекомендацій для нових користувачів через відсутність детальної інформації про них. Для вирішення цієї проблеми необхідно поєднати інформацію про користувача та його поведінку під час вибору товарів та послуг.

Це свідчить про актуальність обговорюваної проблеми в контексті підвищення ефективності рекомендацій для нових користувачів у системах e-commerce.

Об'єктом дослідження даної роботи є процес побудови рекомендацій для нових користувачів систем e-commerce.

Предметом дослідження являються гібридні методи побудови рекомендацій для нових користувачів.

Метою даної роботи є дослідження методів гібридних методів побудови рекомендацій у процесах продажів в системі e-commerce.

В рамках дослідження було розроблено удосконалений метод методів побудови рекомендацій для нових користувачів e-commerce – систем. Інформація про вподобання нових користувачів або про зміни потреб користувачів часто є недоступною для системи e-commerce. Тому в роботі запропоновано удосконалений гібридний метод, який поєднує колаборативну фільтрацію з використанням демографічних даних та з урахуванням контексту вибору користувача.

1 ДОСЛІДЖЕННЯ МЕТОДІВ ПОБУДОВИ РЕКОМЕНДАЦІЙ В СИСТЕМАХ E-COMMERCE

1.1 Дослідження властивостей систем e-commerce

Глобальна інформатизація суспільства призвела до розширення сфери дії онлайн-бізнесів, особливо в рамках системи e-commerce. Електронна комерція визначається як процес отримання прибутку за допомогою Інтернет-технологій, що передбачає вибір і замовлення товарів через комп'ютерні мережі, а також розрахунки між покупцем і постачальником за допомогою електронних документів та/або коштів платежу. Цей процес може охоплювати як приватних осіб, так і організації.

Електронна комерція поділяється на кілька основних категорій:

Business-to-Business (B2B): Включає в себе взаємодію між компаніями, де можуть використовуватися спеціальні технології, такі як Electronic Data Interchange (EDI).

Business-to-Consumer (B2C): Основна частина цього напрямку - електронна роздрібна торгівля, де товари та послуги пропонуються кінцевим споживачам через Інтернет.

Business-to-Administration (B2A): Включає в себе ділові зв'язки між комерційними структурами та адміністрацією, від місцевої влади до міжнародних організацій.

Consumer-to-Administration (C2A): Найменш розвинена категорія, але з високим потенціалом для взаємодії державних структур і споживачів, зокрема в сфері соціальних і податкових послуг.

Consumer-to-Consumer (C2C): Можливість взаємодії між споживачами для обміну комерційною інформацією та досвідом.

Приклад B2B може включати компанії, які використовують Інтернет для замовлень постачальникам, обміну рахунками та оплати. B2C, у свою чергу,

охоплює електронну роздрібну торгівлю, де товари та послуги доступні кінцевим споживачам через мережу.

Зазначені категорії відображають не лише розмаїття електронної комерції, але й підкреслюють важливість взаємодії між бізнесом, адміністрацією та споживачами в цифровому середовищі.

Зростання обсягу фінансових операцій в загальнодоступних мережах свідчить про швидкий розвиток електронної комерції. Компанії активно використовують Інтернет для перетворення ключових бізнес-процесів та взаємодії зі споживачами та адміністрацією.

Елементи електронного бізнесу вже давно присутні в сучасному бізнес-середовищі, включаючи використання Electronic Data Interchange (EDI), Electronic Fund Transfer (EFT) та Enterprise Resource Planning (ERP). Сучасні компанії, зокрема завдяки Інтернету, можуть проводити операції онлайн, що дозволяє знижувати витрати і впроваджувати нові бізнес-схеми.

Важливість безпеки в електронній комерції вимагає публічних засобів захисту інформації, які гарантують ідентифікацію, конфіденційність і цілісність повідомлень у звичайній та публічній мережах.

У підсумку електронна комерція визначається не лише різноманіттям форматів та категорій, але і постійним зростанням її впливу на сучасне бізнес-середовище.

Вимоги до оплати в системах e-commerce:

- довіра;
- достовірність;
- безпека електронні платежів.

В сфері системи e-commerce довіра визначається як ключовий аспект операцій. Кожна транзакція має бути здійснена за умови довіри до особи або організації. У звичайній комерції це досягається за допомогою документів, підписаних відповідальними особами. В e-commerce використовується цифровий механізм підпису, який забезпечує ідентифікацію повідомлення, отриманого від сторонньої організації або приватної особи.

Достовірність інформації має велике значення для ефективності операцій в системах e-commerce. Механізм Electronic Data Interchange (EDI) вирішує завдання забезпечення достовірності передачі інформації, включаючи документи, і має власні стандарти, які регулюють цей процес.

Однією з ключових проблем у системах e-commerce є забезпечення безпеки та ефективності електронних платежів. Традиційна схема включає передачу зашифрованої інформації про кредитну картку покупця відповідно до замовлення, яке формує торговець. Цей процес вимагає високого рівня довіри від користувачів і залишається вразливим перед атаками та недоброзичливцями. Важливим вдосконаленням є розробка механізмів, що гарантують інформування банку та наявність необхідної суми на рахунку покупця для забезпечення надійності платежів.

Ці аспекти не тільки визначають вимоги до оплати в системах e-commerce, але також підкреслюють важливість надійності та ефективності всіх етапів торгівлі та фінансових операцій у цифровому середовищі.

В сучасному бізнес-середовищі системи e-commerce вирізняються рядом значущих переваг, які визначають їх ефективність та конкурентоспроможність (рис. 1.1).

Впровадження систем e-commerce дозволяє значно прискорити процес отримання інформації, особливо у випадку міжнародних операцій. Це сприяє оперативності прийняття рішень та ефективному управлінню бізнес-процесами.

Завдяки системам e-commerce усувається необхідність повторного введення документів, що суттєво скорочує цикл виробництва і продажу. Подача та обробка інформації стають більш ефективними, а ймовірність помилок введення зменшується.

Використання більш доступних та ефективних засобів комунікацій в системах e-commerce призводить до суттєвого зменшення витрат, пов'язаних з обміном інформацією. Це сприяє оптимізації внутрішніх та зовнішніх комунікацій.

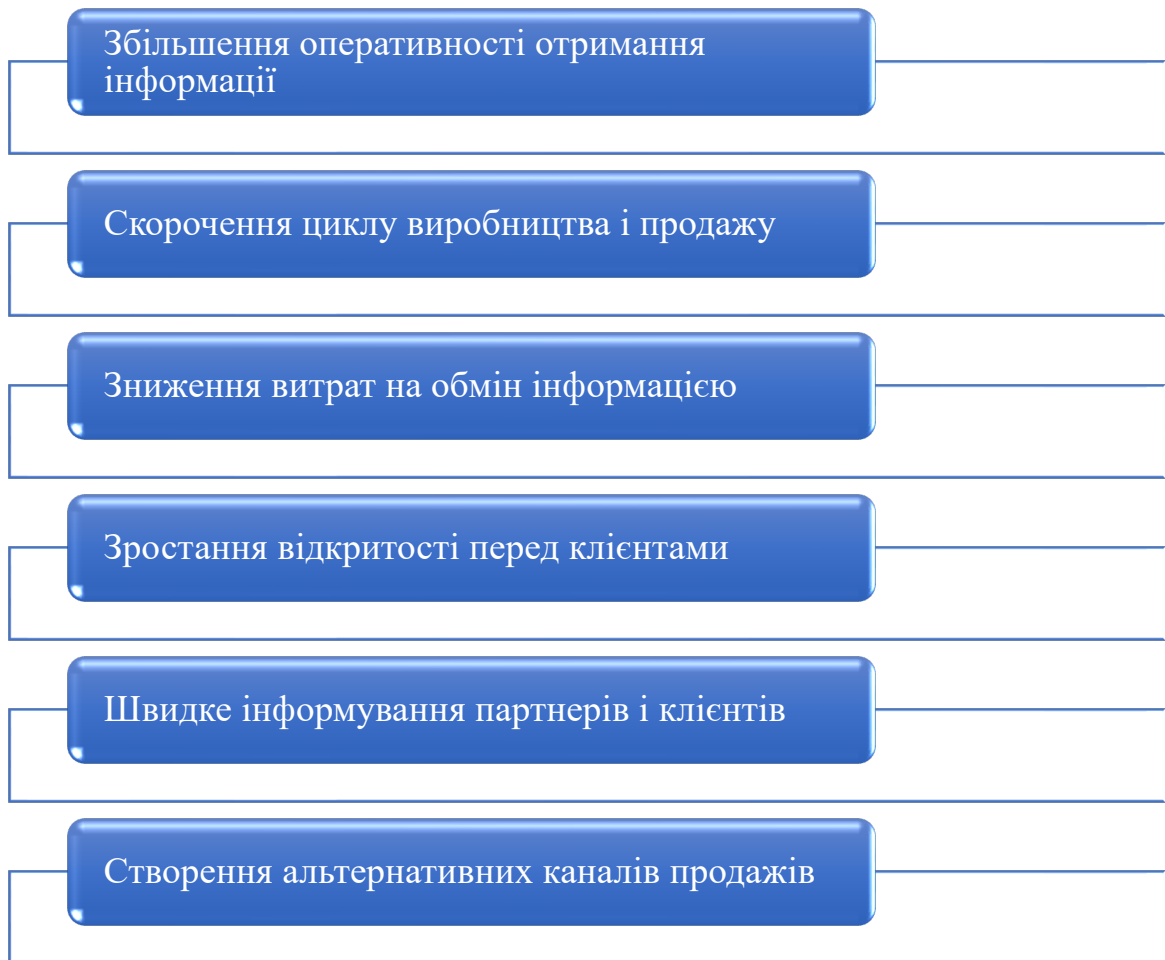


Рисунок 1.1 – Переваги систем e-commerce

Використання інтернет-технологій електронної комерції дозволяє підприємствам стати більш відкритими перед клієнтами. Забезпечення доступу до інформації про продукцію та послуги сприяє покращенню спілкування з клієнтами та підвищує рівень їхньої довіри.

Системи e-commerce дозволяють легко та оперативно інформувати партнерів і клієнтів про нові продукти та послуги. Це сприяє підтримці високої рівня взаємодії та сприяє розвитку взаємовідносин з клієнтською базою.

Використання систем e-commerce відкриває можливості для створення альтернативних каналів продажів. Наприклад, можливість ведення електронного магазину на корпоративному сайті розширює можливості компанії щодо реалізації продукції та послуг.

1.2 Порівняльний аналіз платформ e-commerce

Виконаний порівняльний аналіз платформ e-commerce з відкритим вихідним кодом OsCommerce, PrestaShop, OpenCart, VirtueMart, Magento, Zen Cart, Spree Commerce, Drupal Commerce. Ключові результати аналізу представлено в таблиці 1.1.

OsCommerce, створена Харальдом Понсе де Леоном і вперше запущена в 2000 році в Німеччині, є однією з найдавніших платформ для електронної комерції. Переваги та особливості цієї системи включають:

- OsCommerce є безкоштовним інструментом для створення сайтів електронної комерції, має відкритий вихідний код;
- багатофункціональність, з можливістю встановлення різноманітних додаткових плагінів і налаштувань, система може задовольнити різноманітні потреби користувачів;
- міжнародна підтримка користувачів, OsCommerce підтримує багатомовність і міжнародні платіжні системи;
- зручність для користувачів, для покупців, система пропонує можливість оформлення замовлення без реєстрації, можливість дивитися історію замовлень і встановлювати декілька адрес доставки.

PrestaShop є сучасною та популярною системою для електронної комерції. Основні переваги PrestaShop включають:

- PrestaShop надає безкоштовну можливість для розвитку електронних магазинів;
- розширюваність та функціональність, з великою кількістю модулів і плагінів, PrestaShop пропонує широкий спектр функцій для управління товарами, обліковими записами клієнтів та іншими аспектами бізнесу.
- можливості SEO та маркетингу: PrestaShop має потужні інструменти для SEO і маркетингу, що полегшує просування продуктів у пошукових системах та залучення клієнтів;

– підтримка клієнтів з використання панелі адміністрування сприяє ефективному взаємодії з покупцями.

Таблиця 1.1 Результати порівняльного аналізу платформ e-commerce

Платформа	Ключові переваги та особливості
OsCommerce	Багатофункціональність, з можливістю встановлення різноманітних додаткових плагінів і налаштувань; міжнародна підтримка користувачів
PrestaShop	Розширюваність та функціональність, з великою кількістю модулів і плагінів, можливості SEO та маркетингу
OpenCart	Підтримує необмежену кількість продуктів, категорій, клієнтів і угод; надає звіти по продажам, переглянутим товарам і придбаним продуктам; забезпечує SEO оптимізацію та безкоштовні оновлення
VirtueMart	Гнучке ціноутворення, дозволяє встановлювати різні ціни для різних категорій покупців; підтримує різні способи оплати; детальна статистика та аналітика
Magento	Адаптивний веб-дизайн; можливість управління кількома магазинами з одного центрального пункту управління
Zen Cart	Простота розгортання
Spree Commerce	Може вимагати від користувачів технічних навичок та зусиль при впровадженні; має обмежену кількість розширень
Drupal Commerce	Автоматизоване встановлення базового функціоналу; SEO оптимізація

OpenCart представляє собою високофункціональну систему e-commerce з відкритим вихідним кодом, наділена низкою переваг, які сприяють успішній експлуатації в електронній комерції.

Ця платформа може бути використана на будь-якому сервері, на якому вже встановлені PHP і MySQL.

Основні переваги OpenCart:

- OpenCart пропонує безкоштовне рішення для створення інтернет-магазину, дозволяючи економити витрати на програмне забезпечення.

- зручний інтерфейс OpenCart робить його легким у встановленні та управлінні;

- OpenCart підтримує необмежену кількість продуктів, категорій, клієнтів і угод; керування мовами і валютами здійснюється з одного інтерфейсу адміністратора;

- платформа надає звіти по продажам, переглянутим товарам і придбаним продуктам, що дозволяє вам збирати та аналізувати важливі дані;

- OpenCart дозволяє приймати платежі через різні методи, такі як PayPal, кредитні картки та інші.

- SEO оптимізація: зручна для пошукових систем структура сприяє ефективному просуванню магазину в Інтернеті.

- безкоштовні оновлення: Розробники OpenCart надають безкоштовні оновлення, що гарантує актуальність функціоналу.

VirtueMart — це рішення, спеціально розроблене для роботи з Joomla, і є одним з провідних безкоштовних компонентів для створення інтернет-магазинів.

Основні переваги VirtueMart:

- простота у встановленні та використанні;

- платформа дозволяє створювати необмежену кількість продуктів і категорій, забезпечуючи швидкий пошук і фільтрацію;

- гнучке ціноутворення, VirtueMart дозволяє встановлювати різні ціни для різних категорій покупців і виділяти спеціальні товари або товари на розпродаж;

- мультивалютність: встановлення різних валют для кожного продукту, з можливістю автоматичного конвертування цін;
- VirtueMart підтримує різні способи оплати, забезпечуючи зручність для покупців;
- детальна статистика та аналітика, власники магазину отримують докладну інформацію про замовлення, нових клієнтів і доходи за різні періоди;
- безкоштовні та платні додатки, оскільки розширення VirtueMart розвивається громадою розробників, і існує велика кількість безкоштовних і платних додатків для покращення функціоналу.

Magento Content Management System (CMS) на сьогоднішній день визнана однією з найпопулярніших систем управління контентом для інтернет-магазинів у світі. У 2011 році цю CMS придбала американська компанія eBay, і з того часу Magento стала вибором для таких великих компаній, як Olympus, 20th Century Fox, Time Out та інші.

Наразі на платформі Magento працює понад 250 тисяч інтернет-магазинів по всьому світу.

Magento CMS відрізняється відкритим вихідним кодом і написана на PHP з використанням Zend Framework, в якості бази даних використовує MySQL. Однією з основних переваг Magento є адаптивний веб-дизайн, що забезпечує зручний перегляд інтернет-магазину на будь-яких пристроях.

Основні особливості Magento:

- велика кількість користувачів різноманітних інтернет-магазинів в різних галузях;
- використання відкритого вихідного коду сприяє гнучкості та можливості налаштування платформи;
- адаптивний веб-дизайн, зручність перегляду на різних пристроях полегшує власникам магазинів та користувачам;
- Magento надає велику кількість вбудованих функцій, таких як робота з валютою, знижки, купони, звіти, а також розширюється через різноманітні модулі та шаблони;

–можливість управління кількома магазинами з одного центрального пункту керування;

– гнучке управління цінами: власники магазинів можуть легко знижувати ціни, встановлювати знижки, пропонувати акції для збільшення популярності і залучення клієнтів;

–система оцінок і відгуків, користувачам доступна можливість залишати відгуки та оцінки на товари;

– SEO оптимізація, Magento надає інструменти для оптимізації сторінок товарів та створення XML-карти сайту для поліпшення індексації ресурсу пошуковими системами.

Magento CMS є потужним інструментом для розробки та управління інтернет-магазинами, проте використання її варто розглядати з урахуванням особливостей бізнесу та технічних можливостей хостингу.

Zen Cart, розроблена командою Zen Cart Development, є платформою e-commerce, що відрізняється більшою адаптованістю до CSS та вищою простотою в порівнянні з OsCommerce. Ця платформа пропонує систему управління клієнтами через адміністративну панель, полегшуючи налаштування та встановлення.

Spree Commerce, заснована на Ruby On Rails, є модульною платформою e-commerce. Завдяки своїй модульності вона дозволяє легко налаштовувати та оновлювати функції електронної комерції для швидкого розгортання цифрових магазинів, адже може швидко адаптуватися до потреб користувачів.

Однак Spree Commerce може вимагати від користувачів технічних навичок та зусиль при впровадженні. Також дана система має обмежену кількість розширень .

Drupal Commerce є набором необхідних функцій для створення інтернет-магазину, вбудованим у систему управління вмістом Drupal. Drupal, з його кількома десятками модулів він надає потужну базу для будь-якого рівня складності магазину.

Має високу гнучкість та SEO-оптимізацію. Наявність пакету "Commerce Kickstart" полегшує встановлення та налаштування базового функціоналу. Використовується велика кількість додаткових модулів.

1.3 Аналіз структури системи e-commerce

E-commerce є важливим аспектом сучасного торговельного середовища, що передбачає здійснення вибору та замовлення товарів через комп'ютерні мережі, а також проведення розрахунків між покупцем і постачальником за допомогою електронних документів та/або коштів платежу.

У цьому контексті, інтернет-магазин є одним із варіантів реалізації інформаційної системи e-commerce. Узагальнену структуру системи відображено на рисунку 1.2.

Сайт інтернет-магазину виконує роль інтерфейсу для взаємодії продавця із клієнтами, тобто представляє собою Front Office. Функціонально від складається за торгової зали та особистого кабінету клієнта.

Торгова зала має такі функції:

- надає інформацію про пропозиції товарів, умови купівлі-продажу;
- містить навігацію, пошук;
- містить аналітичний інструментарій для порівняльного аналізу пропозицій;
- дозволяє складати опис умов угоди, включаючи в себе схеми взаєморозрахунків, доставку, контактні дані тощо.

Особистий кабінет клієнта має такі базові функції:

- забезпечує обмін повідомленнями між продавцем та покупцем;
- дозволяє управляти персональними даними для оформлення та виконання замовлень;
- додатково забезпечується можливість налаштування оповіщень про події в інтернет-магазині.

Back Office реалізує логіку роботи системи e-commerce і складається із таких підсистем: управління запасами (IMS), взаємодії з клієнтом (CRM), підтримки контенту (PIM), обробки запитів клієнтів (HelpDesk), бухгалтерського обліку, рекомендаційної підсистеми.

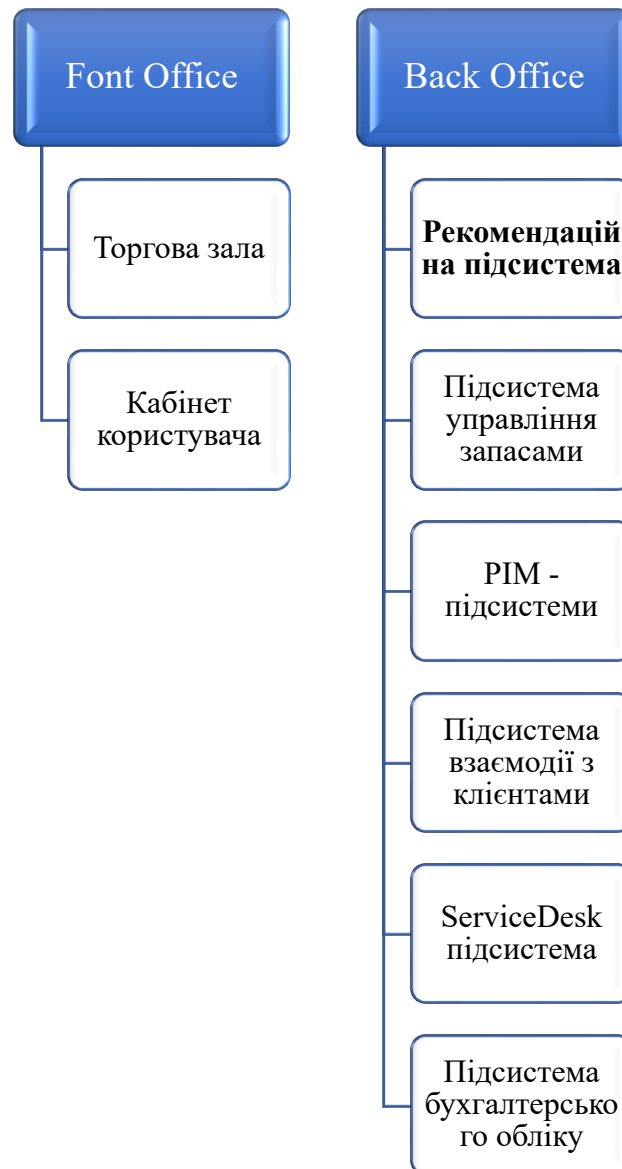


Рисунок 1.2. – Узагальнена структура системи e-commerce

Функціональність підсистем Back Office наведено в таблиці 1.1.

Підсистема управління запасами (Inventory Management System) використовується для ефективного контролю, планування та оптимізації запасів товарів або сировинних матеріалів в організації. Ця система спрямована на

полегшення процесів управління запасами, мінімізацію витрат та забезпечення ефективного використання ресурсів підприємства.

Таблиця 1.1 – Функції підсистем Back Office системи електронної комерції

Підсистема	Функції
Управління запасами	Моніторинг запасів; замовлення та постачання; прогнозування попиту; управління циклами обертання запасів
Взаємодії з клієнтом	збір та зберігання інформації про клієнтів; аналіз даних про клієнтів; автоматизація процесів взаємодії з клієнтами
Підтримки контенту, бухгалтерського обліку	зберігання, оновлення, та розповсюдження даних про товари чи послуги
Обробки запитів клієнтів	реєстрація та відстеження запитів клієнтів; обробка запитів; відстеження термінів та дотримання рівня обслуговування; формування звітності
Бухгалтерського обліку	фінансовий облік; податковий облік; управління рахунками; аналіз фінансової продуктивності

Основні функції підсистеми управління запасами включають: моніторинг запасів; замовлення та постачання; прогнозування попиту; управління циклами обертання запасів.

Дана підсистема надає можливість в режимі реального часу відстежувати рівень наявності товарів або матеріалів на складі. IMS автоматизує процеси

створення замовлень та взаємодії з постачальниками. Це допомагає уникнути надмірного чи недостатнього запасу. IMS може використовувати аналітичні інструменти для прогнозування попиту на товари чи матеріали, що полегшує планування запасів. Також підсистема допомагає визначати і оптимізувати цикли обертання запасів, зменшуючи витрати на зберігання.

Наприклад, магазині e-commerce IMS може автоматично відстежувати рівень наявності товарів на складі, інформувати про необхідність замовлення нових партій товарів від постачальників, а також оптимізувати цикли обертання запасів для максимізації ефективності. Це дозволяє уникнути ситуацій надмірного запасу, що може призвести до додаткових витрат, або недостатнього запасу, що може призвести до втрат продажів та невдоволених клієнтів.

Підсистема взаємодії з клієнтом (Customer Relationship Management) виконує збір, аналіз та управління інформацією про клієнтів компанії. Основна мета CRM - покращення відносин із клієнтами та оптимізація всіх аспектів взаємодії з ними.

Основні функції підсистеми взаємодії з клієнтом включають: збір та зберігання інформації про клієнтів; аналіз даних про клієнтів; автоматизація процесів взаємодії з клієнтами. CRM забезпечує централізоване зберігання даних про клієнтів, таких як контактна інформація, історія взаємодії, покупки та інші важливі атрибути. Підсистема надає інструменти для аналізу даних про клієнтів, що дозволяє визначити їхні потреби, попереджати та реагувати на зміни, а також генерувати звіти для прийняття стратегічних рішень. CRM автоматизує ряд бізнес-процесів, таких як обробка замовлень, ведення обліку взаєморозрахунків, планування маркетингових кампаній та інші, що покращує продуктивність та точність виконання завдань.

Наприклад, CRM може використовуватися для ведення історії покупок клієнтів, створення персоналізованих пропозицій та знижок, а також автоматизації процесів обробки замовлень. В результаті цього підприємство може забезпечити більш ефективну обслуговування клієнтів, підвищити їхню лояльність та збільшити обсяги продажів.

Підсистема підтримки контенту (Product Information Management) в системі e-commerce призначена для ефективного зберігання, оновлення, та розповсюдження даних про товари чи послуги в середовищі системи електронної комерції. Дана підсистема забезпечує централізоване зберігання та керування інформацією про продукти. Така інформація містить у собі характеристики товарів, зображення, описи, ціни, наявність, та інші важливі атрибути. PIM дозволяє уніфікувати дані, щоб забезпечити стандартні формати та структури для всіх продуктів. Підсистема забезпечує автоматизоване оновлення інформації про продукти. Наприклад, зміни в характеристиках або цінах можуть автоматично відобразитися на веб-сайті електронного магазину.

Підсистема обробки запитів клієнтів, відома також як HelpDesk, в системі електронної комерції, є інтегрованим інструментом, спрямованим на надання підтримки та вирішення проблем, що виникають у клієнтів. Основна мета HelpDesk - забезпечення ефективної комунікації між клієнтами та підприємством, вирішення їхніх питань та забезпечення задоволеності від обслуговування.

Основні функції підсистеми HelpDesk в системі електронної комерції включають: реєстрація та відстеження запитів; обробка запитів; відстеження термінів та дотримання рівня обслуговування; формування звітності.

HelpDesk дозволяє клієнтам створювати та реєструвати свої запити чи проблеми. Система веде журнал і відстежує статус та історію кожного запиту. Підсистема автоматизує процес обробки запитів, розподіляє їх між відповідальними співробітниками та надає можливість ефективно вирішувати проблеми клієнтів. Дана підсистема дозволяє взаємодіяти з клієнтами через різні канали, такі як електронна пошта, чат, телефон, а також забезпечує внутрішню комунікацію серед персоналу. HelpDesk враховує терміни вирішення проблем з клієнтами та дотримання рівня обслуговування, забезпечуючи своєчасне та якісне обслуговування клієнтів. Також підсистема надає інструменти для аналізу та формування звітів про ефективність своєї роботи, що дозволяє виявляти та вирішувати можливі покращення.

Наприклад, підсистема HelpDesk може бути використана для обробки запитів щодо замовлень, стану доставки, повернень товарів, технічної підтримки та інших питань, які виникають у клієнтів під час їхніх взаємодій з електронним магазином.

Підсистема бухгалтерського обліку реалізує автоматизацію бухгалтерських процесів та забезпечення їх відповідності стандартам бухгалтерського обліку. Основні функції підсистеми бухгалтерського обліку в системі електронній комерції включають: фінансовий облік; податковий облік; управління рахунками; аналіз фінансової продуктивності.

Дана підсистема здійснює ведення фінансового обліку, включаючи реєстрацію операцій з продажу, покупок, оплат та інших фінансових транзакцій. Підсистема також забезпечує відслідковування податкових зобов'язань та автоматизацію процесу формування фінансової звітності для податкових органів та дозволяє виписувати рахунки для клієнтів, а також здійснювати контроль за відповідними платежами. Підсистема бухгалтерського обліку містить інструменти для аналізу фінансової продуктивності, формування звітності та виявлення ключових показників ефективності.

Так, у електронному магазині підсистема бухгалтерського обліку має автоматично реєструвати всі операції з продажу та покупок товарів, автоматично формувати рахунки та відслідковувати оплати.

Рекомендаційна підсистема в системі електронної комерції використовується для надання персоналізованих рекомендацій клієнтам щодо продуктів чи послуг. Вона дає можливість підвищити рівень задоволеності клієнтів, збільшити конверсію та збут шляхом аналізу покупок та інших даних клієнта.

Основні функції рекомендаційної підсистеми в електронній комерції включають: персоналізовані рекомендації; підтримка крос-продажів та up-selling; аналіз поведінки клієнта;

Підсистема аналізує історію покупок, переглядів товарів та інші дані клієнта для створення персоналізованих рекомендацій, що відповідають його

індивідуальним інтересам та потребам. Використовуючи алгоритми встановлення схожості покупців, система може пропонувати товари або послуги, які були цікаві іншим клієнтам, подібним до даного. Система оновлює рекомендації в реальному часі, враховуючи зміни у виборі та поведінці клієнта.

Підтримка крос-продажів полягає в тому, що рекомендаційна підсистема пропонує клієнтам додаткові товари або послуги, які можуть бути цікавими на основі їхнього попереднього вибору чи покупок.

Система враховує патерни та тенденції у поведінці клієнта для прогнозування його майбутніх виборів та надання відповідних рекомендацій.

Так, в інтернет-магазині рекомендаційна система може рекомендувати клієнту товари, які доповнюють його попередні покупки (крос-продажі), або пропонувати високоякісні товари за більш високу ціну (up-selling). Наприклад, якщо клієнт придбав камеру, система може рекомендувати об'єктиви, фільтри чи штативи, щоб покращити його фотографічний досвід.

1.4 Дослідження процесу побудови рекомендацій в системах e-commerce

Побудова рекомендацій для користувачів системи e-commerce є актуальною та значущою складовою в контексті сучасних тенденцій споживання та електронної комерції. Термін "побудова рекомендацій" визначає промислову практику створення індивідуалізованих пропозицій для кожного користувача з урахуванням його потреб, попередніх покупок та взаємодії з платформою.

З інтернет-продажів стає зрозуміло, що конкуренція на цьому ринку є вкрай високою. У цьому контексті побудова рекомендацій виявляється ефективним інструментом для залучення та утримання клієнтів. В основі цього процесу лежить аналіз даних, таких як регіон проживання, перегляд товарів і історія покупок. Ці дані стають фундаментальною основою для формування індивідуалізованих пропозицій, спрямованих на задоволення унікальних потреб кожного конкретного користувача.

У світі електронної комерції побудова рекомендацій використовується для створення сприятливого середовища для покупців. Зокрема, регіональна персоналізація, що враховує місце проживання клієнта, геолокацію та інші параметри, дозволяє оптимізувати інформаційний контент для кожного регіону. Це стосується не лише представлення контактної інформації та адрес магазинів, але й тарифів на доставку.

Додатково, побудова рекомендацій базується на аналізі переглядів товарів та історії покупок. Це дозволяє системі підготувати індивідуалізований список схожих товарів, спрямованих на задоволення конкретних інтересів користувача. Наприклад, на основі інформації про купівлю книг про історію Стародавнього Риму може створюватися персоналізований список пропозицій, що включає подібну тематику

Іншими аспектами побудови рекомендацій є їх використання у заголовках, контекстній рекламі та банерах. Це спрощує навігацію для користувача та забезпечує швидкий доступ до шуканого.

Поточним трендом щодо побудови рекомендацій в електронній комерції є формування індивідуальних пропозицій для кожного користувача. Такі пропозиції надаються в рамках спеціальних пропозицій для нових відвідувачів, або для постійних клієнтів на основі їхньої історії покупок.

Поточні рішення також орієнтовані на створення індивідуальних сторінок для кожного користувача, де основний контент адаптується під його унікальні потреби. В ході цього процесу, параметри, такі як стать, вік та цінове розмежування, використовуються для персоналізації головних сторінок сайту.

На сьогодні реалізуються практики побудови рекомендацій, орієнтованих на максимальне задоволення потреб і очікувань користувачів в сучасних системах e-commerce.

Процес побудови рекомендацій можна розглядати в контексті етапів, які включають ідентифікацію, зацікавленість, залучення, продаж та утримання (рис. 1.3).

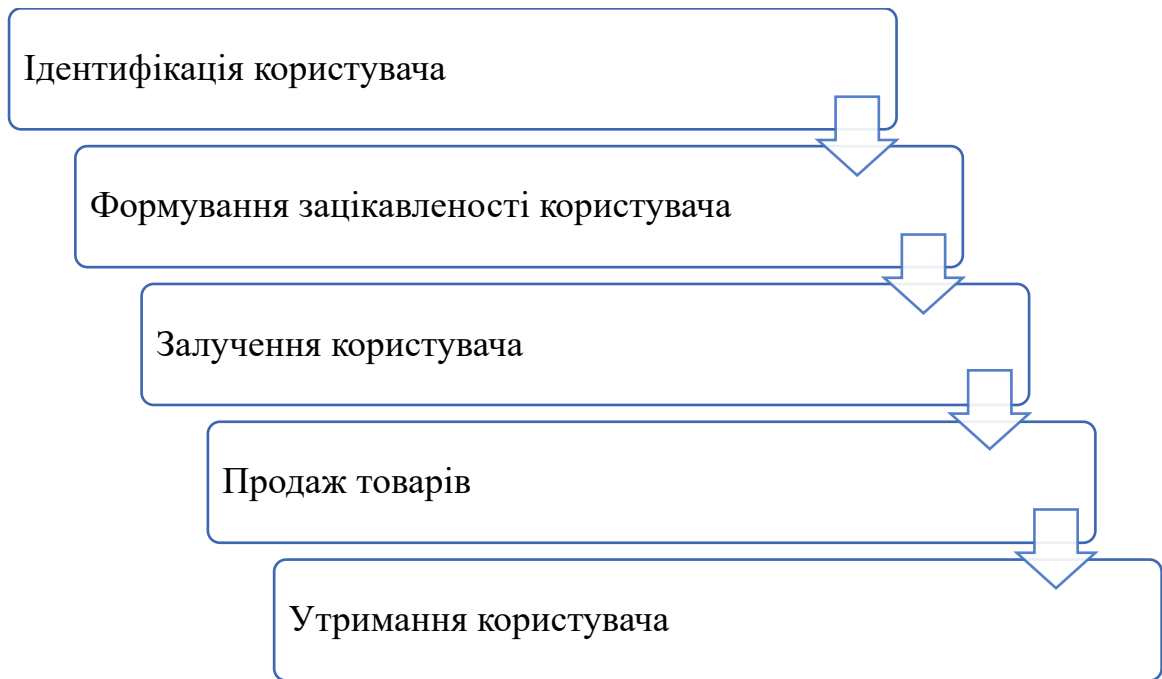


Рисунок 1.3 – Процес продажів з використанням рекомендацій в системах e-commerce

Цей логічний порядок дозволяє враховувати різні аспекти взаємодії з користувачем та максимально оптимізувати його досвід від відвідування сайту. Кожен із етапів відіграє важливу роль у покращенні користувацького досвіду та забезпеченні ефективності продажів.

Етап Ідентифікації. На першому етапі процесу побудови рекомендацій важливо провести ідентифікацію покупця. Це включає збір різноманітної інформації, такої як особисті дані (ім'я, контактні дані), місце проживання, переваги в покупках та рівень платоспроможності. Ці дані становлять основу для подальшого персоналізованого підходу до кожного користувача.

Етап Зацікавленості. Другий етап фокусується на зацікавленості користувача. Редактори сайту мають завдання створити такі умови, які утримають покупця на веб-ресурсі протягом значного часу. Це надто важливо, оскільки статистика показує, що користувач витрачає лише десять секунд на сайті, якщо не знаходить цікавого для себе контенту. Тому важливо створити такі умови, щоб вони зацікавили користувача і утримали його на сайті.

Етап Залучення. Третій етап спрямований на залучення користувача. Його мета - підняти рівень зацікавленості до рівня бажання здійснити покупку. Для досягнення цієї мети використовуються різноманітні техніки та інструменти, щоб зробити процес взаємодії максимально привабливим та простим для користувача.

Четвертий етап - це етап продажу, на якому створюються умови для максимально комфортного здійснення покупки. Забезпечення зручного та швидкого процесу узгодження продажу є ключовим завданням цього етапу.

Етап Утримання. Останній, п'ятий етап зосереджений на утриманні користувача. Це включає в себе створення всіляких стимулів, таких як знижки та акції, для того, щоб покупець бажав повернутися на сайт і здійснювати тут регулярні покупки.

Ця послідовність етапів визначає стратегію побудови рекомендацій для користувачів у системах e-commerce, спрямовану на оптимізацію користувацького досвіду та збільшення обсягу продажів.

2 ОГЛЯД МЕТОДІВ ТА ТЕХНОЛОГІЙ, ЯКІ ЗАСТОСОВУЮТЬСЯ В ПРЕДМЕТНІЙ ОБЛАСТІ

2.1 Опис профілю та оцінка користувача при побудові рекомендацій

При побудові рекомендацій в системах e-commerce профіль користувача визначається через набір оцінок, які користувач призначає різним товарам в системі електронної комерції.

Явною формою вираження цих оцінок слугує числовий рейтинг, що виставляється користувачем системи. Рейтинг представлений числом з дискретної або безперервної шкалою із заздалегідь визначеним діапазоном.

Шкала дискретного рейтингу може включати в себе, наприклад, відзначення від нуля до п'яти зірок чи використання відповідей за шкалою психолога Ренсіса Лікерта. Остання представляє собою систему з числовими значеннями, яка дозволяє визначити ступінь схвалення чи незгоди особи з конкретним твердженням. Як правило, використовується шкала з числовими значеннями від 1 до 5 або від 1 до 7, де: 1 вказує на абсолютну незгоду чи найменше ступінь позитивного відношення; 3 (або 4) може вказувати на нейтральну позицію чи невизначеність; 5 (або 7) вказує на абсолютну згоду чи найвищий ступінь позитивного відношення.

Безперервна шкала оцінки візуалізується, наприклад, у формі повзунка. Після встановлення показника користувачем, його значення конвертується в числові величини.

Додатково, існують інші форми прийняття рейтингів, такі як бінарна шкала, яка дозволяє клієнтам класифікувати елементи в дві категорії (подобається/не подобається), і порядкова шкала, де клієнтам пропонується обрати термін, що найкраще відображає їхню думку, наприклад, "повністю згоден", "згоден", "нейтрально", "не згоден", "категорично не згоден". Зазначена шкала використовується в опитуваннях для отримання детального висловлення поглядів користувача.

Також важливо відзначити унарний рейтинг, який дозволяє користувачам призначати об'єктам лише один клас, зазвичай позитивний (як у випадку кнопки "Лайк" на Facebook).

Оцінки можуть бути явними, наприклад, відзначенням "Лайк", або неявними, такими як придбання продуктів в інтернет-магазині чи переходи за посиланнями на нових сторінках, які можуть слугувати неявними унарними оцінками.

Таким чином, оцінки, що характеризують профіль користувача, можуть бути отримані явним і неявним способом. В першому випадку користувач виставляє рейтинги безпосередньо. В другому випадку фіксуються дії користувача в системі e-commerce.

Приклад матриці рейтингів як явних оцінок користувачів наведено на рисунку 2.1, а неявних – на рисунку 2.2.

	Товари				
	1	2	3	4	5
1	0	3	0	3	0
2	4	0	0	2	0
Користувачі 3	0	0	3	0	0
4	3	0	4	0	3
5	4	3	0	4	0

Рисунок 2.1 – Матриця явних оцінок користувачів

Приклад матриці неявних оцінок користувачів, які відображають факти покупок товарів цими користувачами, наведено на рис. 2.2.

	i_1	Товари				i_8
... u_1 ...	1		1			1
	1					1
Користувачі		1		1	1	
	1		1			1
... u_6 ...		1			1	

Рисунок 2.2 – Приклад матриці неявних оцінок користувачів

Таким чином, в системі e-commerce використовується експліцитний та імпліцитний (неявний) зворотний зв'язок.

Експліцитний (явний) зворотний зв'язок представляє спосіб отримання інформації щодо уподобань клієнта безпосередньо від користувача. Цей експліцитний зворотний зв'язок часто реалізується через запитання, що вимагають від користувача заповнити структурований набір питань, деталізуючи будь-які важливі уподобання.

Також, для отримання або уточнення оцінок можливо використовувати періодичні запитання щодо майбутніх відомостей, орієнтуючись на те, наскільки кожен елемент чи продукт вважається користувачем за цікавий або відповідним.

Методи підтримки неявного отримання зворотного зв'язку спрямовані на створення індивідуалізованої інформації, виводячи рейтинги користувачів із використанням індикаторів інтересів на основі їхньої взаємодії з системою. Цю інформацію можна отримати, спостерігаючи за поведінкою користувачів, коли вони переглядають, вибирають чи використовують певні елементи чи послуги.

Використання отриманих даних дозволяє системі зробити висновки щодо особливостей та реальних уподобань клієнтів на основі їхньої активності в системі електронної комерції.

2.2 Методи побудови рекомендацій

Методи побудови рекомендацій, що використовують наведені матричні вхідні дані, поділяються на дві основні категорії: підходи, орієнтовані на користувача (User-Based), та підходи, орієнтовані на елементи (Item-Based).

У підходах, орієнтованих на користувача, визначається підмножина користувачів, яка подібна до активного користувача. З використанням зваженої комбінації їх оцінок, здійснюється прогноз оцінок вподобань для активного користувача.

Підхід на основі користувача тісно пов'язаний із концепцією схожості користувачів та використовує метрики схожості, такі як коефіцієнт кореляції Пірсона та векторний простір подібності.

Підходи на основі схожості елементів (товарів, послуг) використовують ту ж концепцію, але вони спрямовані на знаходження схожих товарів для кожного окремого елемента.

Порівняння цих двох підходів представлено в таблиці 2.1.

Таблиця 2.1 – Альтернативні підходи до оцінки схожості при побудові рекомендацій

Підхід	Характеристики
Оцінка на основі даних про користувача	Схожість предметів (товарів та послуг) розраховується за схожістю користувачів, що вибирали (неявний зворотний зв'язок) або оцінили (явний зв'язок) ці предмети
Оцінка на основі властивостей товарів	Схожість користувачів встановлюється за їх оцінками предметів, що пропонуються в системі електронної комерції

При побудові рекомендацій на основі схожості користувачів або предметів (колаборативна фільтрація) виконується створення профілів та визначення подібності між користувачами чи предметів.

При вирішенні цієї задачі використовуються метрики, представлені в таблиці 2.2.

Таблиця 2.2 – Метрики оцінки подібності при побудові рекомендацій

Метрика	Характеристика
Косинусна подібність	Використовується для оцінки схожості між двома об'єктами, обчислюється кут між векторами їхніх оцінок
Кореляція Пірсона	Порівнюються рейтинги з усуненням ефектів середнього значення та дисперсії
Коефіцієнт кореляції Спірмена	Використовується для бінарних профілів, оцінює різницю в рангах елементів у профілях
Скоригована косинусна подібність	Компенсує недоліки стандартної метрики косинусу, враховуючи середню оцінку користувача
Середнє квадратичне відхилення	Використовується для оцінки подібності між користувачами, враховуючи різниці у рейтингах
Коефіцієнт Джакард	Вимірює подібність між користувачами з бінарними профілями
Умовна подібність на основі ймовірності	Використовується для топологічних рекомендацій, оцінюючи ймовірність придбання елемента в разі придбання іншого.

Після визначення подібності, алгоритм колаборативної фільтрації здійснює два основних завдання: прогнозування рейтингу та надання рекомендацій. Перше завдання полягає в прогнозуванні рейтингів для невидимих елементів цільового користувача з використанням зваженої суми

рейтингів його найближчих сусідів. Друге завдання полягає у генерації списку рекомендацій для користувача на основі його інтересів та схожості з іншими користувачами.

Для побудови рекомендацій у системах e-commerce широко використовуються представлені на рисунку 2.3. методи.

Перший метод фільтрації на основі змісту (Content-Based Filtering) базується на аналізі контенту, що дозволяє вибирати рекомендації на основі співпадінь змісту.

Використовуючи профіль користувача, який представляє його інтереси, особливо ефективно при роботі з текстовим контентом. Проте, методи виявляють обмеження у роботі з мультимедійним контентом, а також у неможливості оцінити якість товарів.

Демографічний метод (Demographic) класифікує користувачів за демографічними атрибутами для створення рекомендацій. Це важливо для нових користувачів, які ще не надали власних оцінок. Проте, виникають проблеми зі збором даних та узгодженістю інтересів, що може ускладнити точність рекомендацій.

Метод побудови рекомендацій основі знань (Knowledge-Based) пропонує об'єкти на основі зв'язку між потребами та перевагами користувача. Незалежно від обмежень історії даних, він вимагає інженерії знань, але швидко реагує на зміни у виборі користувача.

Четвертий метод базується на семантиці, використовуючи технології Semantic Web для поліпшення якості рекомендацій. Семантико-орієнтовані рекомендаційні підсистеми використовують концептуальні карти та онтології для генерації рекомендацій на основі бази знань, що покращує якість рекомендацій.

Контекстно-залежний підхід (Context-Aware) особливо корисний в динамічних середовищах, де рішення користувачів залежать від контексту. Контекстні системи можуть адаптувати рекомендації до різних ситуацій.

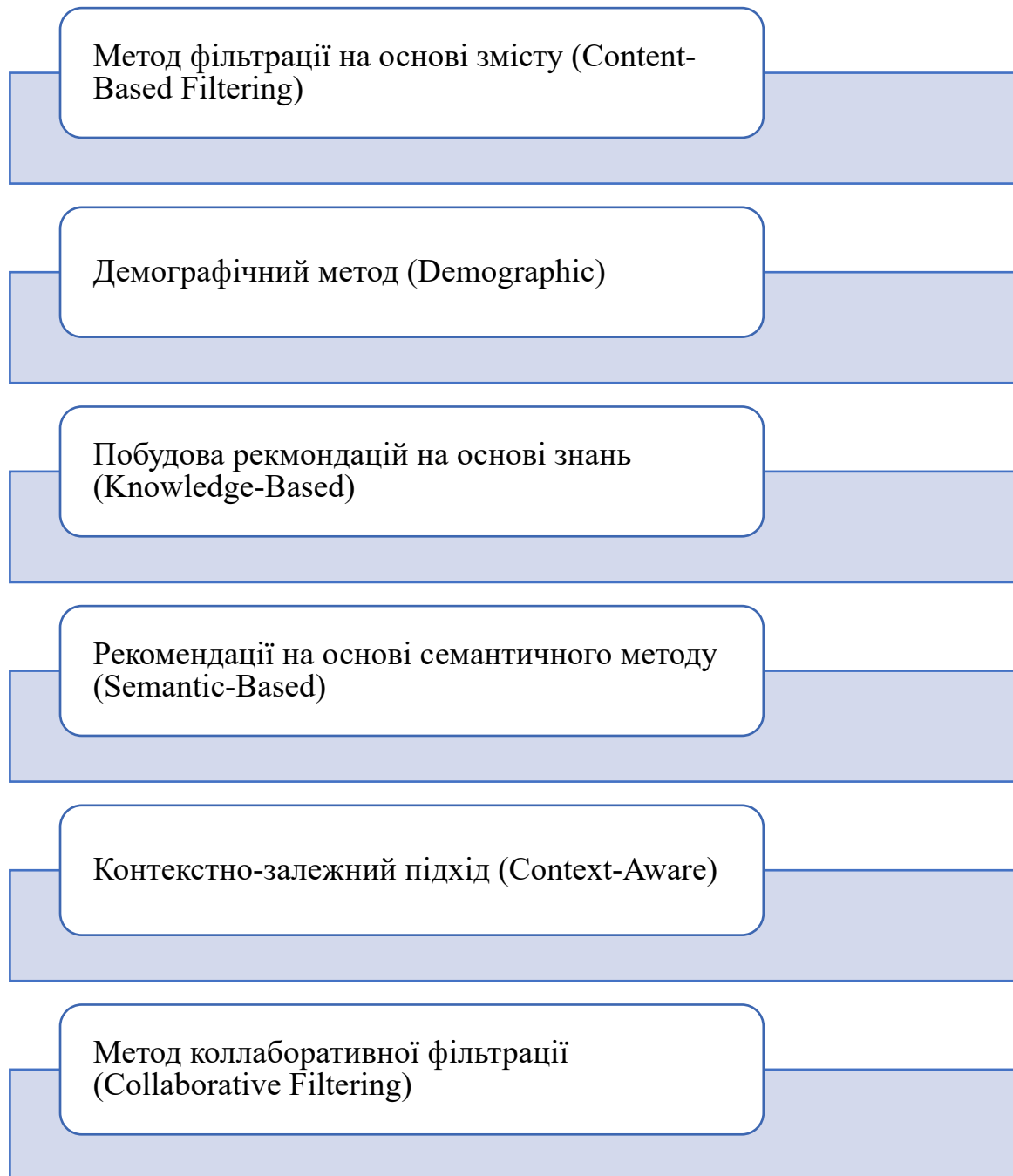


Рисунок 2.3 – Методи побудови рекомендацій

Метод колаборативної фільтрації (Collaborative Filtering) використовує інформацію про інтереси інших користувачів для рекомендацій, що дозволяє пропонувати предмети, що отримали високі рейтинги від користувачів з подібними вподобаннями.

Ці інтереси можуть бути знайдені шляхом встановлення латентних факторів схожості користувачів. Вказані фактори розраховуються шляхом

матричної факторизації. Головна ідея факторизації полягає у розділенні матриці вхідних даних (рисунок 2.1, рисунок 2.2) на дві матриці з виділеними латентними факторами для користувачів та предметів.

В подальшому схожість користувачів встановлюється за схожими латентними факторами.

2.3 Стратегії гібридизації підходів до побудови рекомендацій в системах e-commerce

Гібридні підходи використовують інтеграцію колаборативної фільтрації із аналізом додаткової інформації про користувачів або про предмети.

Узагальнений опис проблем формування рекомендацій, для вирішення яких використовуються гібридні підходи, наведено в таблиці 2.3.

Розрідженість рейтингової матриці виникає внаслідок того, що в більшості рекомендаційних систем, кожен користувач надає оцінки лише обмеженій кількості доступних елементів. Відповідно, виявлення схожості між різними користувачами чи елементами пов'язано із труднощами.

Проблема холодного старту виникає при наданні рекомендацій користувачам, які недавно зареєструвались в системі електронної комерції. У таких випадках користувачі ще не поставили достатньо оцінок, що ускладнює передбачення їхніх інтересів в рекомендаційній підсистемі. Деякі системи розв'язують цю проблему, змушуючи користувачів спочатку оцінювати певний набір елементів.

Однак такий підхід може призвести до упередженості системи. Важливо відзначити, що проблема холодного старту також впливає на нові елементи (товари, послуги), оскільки вони не потрапляють до списку рекомендацій до того моменту, поки достатньо користувачів не оцінили їх.

Проблема шилінг-атак полягає в тому, що рекомендаційні системи можуть постраждати від спам-атак, де користувачі можуть намагатися вводити систему в оману, надаючи фальшиві рейтинги або проводячи інші маніпуляції для сприяння просуванню конкретного продукту.

Таблиця 2.3 – Проблеми побудови рекомендацій, що потребують використання гібридних підходів

Проблема	Характеристика
Розрідженість вхідної рейтингової матриці	Недостатня кількість оцінок для визначення схожості користувачів
Побудова рекомендацій для нових користувачів, або холодний старт	Відсутність інформації про вподобання нових користувачів, які ще не робили вибору в системі електронної комерції
Шилінг-атаки	Навмисне спотворення рейтингів зловмисниками
Неузгодженість інтересів користувачів (проблема сірих овець)	Відсутня схожість цільового користувача з іншими користувачами
Низька масштабованість	Великий розмір матриці вхідних даних для пошуку схожості користувачів і предметів

Проблема узгодженості інтересів окремих клієнтів системи електронної комерції виникає з користувачами, чії уподобання не узгоджуються або не мають спільності з будь-якою групою клієнтів системи.

Така ситуація не дає можливість врахувати схожість користувачів і виконати колаборативну фільтрацію, оскільки отримані в результаті факторизації значення латентних факторів цільових клієнтів та інших клієнтів системи електронної комерції не будуть співпадати.

Проблема масштабованості виникає внаслідок того, що при проведенні колаборативної фільтрації аналізується вся база даних рейтингів. Отже, при зростанні кількості користувачів і елементів у базі даних витрати на обчислення подібності суттєво збільшуються.

Для вирішення наведених проблем використовується концепція гібридизації при побудові рекомендації. Гібридні рекомендаційні системи здійснюють об'єднання двох чи більше підходів до рекомендацій з метою максимізації переваг кожного окремого методу.

Ключові стратегії гібридизації можуть бути описані наступним чином:

Перша стратегія полягає в об'єднанні балів, що були отримані від кожного рекомендаційного підходу. Кожному з підходів надається вага при створенні фінальної узагальненої рекомендації.

Друга стратегія полягає у перемиканні між рекомендаційними підходами залежно від конкретної ситуації чи контексту, забезпечуючи гнучкість у виборі оптимального підходу.

Третя стратегія полягає у використанні одночасному результатів різних рекомендаційних підходів, щоб створити комплексну рекомендацію.

Четверта стратегія використовує комбінування властивостей функцій з різних підходів у єдину рекомендаційну стратегію, сприяючи оптимальному використанню їхніх переваг.

Каскадна стратегія покращує результати одного рекомендаційного підходу за допомогою іншого підходу, забезпечуючи ітеративну оптимізацію рекомендацій.

Стратегія збільшення використовується для підсилення внеску одного підходу шляхом урахування результатів іншого, забезпечуючи синергію між ними.

Метарівнева стратегія передбачає, що модель, створена за однією рекомендаційною стратегією, використовується як вхід для іншої, підкреслюючи взаємодію між різними рівнями моделювання.

Узагальнене порівняння стратегій гібридизації наведено в таблиці 2.4.

Таблиця 2.4 – Стратегії гібридизації підходів до побудови рекомендацій

Стратегія	Особливості
1. Згортка результатів декількох підходів	Кожному з підходів назначається ваговий коефіцієнт; бали додаються з урахуванням ваг
2. Перемикання між підходами	Перемикання відбувається в залежності від контексту побудови рекомендацій; контекст може залежати від часу подій, що використовуються для побудови рекомендацій
3. Комплексна рекомендація	Пропонуються одночасно результати декількох підходів
4. Комбінування властивостей	Комбінуються властивості рекомендаційних алгоритмів з урахуванням переваг кожного з них
5. Каскадна	Результати одного підходу подаються на вхід іншого з метою ітеративного уточнення; каскадна стратегія дає можливість відфільтрувати лише актуальні вхідні дані для кожного наступного етапу обробки, що робить результат рекомендації актуальним і персонально-орієнтованим.
6. Покращення результатів	Результати базового підходу доповнюються з урахування результатів іншого підходу
7. Метарівнева стратегія	Модель, створена за одною стратегією, використовується як вхідна для моделі, створеної за іншою стратегією

Із наведеного в таблиці 2.4 порівняння видно, що глобально стратегії розділяються за способом комбінування результатів окремих підходів на послідовні і паралельні групи.

2.4 Постановка задачі дослідження

Об'єктом дослідження в рамках магістерської атестаційної роботи є процес побудови рекомендацій для нових користувачів систем e-commerce.

Предметом дослідження являються гібридні методи побудови рекомендацій для нових користувачів.

Метою даної роботи є дослідження методів гібридних методів побудови рекомендацій у процесах продажів в системі e-commerce.

Актуальність роботи пов'язана з тим, що існуючі гібридні методи колаборативної фільтрації враховують окремо характеристики споживача або характеристики предметної області.

В той же час у зв'язку зі зміною вподобань користувачів, коли вони розглядаються як «нові користувачі», існує практична потреба у побудові рекомендацій з врахуванням змін інтересів користувачів різних демографічних груп.

Для вирішення цієї проблеми необхідно вдосконалити метод колаборативної фільтрації для персоналізації продажів з урахуванням як демографічних характеристик споживача, так особливостей його поведінки при виборі товару та послуг.

Для досягнення мети дослідження, необхідно розглянути наступні питання:

- дослідження особливостей систем e-commerce;
- дослідження методів побудови рекомендацій у системі e-commerce;

– аналіз та удосконалення методу колаборативної фільтрації з неявним зворотним зв'язком для побудови рекомендацій користувачами систем e-commerce.

3. РОЗРОБКА ГІБРИДНОГО МЕТОДУ ПОБУДОВИ РЕКОМЕНДАЦІЙ

3.1 Вибір гібридного підходу до побудови рекомендацій в системах e-commerce

Класифікацію гібридних підходів до побудови рекомендацій в системах електронної комерції з урахуванням наведених стратегій представлено в таблиці 2.5.

Синтез колаборативної фільтрації та фільтрації на основі змісту в гібридних системах рекомендацій представляє собою підхід, який орієнтований на вирішення проблем масштабованості та неузгодженості інтересів споживачів. Ряд наукових досліджень зосереджуються на комбінуванні цих двох підходів, враховуючи їхні переваги та недоліки. Ключова можливість такого гібридного підходу полягає у поєднанні рекомендацій на основі схожості інтересів користувачів на основі схожості властивостей товарів. Така комбінація дає можливість надати рекомендації за схожістю товарів у випадку неузгодженості інтересів цільового користувача, а також інших користувачів рекомендаційної підсистеми системи e-commerce. Також виникає можливість надати рекомендації, хоча б і неточні, у ситуації нових користувачів, тобто в ситуації холодного старту рекомендаційної системи.

Один з методів гібридизації, полягає в зваженому поєднанні користувальницької колаборативної фільтрації та фільтрації на основі змісту. Така система використовується, наприклад, для інтернет-газет, коли ваги обох підходів можуть динамічно змінюватися залежно від еволюції смаків користувача.

Інші приклади включають систему PTV, яка рекомендує телевізійні передачі, і систему фільтрування новин Usenet, яка використовує веб-сайти фільтрів для покращення колаборативної фільтрації.

Гібридні підходи також включають модифікації на рівні контенту та користувацьких профілів. Наприклад, контекстно-посилена колаборативна

фільтрація використовує класифікатор тексту Байєса для перетворення розрідженої матриці оцінок у повну.

Таблиця 2.5 – Гібридні підходи до побудови рекомендацій

Підхід	Проблеми, що вирішуються
Інтеграція колаборативної фільтрації та фільтрації на основі змісту	Проблеми масштабованості та неузгодженості інтересів споживачів
Інтеграція колаборативної фільтрації та демографічної фільтрації	Проблеми розрідженості та нових користувачів.
Інтеграція колаборативної фільтрації та основаних на знаннях підходів	Проблеми, пов'язані з новими користувачами/товарами, розрідженістю вхідних даних, неузгодженістю споживачів, а також проблеми масштабованості
Інтеграція колаборативної фільтрації та систем на основі семантики	У випадку нового товару рекомендаційна система може використовувати семантичну інформацію для надання обґрунтованих рекомендацій користувачам
Контекстно-залежний з використанням колаборативної фільтрації	Динамічна зміна рекомендацій в залежності від різних місць або моментів часу

Інтеграція колаборативної фільтрації та демографічної фільтрації є гібридним підходом, який дозволяє вирішити проблеми розрідженості та нових користувачів.

В гібридних системах, де використовуються методи колаборативної та демографічної фільтрації на основі користувачів, рейтингові та демографічні показники подібності між користувачами інтегруються для точного прогнозування рейтингів елементів. Наприклад, гібридний рекомендаційний алгоритм може вдосконалювати метод подібності, комбінуючи технічні характеристики демографічних рекомендацій та алгоритм колаборативної фільтрації.

Інший аналогічний підхід, який використовує демографічну інформацію для заповнення прогалів у знаннях щодо вподобань користувачів.

Інтеграція колаборативної фільтрації та основаних на знаннях підходів дозволяє вирішити проблеми, що пов'язані з новими користувачами/товарами, розрідженістю вхідних даних, неузгодженістю споживачів, а також проблеми масштабованості. Однак ця система потребує використання методів інженерії знань.

В рамках даного підходу реалізовано зважений гібридний метод, що базується на явних моделях товарів та користувачів, а також інтелектуальну систему, яка забезпечує співставлення між цими моделями. Також реалізована гібридна система використовує агента для вибору відповідної підсистеми (рекомендації на основі знань або користувальницької колаборативної фільтрації), координує операції двох підсистем та взаємодіє з користувачами в невизначених ситуаціях.

Інтеграція колаборативної фільтрації та систем на основі семантики дає можливість додатково врахувати семантичні відносини між різними об'єктами або користувачами, щоб підвищити якість рекомендацій.

Об'єднання семантичної подібності та подібності рейтингів між елементами має дві ключові переваги перед традиційною колаборативною фільтрацією.

По-перше, семантичні атрибути для товарів дозволяють формувати рекомендації на основі причин, що стосуються зацікавленості користувача в конкретному товарі. По-друге, у випадку нового товару або обмеженого обсягу

вхідних даних рекомендаційна система все ще може використовувати семантичну інформацію для надання обґрунтованих рекомендацій користувачам.

При реалізації даного підходу використовуються семантичні подібності між елементами для перетворення розрідженої матриці оцінок в повну матрицю оцінок, а потім використовується колаборативна фільтрація користувачів для надання рекомендацій.

Також використовується колаборативну фільтрацію з використанням семантичних профілів користувачів, які вивчаються за допомогою алгоритму зворотного зв'язку з репрезентативними документами.

В рамках даного напрямку використовуються онтологічні профілі користувачів для формування семантичних сусідів. Прогнози в даному випадку обчислюються як середньозважене відхилення від середнього сусіда, подібність між профілями використовується як значення ваги.

Альтернативний підхід використовує контекстну семантичну схожість предметів для збільшення існуючих користувацьких даних, а потім надає персоналізовані рекомендації з використанням колаборативної фільтрації на основі отриманих даних.

Використовуються також лінійна комбінація семантики та рейтингу подібності елементів, об'єднання семантичної та рейтингової подібності для генерації прогнозів, комбінована подібність для прогнозування відсутніх рейтингів.

Вказані алгоритми можуть виконуватись в залежності від розрідженості матриці вхідних даних.

Використання контекстно-залежного підходу виявляється надзвичайно ефективним у випадках динамічних рекомендацій, де користувальницькі вибори залежать від різноманітних аспектів їхнього оточуючого контексту.

Тут під контекстом розуміється будь-яка інформація (наприклад, географічне положення чи час), що може використовуватися для аналізу ситуації суб'єкта.

Системи, орієнтовані на контекст, можуть динамічно пристосовувати свої рекомендації в залежності від різних місць або моментів часу. Колаборативна фільтрація в подальшому враховуватиме контекст, коли запропонує користувачеві елементи, що отримали високі оцінки від інших користувачів з схожими смаками та інтересами.

3.2 Удосконалений гібридний метод побудови рекомендацій з урахуванням демографічних характеристик та контексту вибору споживачів

Удосконалений метод колаборативної фільтрації з неявним зворотнім зв'язком, представлений у даній роботі, враховує не лише результати неявного зворотного зв'язку від користувачів, їх демографічні характеристики, а також періодичні зміни інтересів клієнтів рекомендаційною системою.

Узагальнено етапи гібридного методу побудови рекомендацій представлено на рисунку 3.1.

Метод використовує каскадну стратегію гібридизації, згідно якої результати кожного попереднього етапу використовуються в якості вхідних даних на наступному етапі.

В рамках даної стратегії було вибрано контекстно-залежний підхід. Згідно даного підходу виділяються такі особливості контексту для надання рекомендацій:

– динамічний аспект: мають бути використані лише актуальні дані щодо проджів;

– статичний аспект: використання даних щодо користувачів системи електронної комерції.

Необхідність використання обох аспектів полягає в тому, що при побудові рекомендацій для користувачів, щодо яких мало або немає даних про покупки в рекомендаційній системі, доцільно враховувати схожість їх вподобань через демографічні характеристики.

З іншого боку, вподобання можуть швидко змінюватись з часом внаслідок сезонності або актуальних подій у місці продажу і тому необхідно в першу чергу враховувати актуальні продажі для побудови рекомендацій.

Тому поєднання статичної і динамічної компоненти даних дає можливість сформулювати пояснення з різним ступенем точності в залежності від наявності даних про користувача.

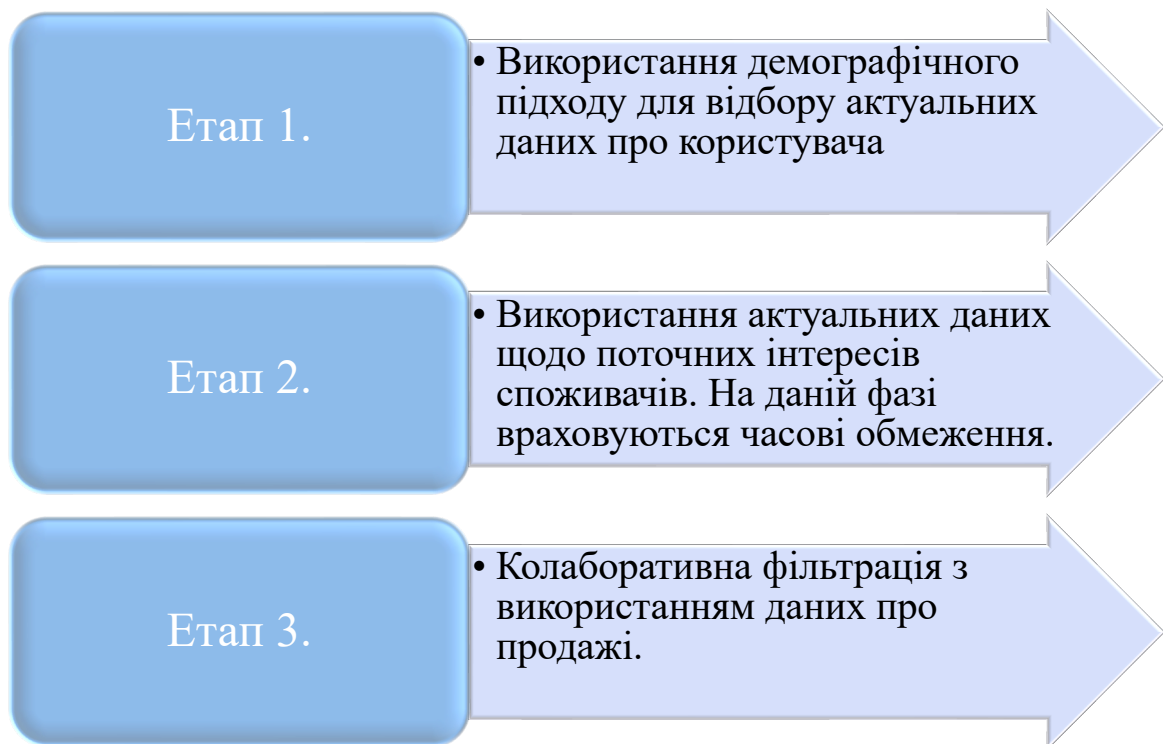


Рисунок 3.1 – Основні фази гібридного методу побудови рекомендацій

Розроблений гібридний метод побудови рекомендацій використовує такі вхідні дані про товари та споживачів.

Дані про продажі мають такий вигляд:

$$\begin{aligned}
 B &= \{b_i\}, \\
 b_i &= \{p_j, u_i, t_j\},
 \end{aligned}
 \tag{2.1}$$

де b_i – запис про продаж товару або послуги;

p_j – проданий предмет, товар, послуга;

u_l – користувач (клієнт);

t_m – часова мітка продажі (timestamp).

У виразі (2.1) представлено мінімальний набір даних, необхідних для роботи методу. Додаткові дані, наприклад щодо групи товарів дають можливість уточнити рекомендації.

Кожен клієнт рекомендаційної системи має наступні демографічні властивості:

$$u_l = \{Id_l, d_l, s_l\}, \quad (2.2)$$

де Id_l – ідентифікатор клієнта в рекомендаційній системі;

d_l – дата народження клієнта;

s_l - стать клієнта.

Проданий клієнтові товар або послуга має такі базові характеристики, як код, назва, а також ціна:

$$p_j = \{c_j, n_j, v_j\}, \quad (2.3)$$

де c_j – унікальний код предмету в системі;

n_j – назва предмету;

v_j – вартість предмету.

Додатково використовується такі дані щодо часу:

ΔT – інтервал дат народження клієнтів рекомендаційної системи;

ΔT –інтервал часу покупок для клієнтів рекомендаційної системи; даний інтервал відображає типову періодичність схожих продажів, наприклад сезонних продажів одягу.

Розроблений метод складається з послідовності етапів, що відображає наведені на рис. 2.1 фази методу.

Етап 1. Відбір вхідних даних з урахуванням дати народження клієнтів та їх статі:

$$B_1 = \{b_i : (\forall l) d_l \in \Delta T, s_l = S\}, \quad (2.4)$$

де S – цільова стать споживачів, з урахування якої формується рекомендація.

Результатом першого етапу є підмножина покупок B_1 , які є типовими для клієнтів схожого віку та статі.

Етап 2. Відбір підмножини B_2 записів про продажі із підмножини B_1 , для яких виконується умова циклічності і які характеризують цикли вибору споживачів із отриманої на першому етапі підмножини покупок людей, що є схожими за демографічними ознаками:

$$B_2 = \{b_i : b_i \in B_1, (\forall i) t_j \in \Delta T\}. \quad (2.5)$$

Результатом даного етапу є підмножини вхідної множини записів про продажі, що має наступні властивості:

– містить лише записи про продажі людям із визначеними демографічними ознаками;

– містить лише записи про продажі в межах поточного циклу, що є характерним, наприклад, для сезонних продажів.

Етап 3. Побудова рекомендацій на визначеному наборі даних методом колаборативної фільтрації.

На даному етапі використовуються дані про продажі, тобто дані неявного зворотного зв'язку. Це дає можливість більш об'єктивно сформулювати рекомендації, оскільки ці дані відображають витрати покупців.

На відміну від неявно визначених даних про продажі, явні дані про рейтинги можуть бути спотворені в результаті формування фальшивих відгуків та рейтингів предметів.

В результаті даного етапу отримуємо рекомендований перелік предметів P_1 . Цей перелік містить товари, які використовують споживачі зі схожими демографічними даними та схожими вподобаннями.

Перелік містить актуальні товари, які відповідають поточному циклу продажів або ж поточним актуальним подіям. Прикладами таких подій є: футбольний матч, концерт, тощо. Такі події змінюють вибір користувачів на невеликому часовому інтервалі, який має максимальну протяжність тижні або місяці.

Наприклад, продаються предмети з певними логотипами серед цільової аудиторії відповідного віку. Або ж змінюються модні тенденції і молодь починає купувати мобільні телефони іншого виробника.

Даний метод надає можливість формувати рекомендацію на основі схожості користувачів у випадку наявності актуальних даних про вибір цільового користувача.

В тому ж випадку, якщо дані відсутні, то метод формує рекомендацію на основі актуального вибору користувачів зі схожими демографічними ознаками.

Точність у другому випадку буде вищою.

Перевага даного методу базується на об'єднанні кількох підходів до фільтрації, що дозволяє об'єднати можливості кожного трьох методів з метою та підвищити ефективність надання рекомендацій.

Вдосконалений метод використовує демографічні характеристики користувачів і враховує циклічні зміни їхніх вподобань, що створює умови для

побудови рекомендацій у випадку "холодного старту", а також при розрідженій вхідній матриці.

Запропоновані рекомендації відповідають унікальним вподобанням та потребам користувачів у конкретний момент часу та в конкретному демографічному контексті.

Обмеження по вхідним даним на перших двох фазах методу дає можливість знизити обчислювальні витрати.

4. ПРОВЕДЕННЯ ЕКСПЕРИМЕНТАЛЬНОЇ ПЕРЕВІРКИ ГІБРИДНОГО МЕТОДУ

4.1 Визначення показника оцінювання результатів роботи методу

Для оцінки роботи розробленого методу вибраний показник AUC. Показник AUC (Area Under the Curve) є важливим метричним інструментом в області оцінки продуктивності моделей класифікації. Він визначається як площа під кривою ROC (Receiver Operating Characteristic) та AUC надає кількісну міру якості розділення класів моделлю.

Крива ROC представляє собою графік, де відображається залежність чутливості (true positive rate) від специфічності (false positive rate) при зміні порогу вирішення задачі класифікації.

Значення AUC лежить в діапазоні від 0 до 1, де значення 1 вказує на ідеальну дискримінацію між класами, тоді як значення 0.5 вказує на випадковий вибір класифікатора. Високе значення AUC свідчить про ефективність моделі у розрізненні між позитивними та негативними прикладами, незалежно від конкретного порогу класифікації.

Показник AUC є особливо корисним в ситуаціях, де набір даних незбалансований, оскільки він нечутливий до змін у розподілі класів. Його використання дозволяє отримати узагальнену оцінку ефективності класифікатора, сприяючи порівнянню різних моделей та вибору оптимального рішення в контексті конкретного завдання класифікації.

Остання властивість даного показника є ключовою для його використання з вхідними даними про продажі, представленими достатньо розрідженою матрицею. Ці вхідні дані є результатом неявного зворотного зв'язку від клієнтів системи електронної комерції. Продажі товарів розподіляються нерівномірно, що свідчить про незбалансованість вхідного набору даних.

Дослідження включає побудову рекомендацій і, відповідно, визначення трьох значень AUC для різних варіантів класифікації вхідних предметів:

– колаборативна фільтрація на основі використання даних з неявного зворотного зв'язку;

– гібридна колаборативна фільтрація, інтегрована з демографічною фільтрацією;

– гібридна колаборативна фільтрація, поєднана з демографічною фільтрацією та врахуванням актуальності даних.

Для кожного варіанту побудови рекомендацій обчислюється AUC. Високі значення AUC свідчатимуть про якісну роботу класифікатора, тоді як значення, близьке до 0,5, свідчатимуть про випадкові результати, а значення менше 0,5 вказуватимуть на обернену ефективність класифікатора.

Отримані результати дозволять зробити висновки щодо ефективності гібридного методу.

4.2 Реалізація гібридного методу

Послідовність виконання методу з урахуванням додаткових операцій зі вхідними даними містить такі кроки:

1) завантаження вхідних даних та побудова матриці даних для подальших розрахунків;

2) матрична факторизація й розрахунок латентних властивостей користувачів;

3) формування тестового переліку предметів, які можуть цікавити користувача;

4) побудова тестової рекомендації з використанням лише колаборативної фільтрації;

5) побудова рекомендації з використанням комбінації демографічної та колаборативної фільтрації;

6) побудова рекомендацій з використанням всіх трьох етапів методу;

7) розрахунок показників AUC для всіх трьох варіантів реалізації методу.

Код першого етапу експериментальної перевірки показано на рисунку 4.1.

```
import scipy.sparse as sparse
from scipy.sparse.linalg import spsolve
from sklearn.preprocessing import MinMaxScaler
#-----

raw_data = pd.read_table('data/chokostore.tsv')
raw_data = raw_data.drop(raw_data.columns[1], axis=1)
raw_data.columns = ['user', 'products', 'purchases']

data = raw_data.dropna()

# Конвертування назву продуктів у числові ID
data['user_id'] = data['user'].astype("category").cat.codes
data['product_id'] = data['product'].astype("category").cat.codes

# Створіть рамку пошуку, щоб ми могли знову знайти назву продуктів
item_lookup = data[['product_id', 'product']].drop_duplicates()
item_lookup['product_id'] = item_lookup.product_id.astype(str);

data = data.drop(['user', 'product'], axis=1);

# Видалити рядки в яких 0 покупок
data = data.loc[data.purchases != 0]

# Отримати список усіх користувачів, продуктів та продажів
users = list(np.sort(data.user_id.unique()));
products = list(np.sort(data.product_id.unique()));
purchases = list(data.purchases);

# Отримати рядки і стовпці для нашої нової матриці
rows = data.user_id.astype(int);
cols = data.product_id.astype(int);

# Створіть розріджену матрицю для наших користувачів і елементів, що містять кількість продажів
data_sparse = sparse.csr_matrix((purchases, (rows, cols)), shape=(len(users), len(products)));
```

Рисунок 4.1 – Підготовка вхідних даних

В даному фрагменті завантажуються бібліотеки, завантажуються дані про продажі та створюється вхідна матриця для роботи методу побудови рекомендацій.

Код другого етапу перевірки наведено на рис. 4.2.

На даному етапі виконується підготовка до побудови рекомендацій.

```

def implicit_als(sparse_data, alpha_val=40, iterations=10, lambda_val=0.1, features=10):
#Впровадження змінних найменших квадратів з неявними даними. Ми повторюємо
# обчислюйте вектори користувача (x_u) та item (y_i), використовуючи наступні формули:

x_u = ((Y.T*Y + Y.T*(Cu - I) * Y) + lambda*I)^-1 * (X.T * Cu * p(u))
y_i = ((X.T*X + X.T*(Ci - I) * X) + lambda*I)^-1 * (Y.T * Ci * p(i))

# Обчислити значення для кожного значення в наших даних
confidence = sparse_data * alpha_val

# Отримаємо розмір рядків користувача та стовпця елемента
user_size, item_size = sparse_data.shape

# Ми створюємо користувацькі вектори X розміру користувацьких функцій, елементів векторів
# Y розміру елементи по особі i випадково призначити значення.
X = sparse.csr_matrix(np.random.normal(size = (user_size, features)))
Y = sparse.csr_matrix(np.random.normal(size = (item_size, features)))

#Вчислення I та lambda * I
X_I = sparse.eye(user_size)
Y_I = sparse.eye(item_size)
I = sparse.eye(features)
lI = lambda_val * I

```

Рисунок 4.2 – Підготовка вхідних даних

На третьому етапі формується список предметів, що є подібними до предмету з `item_id=5`, як показано на рисунку 4.3..

```

# Знайдемо подібні елементи

item_id = 5
item_vec = item_vecs[item_id].T

# Розрахуйте оцінку подібності
# і виберіть 5 найбільш схожих.
scores = item_vecs.dot(item_vec).toarray().reshape(1,-1)[0]
top_5 = np.argsort(scores)[::-1][:5]

products = []
products_scores = []

# Get and print the actual products and scores
for idx in top_5:
    products.append(item_lookup.product.loc[item_lookup.product_id == str(idx)].iloc[0])
    products_scores.append(scores[idx])

similar = pd.DataFrame({'product': products, 'score': products_scores})

print similar

```

Рисунок 4.3 – Формування тестового списку предметів даних

На четвертому етапі створюємо рекомендації для користувача з Id=7 з використанням колаборативної фільтрації (рисунок 4.4).

```

user_id = 7
consumed_idx = data_sparse[user_id,:].nonzero()[1].astype(str)
consumed_items = item_lookup.loc[item_lookup.product_id.isin(consumed_idx)]
print consumed_items

def recommend(user_id, data_sparse, user_vecs, item_vecs, item_lookup, num_items=5):
    # Отримання всіх взаємодій користувача
    user_interactions = data_sparse[user_id,:].toarray()

    user_interactions = user_interactions.reshape(-1) + 1
    user_interactions[user_interactions > 1] = 0

    # Тут ми обчислимо рекомендацію, взявши
    # точку продукту користувацьких векторів з векторами елементів.
    rec_vector = user_vecs[user_id,:].dot(item_vecs.T).toarray()

    # Давайте оцінюємо нашу оцінку від 0 до 1, щоб все було легше інтерпретувати.
    min_max = MinMaxScaler()
    rec_vector_scaled = min_max.fit_transform(rec_vector.reshape(-1,1))[:,0]
    recommend_vector = user_interactions*rec_vector_scaled

    # Отримаємо всі показники продуктів у порядку рекомендацій (в порядку спадання) і.
    # виберемо лише верхні елементи "num_items"
    item_idx = np.argsort(recommend_vector)[::-1][:num_items]

    products = []
    scores = []

    # Перебір рекомендованих показників продуктів і отримаємо реальну назву продукта
    for idx in item_idx:
        products.append(item_lookup.product.loc[item_lookup.product_id == str(idx)].iloc[0])
        scores.append(recommend_vector[idx])

    # Створення нового dataframe із рекомендованими продуктами та оцінками
    recommendations = pd.DataFrame({'product': products, 'score': scores})

    return recommendations

# Сгенеруємо та виведемо наші рекомендації
recommendations = recommend(user_id, data_sparse, user_vecs, item_vecs, item_lookup)
print recommendations

```

Рисунок 4.4 – Формування рекомендацій для користувача з використанням колаборативної фільтрації

Наступний етап – уточнення рекомендацій з використанням демографічних даних. Додатковий код для фільтрації за демографічними даними наведено на рис. 4.5

```

for user in users:
    if user.age = users[x].age and user.gender = users[x].gender
        similar_users.append(user)

```

Рисунок 4.5 – Фільтрація користувачів за демографічними даними

Код наступного етапу – фільтрації у часовому аспекті наведено на рис.4.6.

```

now = datetime.now();
Boundary_month1=now.month-1;
Boundary_month2=now.month+1;

# Перебір покупок, які задовольняють двом значенням одразу

for purchase in recommend_product_purchase:
    if purchase.date.month>boundary_month1 and purchase.date.month<boundary_month2
        recommend_context_products.append(purchase)

```

Рисунок 4.6 – Фільтрація в часовому аспекті

```

def calc_mean_auc(training_set, altered_users, predictions, test_set):
    store_auc = [] # Пустий список для зберігання AUC для кожного користувача, якого елемент видалено з
    # навчального комплексу
    popularity_auc = [] # Щоб зберегти популярні результати AUC
    pop_items = np.array(test_set.sum(axis = 0)).reshape(-1) # Пошук загальної кількості ітерацій, щоб знайти
    # найбільш популярні
    item_vecs = прогнози [1]
    для користувача в altered_users: # Ітерація через кожного користувача, який змінив елемент
        training_row = training_set[user,:].toarray().reshape(-1) # Пошук рядка тренувань
        zero_inds = np.where(training_row == 0) # Пошук, де взаємодія ще не відбулася
        # Пошук прогнозованих значень на основі наших векторів користувача / елемента
        user_vec = прогнози [0] [користувач:]
        pred = user_vec.dot(item_vecs).toarray()[0, zero_inds].reshape(-1)
        # Пошук елементів, які спочатку були нульовими
        # Вибір усіх рейтингів з прогнозу MF для цього користувача, який спочатку не мав ітерації
        actual = test_set[user,:].toarray()[0, zero_inds].reshape(-1)
        # Виберіть binaryised так / ні пари взаємодії з оригінальними повними даними
        #, що вирівнюється з тими ж парами в навчанні
        pop = pop_items[zero_inds] # Пошук популярності елемента для наших вибраних елементів
        store_auc.append(auc_score(попередній, фактичний)) # Обчислення AUC для даного користувача і збереження
        # його
        popularity_auc.append(auc_score(pop, actual)) # Обчислення AUC з використанням найпопулярніших і оцінка
        # його
    # Ітерація кінцевих користувачів
    return float('% .3f'% np.mean(popularity_auc))

```

Рисунок 4.7 – Розрахунок AUC

Наведена послідовність етапів дає можливість оцінити переваги гібридного методу перед традиційною колаборативною фільтрацією.

4.3 Результати експериментальної перевірки

Результати перевірки методу наведено в таблиці 4.1.

Метод	AUC
КФ на основі даних про продажі	0.83
Гібридний метод з доповнення колаборативної фільтрації демографічною фільтрацією	0.874
Повний гібридний метод з попередньою демографічною та контекстною фільтрацією	0.902

Результати експериментальної перевірки показують, що доповнення колаборативної фільтрації врахуванням демографії та актуальних даних про продажі (контекстна фільтрація) дає можливість досягти 90% точності рекомендацій. Така точність свідчить про можливість практичного застосування гібридного методу в рекомендаційній підсистемі системи e-commerce.

ВИСНОВКИ

В магістерській атестаційній роботі виконано удосконалення гібридного методу колаборативної фільтрації для побудови рекомендацій клієнтам систем e-commerce.

Актуальність роботи пов'язана з необхідністю враховувати поточні зміни вподобань користувачів, пов'язані із сезонними та подієвими змінами. Проте існуючі гібридні методи колаборативної фільтрації, які є ключовими на сьогодні при побудові рекомендацій, враховують окремо характеристики споживача або ж характеристики предметної області.

В при вирішенні задачі побудови рекомендацій при зміні вподобань користувачів вони можуть розглядатись як «нові користувачі». Тому інформація про їх минулий вибір не є повністю коректними даними для побудови рекомендацій. Також важливою є і інформацію про поточні вподобання користувачів з урахуванням їх демографічних характеристик.

В роботі досліджено процес побудови рекомендацій для нових користувачів систем e-commerce. Проаналізовано гібридні стратегії, підходи та методи побудови рекомендацій.

Удосконалено гібридний метод побудови рекомендацій, що використовує каскадну стратегію, шляхом додаткового відбору актуальних даних про продажі.

Перевага даного методу базується на об'єднанні кількох підходів до фільтрації, що дозволяє об'єднати можливості кожного трьох методів з метою та підвищити ефективність надання рекомендацій.

Вдосконалений метод використовує демографічні характеристики користувачів і враховує циклічні зміни їхніх вподобань, що створює умови для побудови рекомендацій у випадку "холодного старту", а також при розрідженій вхідній матриці.

Запропоновані рекомендації відповідають унікальним вподобанням та потребам користувачів у конкретний момент часу та в конкретному демографічному контексті.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Методичні вказівки до організації виконання та захисту атестаційної роботи на здобуття першого (бакалаврського) рівня вищої освіти для студентів усіх форм навчання спеціальності 151 – «Автоматизація та комп'ютерноінтегровані технології» за освітньою програмою «Системна інженерія» / Упорядники: І.В. Гребеннік, В.Г. Іванов, Б.О. Колесник, А.С. Нечипоренко, П.Е. Ситнікова, О.С.Чорна – Харків: ХНУРЕ, 2019. – 58 с.

2. Adomavicius, G. and A. Tuzhilin, 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE T. Knowl. Data En.*, 17(6): 734-749.

3. De Campos, L. M., Fernández-Luna, J. M., Huete, J. F. and Rueda-Morales, M. A. (2010). Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks. *International Journal of Approximate Reasoning*, Vol. 51, No. 7, pp. 785–799.

4. Song, R. P., Wang, B., Huang, G. M., Liu, Q. D., Hu, R. J. and Zhang, R. S. (2014). A hybrid recommender algorithm based on an improved similarity method. *Applied Mechanics and Materials*, Vol. 475, pp. 978–982.

5. Xia, W., He, L., Gu, J. and He, K. (2009). Effective Collaborative Filtering Approaches Based on Missing Data Imputation. In *INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on*, pp. 534–537.

6. Vozalis, M. and Margaritis, K. G. (2004). Enhancing collaborative filtering with demographic data: The case of item-based filtering. In *4th International Conference on Intelligent Systems Design and Applications*, pp. 361–366.

7. Vozalis, M. and Margaritis, K. G. (2004). Enhancing collaborative filtering with demographic data: The case of item-based filtering. In *4th International Conference on Intelligent Systems Design and Applications*, pp. 361–366.

8. Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-adapted Interaction*, Vol. 12, No. 4, pp. 331–370.
9. Melville, P. and Sindhvani, V. (2010). Recommender systems. *Encyclopedia of Machine Learning*, Vol. 1, pp. 829–838.
10. Blanco-Fernández, Y., Pazos-Arias, J. J., Gil-Solla, A., Ramos-Cabrera, M., López-Nores, M., García-Duque, J., Fernández-Vilas, A., Díaz-Redondo, R. P. and Bermejo-Muñoz, J. (2008). A flexible semantic inference methodology to reason about user preferences in knowledge-based recommender systems. *Knowledge-Based Systems*, Vol. 21, No. 4, pp. 305-320.
11. Chen, T. and He, L. (2009). Collaborative filtering based on demographic attribute vector. In *Future Computer and Communication, 2009. FCC'09. International Conference on*, pp. 225–229.
12. Yapriady, B. and Uitdenbogerd, A. L. (2005). Combining demographic data with collaborative filtering for automatic music recommendation. In *Knowledge-Based Intelligent Information and Engineering Systems*, pp. 201–207.
13. Carrer-Neto, W., Hernández-Alcaraz, M. L., Valencia-García, R. and García-Sánchez, F. (2012). Social knowledge-based recommender system. Application to the movies domain. *Expert Systems with Applications*, Vol. 39, No. 12, pp. 10990–11000.
14. Antoniou, G. and Harmelen, F. V. (2004). *A semantic Web primer*. MIT Press.
15. Yang, R., Hu, W. and Qu, Y. (2013). Using Semantic Technology to Improve Recommender Systems Based on Slope One. In *Semantic Web and Web Science*, Springer, pp. 11–23.
16. Shambour, Q. and Lu, J. (2011). A hybrid multi-criteria semantic-enhanced collaborative filtering approach for personalized recommendations. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on*, Vol. 1, pp. 71–78.
17. Peis, E., Morales-del-Castillo, J. M. and Delgado-López, J. A. (2008). *Semantic Recommender Systems. Analysis of the state of the topic*. Hipertext.net.

Retrieved April 16, 2013, from <http://www.upf.edu/hipertextnet/en/numero-6/recomendacion.html>.

18. Mobasher, B., Jin, X. and Zhou, Y. (2004). Semantically enhanced collaborative filtering on the web. In *Web Mining: From Web to Semantic We*, Springer, pp. 57–76.

19. El-Dosuky, M. A., Rashad, M. Z., Hamza, T. T. and El-Bassiouny, A. H. (2012). Food Recommendation Using Ontology and Heuristics. In *Advanced Machine Learning Technologies and Applications*, Springer, pp. 423–429.

20. Daramola, O., Adigun, M. and Ayo, C. (2009). Building an Ontology-based Framework for Tourism Recommendation Services. In *Information and Communication Technologies in Tourism 2009*, Springer, pp. 135–147.

21. Blanco-Fernández, Y., Pazos-Arias, J. J., Gil-Solla, A., Ramos-Cabrer, M., Barragáns-Martínez, B., López-Nores, M., García-Duque, J., Fernández-Vilas, A. and Díaz-Redondo, R. P. (2004). AVATAR: An advanced multi-agent recommender system of personalized TV contents by semantic reasoning. In *Web Information Systems–WISE 2004*, Springer, pp. 415–421.

22. Davoodi, E., Kianmehr, K. and Afsharchi, M. (2013). A semantic social network-based expert recommender system. *Applied Intelligence*, Vol. 39, No. 1, pp. 1–13.

23. Middleton, S. E., Alani, H. and De Roure, D. C. (2002). Exploiting synergy between ontologies and recommender systems. *arXiv Preprint Cs/0204012*.

24. Oldakowski, R. and Bizer, C. (2005). SemMF: A framework for calculating semantic similarity of objects represented as RDF graphs. In *Poster at the 4th International Semantic Web Conference (ISWC 2005)*.

25. Dey, A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing*, Vol. 5, No. 1, pp. 4–7.

26. Abbar, S., Bouzeghoub, M. and Lopez, S. (2009). Context-aware recommender systems: A service-oriented approach. In *VLDB PersDB workshop*, pp. 1–6.

27. Lee, D., Park, S. E., Kahng, M., Lee, S. and Lee, S. (2010). Exploiting contextual information from event logs for personalized recommendation. In *Computer and Information Science 2010*, Springer, pp. 121–139.

28. Lee, J. S. and Lee, J. C. (2007). Context awareness by case-based reasoning in a music recommendation system. In *Ubiquitous Computing Systems*, Springer, pp. 45–58.

29. Lee, J. S. and Lee, J. C. (2006). Music for my mood: A music recommendation system based on context reasoning. In *Smart sensing and context*, Springer, pp. 190–203.

30. Ma, H., Zhou, T. C., Lyu, M. R. and King, I. (2011). Improving recommender systems by incorporating social contextual information. *ACM Transactions on Information Systems (TOIS)*, Vol. 29, No. 2, p. 9.

31. Wang, J., De Vries, A. P. and Reinders, M. J. (2006). Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 501–508.

32. Billsus, D. and M.J. Pazzani, 1998. Learning collaborative information filters. *Proceeding of the 15th International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp: 46-54.

33. Billsus, D. and M.J. Pazzani, 2000. User modeling for adaptive news access. *User Mod. User-adapted Interac.*, 10(2-3): 147-180.

34. Claypool, M., P. Le, M. Wased and D. Brown, 2001. Implicit interest indicators. *Proceeding of International Conference on Intelligent User Interfaces*, pp: 33-40.

35. Pazzani, M.J. and D. Billsus, 2007. Content-based Recommendation Systems. In: Brusilovsky, P., A. Kobsa and W. Nejdl (Eds.), *the Adaptive Web*. LNCS 4321, Springer-Verlag, Berlin Heidelberg, pp: 325-341.

36. Herlocker, J.L., J.A. Konstan, A. Borchers and J. Riedl, 1999. An algorithmic framework for performing collaborative filtering. *Proceedings of the 22nd Annual*

International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR). Berkeley, California, USA, ACM, pp: 230-237.

37. Guo, G. (2013, August). Improving the performance of recommender systems by alleviating the data sparsity and cold start problems. In Proceedings of the Twenty-Third international joint conference on Artificial Intelligence, AAAI Press, pp. 3217-3218.