

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук

Кафедра Програмної інженерії

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

другий (магістерський)
(рівень вищої освіти)

Дослідження методів побудови рекомендаційної системи
для реалізації товарів в інтернет-просторі

Виконав:

Студент 2 курсу групи ІІЗМ-21-4

Новік В. Є.

(прізвище, ініціали)

Спеціальність 121 – Інженерія

програмного забезпечення

Тип програми Освітньо-наукова

Керівник проф. Власенко Л. А.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. Кафедри

З. В. Дудар

2023

Харківський національний університет радіоелектроніки

Факультет	<u>Комп'ютерних наук</u>
Кафедра	<u>Програмної інженерії</u>
Рівень вищої освіти	<u>другий (магістерський)</u>
Спеціальність	<u>121 – Інженерія програмного забезпечення</u> (код і повна назва)
Тип програми	<u>освітньо-наукова програма</u>
Освітня програма	<u>Інженерія програмного забезпечення</u>

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

«__» _____ 2023 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

студенту _____ Новіку Вадиму Євгеновичу _____
(прізвище ім'я по батькові студента)

1. Тема роботи «Дослідження методів побудови рекомендаційної системи для реалізації товарів в інтернет просторі»

затверджена наказом університету від «10» 04 2023 р. № 340Ст _____

2. Термін подання студентом роботи до екзаменаційної комісії «18» 05 2023 р.

3. Вихідні дані до роботи методи рекомендаційної системи для реалізації товару в інтернет просторі, опис алгоритму контентної фільтрації.

4. Перелік питань, що потрібно опрацювати в роботі аналіз предметної галузі, огляд існуючих методів, архітектура програмного забезпечення, дослідження наукових матеріалів, методи побудови рекомендаційної системи для реалізації товарів в інтернет просторі.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Аналіз предметної галузі	15.02.2023	виконано
2	Виявлення проблем в сучасних рішеннях	20.02.2023	виконано
3	Постановка задачі	04.03.2023	виконано
4	Формування вимог до програмної системи	10.03.2023	виконано
5	Огляд наукової літератури	01.04.2023	виконано
6	Проектування програмного забезпечення	10.04.2023	виконано
7	Опис прийнятих програмних рішень	15.04.2023	виконано
8	Аналіз результатів дослідження	25.04.2023	виконано
9	Впровадження результатів дослідження	27.04.2023	виконано
6	Підготовка пояснювальної записки	08.05.2023	виконано
7	Підготовка презентації та доповіді	09.05.2023	виконано
8	Перевірка на плагіат	13.05.2023	виконано
9	Нормоконтроль та рецензування	14.05.2023	виконано
10	Попередній захист	15.05.2023	виконано
11	Архівування	16.05.2023	виконано
12	Допуск до захисту у зав. кафедри	17.05.2023	виконано

Дата видачі завдання _____ 1 лютого _____ 2023 р.

Студент _____
(підпис)

Керівник роботи _____ проф. Власенко Л. А.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Кваліфікаційна робота магістра містить 70 с., 15 рис., 18 джер.

РЕКОМЕНДАЦІНА СИСТЕМА, ПРОДАЖ ТОВАРІВ В ІНТЕРНЕТ,
КОСИНУСНА СХОЖІСТЬ, ASP.NET, C#, TF-IDF

Об'єктом дослідження є методи побудови рекомендаційної системи для реалізації товарів в інтернет-просторі.

Метою роботи є дослідження та аналіз методів побудови рекомендаційної системи для реалізації товарів за допомогою мережі Інтернет.

Методи дослідження, які використовувались: методи для реалізації збору даних, які стосуються комплектуючих комп'ютерів, методи пропозиції найкращих варіантів комп'ютерів для потреб користувача.

У результаті роботи були досліджені та проаналізовані методи для створення рекомендаційної системи, яка допоможе користувачам обирати найкращі варіанти комп'ютерів або комплектуючих для них, на основі власних побажань.

RECOMMENDATION SYSTEM, SELLING GOODS ON THE INTERNET,
COSINE SIMILARITY, ASP.NET, C#, TF-IDF

The object of the study is the methods of building a recommendation system for the sale of goods in the Internet space.

The purpose of the work is research and analysis of methods of building a recommendation system for the sale of goods using the Internet.

Research methods used: methods to implement data collection related to computer components, methods to suggest the best options of computers for the user's needs.

As a result of the work, methods were researched and analyzed to create a recommender system that will help users choose the best options for computers or components for them, based on their own preferences.

Я, Новік Вадим Євгенович
(прізвище, ім'я, по батькові)

студент(ка) групи ІІЗМ-21-4 здобувач вищої освіти на другому
(магістерському) рівні

кафедра програмної інженерії,
(повна назва кафедри)

заявляю: моя кваліфікаційна робота на тему

Дослідження методів побудови рекомендаційної системи для реалізації
товарів в інтернет-просторі,

що буде представлена до ЕК для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу EIArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений(а) з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

ЗМІСТ

ВСТУП.....	7
1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ.....	10
1.1 Загальний аналіз предметної галузі	10
1.2 Виявлення переваг та недоліків	13
1.3 Постановка задачі	21
2 ФОРМУВАННЯ ВИМОГ ДО ПРОГРАМНОЇ СИСТЕМИ	23
2.1 Функціональні можливості програмної системи.....	23
2.2 Нефункціональні вимоги	24
2.3 Допущення та залежності	25
3 ОГЛЯД НАУКОВОЇ ЛІТЕРАТУРИ.....	26
3.1 Загальні відомості	26
3.2 Огляд наукових публікацій.....	31
4 АРХІТЕКТУРА ТА ПРОЕКТУВАННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ	43
4.1 Проектування архітектури програмного забезпечення.....	43
4.2 Проектування UML.....	44
4.3 Проектування структури бази даних.....	47
5 ОПИС ПРИЙНЯТИХ ПРОГРАМНИХ РІШЕНЬ	48
5.1 Опис засобів розробки	48
5.2 Опис прийнятих програмних рішень	48
6 АНАЛІЗ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ	53
7 ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ	56
ВИСНОВОК	57
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	59
ДОДАТОК А Перелік джерел посилання за науковими напрямками керівника та науковців кафедри програмної інженерії	61
ДОДАТОК Б Слайди презентації.....	62
ДОДАТОК В Звіт з результатами перевірки на унікальність тексту.....	69
ДОДАТОК Г Експертний висновок результатів перевірки кваліфікаційної роботи	70

ВСТУП

На сьогоднішній день технології мають великий вплив на усе людство. Кожного дня виконуються дослідження та знаходяться нові рішення, які були би спрямованні на допомогу людям в тій чи іншій сфері діяльності, будь то рішення для створення більш безпечного навколишнього середовища для працівника, або виконання тих процесів на, які звичайній людині потрібно було би в декілька, або в сотню разів більше часу ніж автоматизованій машині. Різноманітність технологічних рішень дуже велика і може охоплювати майже всі, а може навіть і всі, побажання і потреби людини.

Більшість з усіх можливостей, які ми маємо, це досягнення за допомогою великого об'єму роботи науковців, дослідників та звичайних працівників, які хотіли поліпшити життя усього суспільства, люди змогли всього за декілька десятків років досягти неймовірних результатів в тому числі в галузі електронних обчислювальних машин, прогрес в якій, зробив переважно позитивний вплив на усі сфери діяльності людини.

Раніше комп'ютери могли досягати великих розмірів і займали майже весь простір в приміщеннях, багату кількість комплектуючих і це все для того, щоб виконувати не дуже складні алгоритми. Тепер же прогрес сягнув такої відмітки, що ми маємо комп'ютери трохи більше розмірів книжки, або журналу і які, можуть виконувати багато-мільйонні обробки запитів за дуже невеликий проміжок часу. Також з'явилося багато професій і навіть ціла галузь, яка називається Інформаційні Технології. Існує багато професій в галузі ІТ, які дуже міцно закріпилися на трудовому ринку і які дуже сильно впливають на наше життя і його розвиток. Такі професії використовують засоби обчислювальної техніки для досягнення певних результатів і виконання роботи. Для виконання роботи в таких професіях зазвичай потрібні дуже комп'ютери, які зможуть без труднощів обробляти великі обсяги даних, та виконувати процеси досить швидко. А для професій, які спеціалізуються наприклад на створенні

графіки, потрібні досить потужні пристрої, щоб без проблем обробити дуже велику кількість даних. Але для гарної роботи комп'ютера необхідно правильно обрати апаратне забезпечення або просто комплектуюче для персонального комп'ютера.

Існують фізичні магазини де вам можуть запропонувати той чи інший прилад, який би більш точніше підійшов для тих чи інших потреб. Але тут також не все ідеально, тому що, в більшості консультанти-продавці будуть пропонувати вам найдорожчі варіанти для своєї вигоди від продажу, або просто людина не буде знати тієї інформації, за допомогою якої можна було би вибрати найпродуктивніший прилад по вигідній ціні, так як вона частіше не є спеціалістом в даній галузі. Також є інтернет-магазини, які пропонують дуже великий вибір серед приладів та комплектуючих, ніж в фізичному магазині і саме вони є більш привабливими для користувача. Особливу популярність вони також набули під час пандемії, коли кожен намагався мінімізувати своє знаходження в місцях де може бути багато людей. Але і в інтернет-магазинах не так просто обрати серед безлічі різних комплектуючих та комп'ютерів той, який буде вирішувати ваші потреби. Також є проблемою те, що навіть готові зібрані комп'ютери можуть рідко бути вигідними для користувача. Наприклад, дуже часто буває коли клієнт, не хоче бачити в своєму майбутньому приладі, те чи інше комплектуюче і в той же час знає, яке йому потрібно. Але в магазині таких приладів може просто не бути і єдиний вихід це шукати маловідомі сайти компаній, магазинів, де вам допомогли вирішити це питання. Та надійність таких магазинів занадто менше і це може бути набагато дорожче.

Тому для вирішення таких проблем було прийнято рішення для створення проекту, який мав би, з одного боку звичний для клієнтів вигляд та функціонал інтернет-магазину з продажу комп'ютерів та комплектуючих, та з другого боку мав в собі методи, які використовувались для створення рекомендаційної системи для реалізації в інтернет-просторі.

Таким чином створення рекомендаційної системи допоможе користувачу підібрати найвигідніший для його потреб прилад. Планується також створення

можливості зібрання комп'ютера власноруч на сайті, підібрати всі необхідні комплектуючі, і для цього клієнту навіть не потрібно бути експертом в цій галузі. Тому що, кожен вибір буде супроводжений певними рекомендаціями, щодо сумісності комплектуючих, та їх доцільність використання для тих потреб, які вкаже клієнт перед зібранням власного комп'ютера.

Актуальність роботи зумовлена тим, що є великий попит на купівлю комп'ютерів, в тому числі переважно через мережу Інтернет. І для користувача є дуже важливим придбати пристрій, який буде відповідати саме його цілям і також є важливим заощадити гроші не переплачуючи за ті можливості, які йому не потрібні.

Метою дослідження в рамках цього проекту є, дослідження методів, які допоможуть створити рекомендаційну систему, в свою чергу, яка зможе вирішити сучасні проблеми користувачів, а саме, яка допоможе заощадити гроші, час а найважливіше допоможе підібрати пристрій, який зможе надовго вирішити його потреби.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Загальний аналіз предметної галузі

Світовий ринок в галузі комп'ютерів є дуже великим і має також великий попит серед населення усього світу [1]. Якщо дивитись на це глобально він не є стабільним і з року в рік може мати, як позитивний дохід, так і негативний. Чому так відбувається? Це зумовлено тим, що на дохід впливає велика кількість факторів, як зовнішніх так і внутрішніх. Так наприклад під час пандемії, став актуальним перехід на дистанційну роботу та навчання, для цього потрібно мати прилад з можливістю виходу в мережу Інтернет і найчастіше це буде або стаціонарний комп'ютер, або ноутбук. Тому в цей період світовий ринок комп'ютерів мав дуже великий зріст. Або є ситуації негативного доходу, за приклад можна взяти сучасну тенденцію ринку. Спадок продажів після 2021 року та нині зумовлений тим, що в час пандемії, а саме часовий відрізок між 2020 та 2021 роком, був періодом максимального зросту продажів комп'ютерів та комплектуючих. З цього і виходить актуальний і нині спадок продажів. Тому що, у користувачів немає потреби в заміні своїх пристроїв на нові, так як просто нема потреби в заміні комплектуючих на більш потужні, бо ті комп'ютери з комплектуючими, які були актуальними декілька років тому, в повній мірі можуть виконувати всі необхідні процеси з тою ж продуктивністю, що і раніше. Є ще декілька факторів, які можуть впливати на продаж, наприклад інфляція, яка на сьогоднішній день охопила багато країн світу і користувачі не хочуть витратити гроші на оновлення пристроїв, тому що вже мають відносно нові комп'ютери, які вони купували під час першого року пандемії. І також є фактором - оновлення операційної системи та програмного забезпечення пристроїв. Наприклад нині відбувається активний перехід персональних комп'ютерів на нову операційну систему Windows 11, але цей перехід можливий, якщо користувач вже має досить потужний пристрій для підтримки цієї версії. Також безліч людей будуть хотіти просто отримати поновлення для нових емоцій та можливостей.

Так чи інакше галузь комп'ютерів та їх комплектуючих була, є та буде дуже перспективною та користуватись попитом серед населення світу. Тому що, є багато сучасних професій та в цілому занять людини, які потребують наявності сучасного, досить потужного персонального комп'ютера. Основними способами продажу, як в цілому і для всіх товарів, є продаж в фізичному магазині, де людина може прийти, подивитись, по-взаємодіяти з пристроєм або звернутись до консультанта-продавця за допомогою у виборі. І також є варіант, який отримав більшу актуальність та попит серед людей, під час пандемії та перебудовування системи роботи та навчання на дистанційну, які хотіли придбати собі будь-який пристрій, це спеціалізовані інтернет-магазини. Це можуть бути, як магазини компаній, які безпосередньо виготовляють ту чи іншу модель персональних комп'ютерів, або інтернет-магазини такі, як Розетка, що виступають посередником між продавцем та покупцем. Таким чином людям, які ведуть бізнес по продажу персональних комп'ютерів та комплектуючих до них, можна без наявності фізичного магазину або інтернет-магазину, звертатись до таких посередників, які будуть розміщувати товари фізичної особи-підприємця за сплату певного проценту з продажів.

В рамках цього проекту будуть досліджуватись методи для створення рекомендаційної системи для реалізації товару безпосередньо в інтернет-просторі і вирішуватись проблеми, які мають сучасні інтернет-магазини. А тому як, існує дуже велика кількість різноманітних магазинів з продажу комп'ютерів, ми відразу відкинемо розгляд і докладний аналіз фізичних магазинів і сконцентруємо увагу на аналізі магазинів або сервісів в інтернет-просторі. Тож раніше було досить багато магазинів де майбутній власник сучасного та потужного персонального комп'ютера міг придбати його собі і вирішити низку потреб. Під час пандемії, таких магазинів та сервісів з'явилося в рази більше тому, як попит у людей зростав на покупку комп'ютерів для дистанційної роботи або навчання і важно було робити закупи також дистанційно. Тому і кожен хто хотів відкрити власний невеличкий або інший в розмірах бізнес, з радістю це робив. Але не зважаючи на ту велику кількість магазинів, які з'явилися вони все одно не змогли вирішити всі

проблеми користувачів, які могли виникати при покупці персонального комп'ютера. Звичайно вони вирішили актуальну проблему з дистанційною покупкою, шляхом створення більшої варіації магазинів з унікальним або таким собі дизайном, який міг би задовольнити будь-якого покупця. Також, якщо казати про Україну, то це розширило ринок, тому як раніше, не було такої варіативності в інтернет-магазинах і просто не було потреби у тих компаній, які мали вже фізичні магазини, вносити досить велику кількість грошей на покращення та подальший розвиток інтернет версії, тому що, на той час люди мабуть більше віддавали пріоритету на візит фізичного магазину. Тому раніше і були на слуху тільки Розетка та Хотлайн, але останній використовувався і використовується нині, як пошук певної одиниці товару в усіх магазинах країни для знаходження вигідної ціни для клієнта. Але низька ціна в невідомому магазині, не завжди може стати пріоритетом, ніж покупка хоч трохи і подороже, але в провіреному магазині, де клієнт буде впевнений, що він в досить короткі строки зможе отримати свій товар і буде впевнений за якість виконання доставки. Але основною, на мою думку, проблемою було і досі є те, що клієнт, який заходить на інтернет-сторінку магазину і хоче купити для себе персональний комп'ютер для вирішення своїх потреб, не завжди, а частіше зовсім не може зрозуміти, який персональний комп'ютер йому необхідний. Тому як, існує велика кількість готових збірок комп'ютерів, але за частіше вони є, або дуже дорогі, тому як в них вбудовані найновіші комплектуючі, або комп'ютери де використовується не збалансовані між собою комплектуючі, наприклад коли ми маємо сучасну відеокарту, але вся остання система залишає бажати кращого. І щоб уникнути ситуації в яких, користувач переплачує за той чи інший продукт, або навпаки отримує не ту версію зібраного комп'ютеру, який би міг вирішити низку потреб користувача, потрібно встановити можливість зібрати комп'ютер власноруч. Інтернет-магазини з такою можливістю є, але знову виникає питання, як клієнт зможе з безлічі комплектуючих з нуля зібрати комп'ютер, який мало того, що буде вирішувати всі його потреби, так ще хоча б просто запускався. Тому було вирішено зробити дослідження тих методів, які би використовувались в створенні

рекомендаційної системи в свою чергу, яка допомогла би клієнту на досить зрозумілому рівні порекомендувати, ту чи іншу варіацію комплектуючих, або вже зібраний варіант, який би точно зміг відповідати вказаним потребам клієнта і в свою чергу він не переплачував за те, що йому не потрібно.

1.2 Виявлення переваг та недоліків

На сьогоднішній день існує досить багато інтернет-магазинів, які спеціалізовані на продажі комп'ютерів та комплектуючих. Майже кожен з них має унікальний дизайн та має досить велику кількість можливостей на будь-який смак клієнта. Але не зважаючи на великий ринок, дійсно якісних аналогів не так багато, як для такої величезної галузі.

У якості дослідження, я проводив аналіз ринку в Україні, а також найближчого сусіда з європейського ринку – Польщу. Вибір Польщі у якості аналізу, був зумовлений насамперед тим, що торгові зв'язки між цією країною, та Україною дуже тісні і дуже часто відбувається, що саме електричні прилади поставляються саме з Польщі так як, наприклад багато торговельних марок у галузі комп'ютерів, збирають свої пристрої саме в Польщі. Також нерідко відбувається так, що клієнт замовляє персональний комп'ютер або комплектуюче саме з Польщі, так як попросту це є значно дешевше.

Тож провівши аналіз ринку обраних країн, серед великої кількості аналогів можна виділити декілька, які є досить якісними, в тій чи іншій мірі, та користуються попитом у клієнтів, а також мають можливість замовити комп'ютер з самостійно обраною конфігурацією.

Одним із них є спеціалізований на продажі комп'ютерів та комплектуючих, інтернет-магазин Telemart.ua. Цей інтернет-магазин являється одним із популярних в Україні, які спеціалізуються на продажі персональних комп'ютерів та комплектуючих, він виділяється серед більшості тим, що має в собі можливість

самостійної конфігурації комп'ютера та певні рекомендації, щодо його конфігурації. Цей інтернет-магазин має в собі всі функціональні можливості, які необхідні для стабільної продажі товару і також має досить багато вигідних пропозицій, щодо купівлі товарів, які є досить привабливими для клієнта і є сприятливими для покупки.

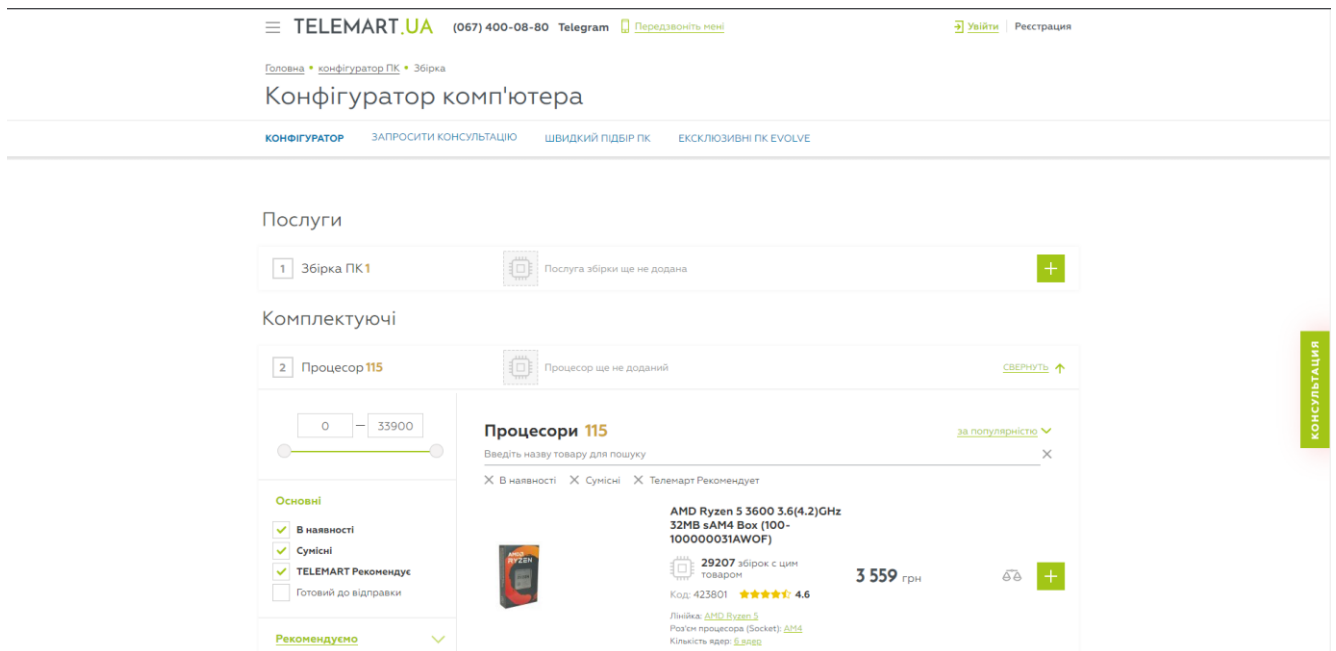


Рисунок 1 – Сторінка конфігурації комп'ютера в Telemart.ua

До переваг цього інтернет-магазину можна віднести те, що він має простий та зрозумілий дизайн. Також сучасна модель Trade-in, яка в Україні на комп'ютерному ринку є досить не звиклим явищем. За цією можливістю клієнт може обміняти свій старий комп'ютер або певні комплектуючі на бонуси вже за якими, зможе купити новенький пристрій. Авжеж одним із критеріїв для отримання обміну, це пристрої в робочому стані та які, не мають пошкоджень. І також можливість зробити власну конфігурацію ПК з самого нуля. Також, якщо клієнт хоче отримати консультацію, щодо цієї можливості, то Telemart.ua надає цю можливість на сторінці перед початком конфігурації. Ще до переваг можна віднести те, що є можливість швидкої пропозиції комп'ютера виходячи з потреб користувача. Для цього клієнту потрібно лише обрати його бюджет, як він хоче

використовувати свій майбутній комп'ютер, тобто для ігор або для роботи і також обирає пріоритетний для себе тип материнської плати та процесора. Також для популярних конфігурацій є можливість подивитись відео, як відчуває себе система під час навантажень, що очевидно є хорошим плюсом. Ще у якості переваги можна виділити те, що є в наявності можливість фільтрації комплектуючих при самостійній конфігурації комп'ютера. Тобто, якщо клієнт обирає материнську плату, яка може підтримувати тільки AMD процесори, то процесори від Intel автоматично не будуть надаватись у якості вибору.

До недоліків можна віднести те, що незважаючи на можливість продажу мобільних персональних комп'ютерів або по prostu ноутбуків, можливість швидко підібрати саме ноутбук відсутня, бо ця можливість працює тільки для стаціонарних комп'ютерів, що є мінусом, так як в сучасному світі люди віддають більше пріоритету мобільності іноді навіть, якщо персональний комп'ютер купляється з ціллю використання його в іграх та важких процесах. Також коли клієнт обирає не сумісні комплектуючі він отримує повідомлення про це, але все одно система дозволяє придбати ці комплектуючі, тобто не блокується можливість оформити покупку. Також, якщо було обрано не всі життєво необхідні комплектуючі для комп'ютера, також система не блокує доступ до покупки. Також нема можливості подивитись наскільки буде гарною продуктивність тієї конфігурації персонального комп'ютера, яку самостійно обрав клієнт. Ще на сторінці конфігурації відображається смужка її прогресу, але на її прогрес разом з обов'язковими комплектуючими, також впливають і аксесуари такі як, крісло, клавіатура, килимок і тощо.

Versum – це інтернет-магазин, який спрямований на продаж стаціонарних персональних комп'ютерів та аксесуарів до нього. Компанія орієнтована на гравців в комп'ютерні ігри, як для хобі так і для професійної діяльності в кіберспорті. Магазин виділяється можливістю розробити власну конфігурацію комп'ютера. Компанія Versum є першим масштабним підприємством в Україні, що спеціалізується на виробництві кіберспортивної екіпіровки. Головна мета компанії задовольнити потреби тих хто живе грою, або тільки знайомиться з

віртуальним світом. Це досягається шляхом аналізу ринку та дослідження сучасних технологій, а також спільною роботою інженерів та дизайнерів, щоб комп'ютер клієнта відповідав високим стандартам ігрових стаціонарних персональних комп'ютерів.

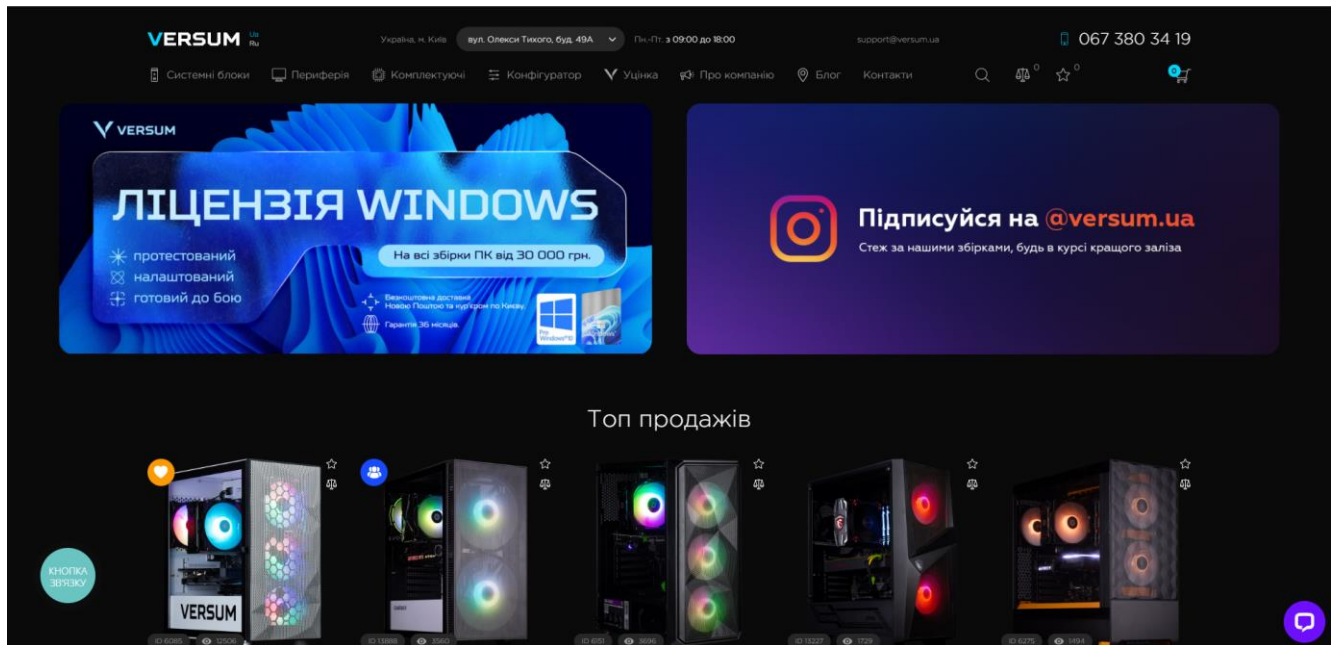


Рисунок 2 – Інтерфейс головної сторінки інтернет-магазину Versum

До переваг цього інтернет-магазину можна віднести те, що його дизайн є досить сучасним та дуже підходить для тематики комп'ютерних ігор та кіберспорту. Також є величезним плюсом те, що компанія має власний канал на платформі YouTube і також канал на відеостримінговій площадці Twitch. Наявність таких можливостей є досить привабливим для клієнта тому що, це дає певне розуміння, що компанія є соціально-активною і буде уникати способів обману своїх клієнтів. А за допомогою каналу на Twitch є можливість дивитись онлайн на збірку комп'ютера майстром. Є можливість створити власну конфігурацію комп'ютера і система блокує купівлю обраної конфігурації, якщо користувач обрав не всі, необхідні для роботи комп'ютера комплектуючі, що є досить добре і це може запобігти зробити помилку клієнтом під час самостійної конфігурації персонального комп'ютера. Також клієнту надається можливість обрати вже зібрані комп'ютери. Кожна конфігурація підтримує можливість

відобразити, яка буде у комп'ютера продуктивність в тій чи іншій грі, що дозволить клієнту обрати саме ту модель, яка буде для нього привабливішою. Ще не маловажним моментом, що є можливість у вже готовій збірці, змінити довільне комплектуюче на бажане для клієнта. Є фільтрація комплектуючих за сумісністю, тобто якщо, клієнт обирає материнську плату з підтримкою AMD, то обрати процесор від Intel він не може із за несумісності комплектуючих. Також можна виділити можливість консультації зі співробітниками інтернет-магазину. Інтернет-магазин підтримує наявність лічильнику переглядів того чи іншого товару, що є досить добре, як для власників інтернет-магазину так і для клієнтів, коли кожна із сторін може розуміти, які товари користуються більшою зацікавленістю серед клієнтів. Є можливість сортування готових конфігурацій за параметром продуктивності в іграх.

До недоліків можна віднести те, що інтернет магазин не вирішує проблеми тих клієнтів, які обирають комп'ютер для роботи або для мобільності. Також при зміні конфігурації у вже зібраному персональному комп'ютері, не реагує на зміну комплектуючих кількість кадрів в секунду, що може викликати нерозуміння зі сторони клієнта і він не зможе зрозуміти наскільки буде продуктивніше комп'ютер з оновленою конфігурацією від нього самого. Також є досить незручно те, що коли клієнт обирає материнську плату, наприклад від AMD і обирає такий самий процесор, то чомусь не блокується можливість обирання процесора від Intel і, якщо клієнт робить цю зміну, то його конфігурація автоматично відновлюється і обирається перша з можливих материнських плат з підтримкою Intel. Також можна виділити, як недолік відсутність рекомендації, щодо конфігурації. Що є незручним відображення ціни окремих комплектуючих, тобто є відображення скільки коштів буде додано за певне комплектуюче до загальної суми конфігурації. Було би набагато краще, якщо просто разом з цією інформацією додавалося значення, скільки коштує окремий компонент. Також є недоліком те, що при самостійній конфігурації комп'ютера, клієнт не може дізнатися, яка буде продуктивність в його майбутньому персональному комп'ютері. Опис одного типу комплектуючих на сторінці конфігурації комп'ютера, є однаковим для всіх

незважаючи на ціну, модель або новизну.

Hyper Ditek – це інтернет-магазин, який спеціалізується на збірці, тестуванню та продажу виключно сучасних та високотехнологічних персональних комп'ютерів. Також компанія займається збіркою ігрових комп'ютерів, графічних та архітектурних станцій, серверів. Створення комп'ютерів під брендом Hyper Ditek здійснюється з урахуванням світових технологій та відповідає усім міжнародним вимогам якості та безпеки. Компанія створена на самперед для справжніх ігromанів. Вона націлена на вирішення різного рівня завдань і пропонує будь-якому геймеру втілити свою мрію в життя, купивши потужний ігровий комп'ютер на будь-який смак та гаманець.

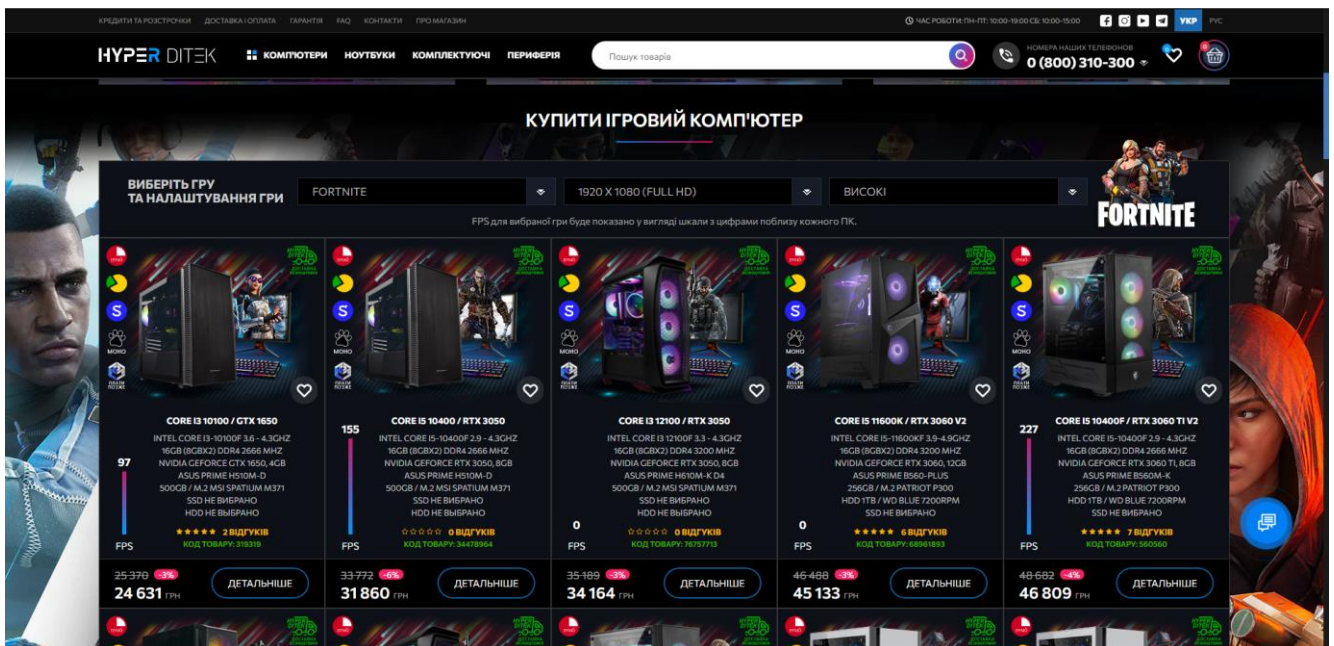


Рисунок 3 – Інтерфейс головної сторінки інтернет-магазину Hyper Ditek

До переваг можна віднести те, що клієнту надається вибір серії комп'ютерів на головній сторінці. В цій можливості комп'ютери поділяються на 3 серії, де «Оптимальна серія», це комп'ютери від 25000 гривень, «Продвинута серія» від 45000 гривень і «Екстримальна серія» від 75000 гривень. Це дає змогу користувачу одразу відсортувати певні збірки персональних комп'ютерів і вже обирати ті, які більш підходять до його гаманця. Є досить приваблива функція динамічного лічильника кадрів в секунду. На головній сторінці з готовими

конфігураціями комп'ютерів є також можливість обрати гру та налаштування гри, щоб подивитись, яка буде продуктивність комп'ютера залежно від його характеристик та налаштувань у грі. Це є досить корисна можливість, особисто для професійних геймерів, які професійно грають в певну гру і для них буде досить зручно подивитись, яка буде продуктивність в їх грі. Також є можливість змінити конфігурацію у вже зібраному комп'ютері і тут також є в наявності динамічний лічильник кадрів в секунду, який реагує на зміни комплектуючих клієнтом. Для готових збірок також за побажанням клієнта можна обрати додаткові аксесуари такі, як монітор, мишку, клавіатуру та навушники. Окрім стаціонарних персональних комп'ютерів магазин також пропонує і мобільні версії, а саме ігрові ноутбуки, це є дуже гарною опцією для гравців, які часто знаходяться у подорожах. Інтернет-магазин має соціальні мережі, які веде активно і це дає клієнту впевненості в якості роботи цієї компанії. Також інтернет-магазин пропонує готову збірку для певної діяльності, наприклад готова конфігурація для 4K геймінгу, для ведення трансляцій на відеостримінгових платформах, або для дизайнерів.

До недоліків можна віднести не дуже привабливий і чіткий для розуміння дизайн, тому що на сторінках дуже багато всього наліплено читати такий дизайн дуже складно для звичайного користувача. Магазин підтримує можливість обрати гру і її налаштування для отримання значення майбутньої продуктивності системи, але налаштування гри мають тільки опції – низькі, високі і ультра. Проблемою є те, що тут є відсутньою опція вибори середніх налаштувань, які є досить популярними серед гравців. При зміні конфігурації клієнту надається дуже мала кількість різноманітності комплектуючих і якщо користувач захоже змінити конфігурацію вже зібраного персонального комп'ютера, то він не зможе замінити процесор на будь-який інший, навіть того ж самого виробника. Для кожного товару є можливість переглянути відгуки і також є в наявності лічильник відгуків, але якщо товар немає відгуків то лічильник все одно зберігає значення один замість нуля. Якщо перейти по посиланню на довільний ігровий ноутбук, то там буде дуже стислий і не інформативний опис характеристик. Магазин має

можливість замовити окремо певне комплектуюче, яке є в наявності, але у деяких товарів можуть бути відсутніми характеристики та опис товару або і те і те одночасно, що є критичним недоліком, так як магазин не інформує клієнта про характеристики того чи іншого товару. Також деякі комплектуючі комп'ютера, які можна подивитись на спеціальній сторінці, мають однаковий опис товару, хоча моделі при цьому відрізняються.

Komputronik – це інтернет-магазин, який спеціалізується на продажі товарів для майже всіх видів діяльності людини. Ця компанія майже двадцять років успішно проводить послідовну політику стабільного та досить динамічного розвитку. Komputronik приділяє особливу увагу якості, як у сфері обслуговування клієнтів так і високих стандартах роботи співпрацівників. Цей інтернет-магазин входить до найпопулярніших інтернет-магазинів Польщі, які також спеціалізуються на продажі комп'ютерної техніки.



Рисунок 4 – Інтерфейс сторінки конфігурації комп'ютера в Komputronik

До переваг цього інтернет-магазину можна віднести зрозумілий та сучасний дизайн. Також не зважаючи на те, що магазин займається продажем майже всього, це не заважає компанії роботи це дуже якісно. Опис товарів є дуже детальним він охоплює всі деталі конфігурації і тому клієнт може більш

детальніше розглядати та порівнювати товари між собою. При виборі вже готових збірок клієнту надається можливість додати до конфігурації встановлення операційної системи або навпаки обрати її відсутність, додати пам'ять диску, а також у випадках, якщо це ноутбук, то можна додати і оперативну пам'ять. Є можливість власноруч створити конфігурацію комп'ютера. Система блокує вибір не сумісних комплектуючих, таким чином, якщо клієнт обирає процесор або материнську плату від Intel, то обрати AMD вже не зможе. І також система одразу фільтрує і набір комплектуючих та показує тільки ті, які він може обрати. Якщо клієнт обирає не всі комплектуючі, які необхідні для роботи комп'ютера, то система попереджує про це і блокує покупку як повноцінної збірки та пропонує купити, як окремі частини. Також при замовленні пропонується обрати можливість збірки комп'ютера на основі обраних комплектуючих, за додаткову плату, майстром від магазину. Ще клієнт може додати до обов'язкових комплектуючих, аксесуари до свого майбутнього персонального комп'ютера.

До недоліків можна віднести відсутність рекомендації, щодо обрання мобільних персональних комп'ютерів. Така можливість реалізована тільки для стаціонарних комп'ютерів. Немає можливості подивитись наскільки буде продуктивною обрана конфігурація комп'ютера і в такому випадку клієнтові власноруч потрібно шукати, якусь інформацію або відео, щоб зрозуміти зможе та конфігурація, яку він обрав, вирішити всі його побажання та потреби.

1.3 Постановка задачі

Враховуючи аналіз предметної галузі та виявлені переваги і недоліки аналогів, було прийнято рішення, щодо дослідження методів, які допоможуть спроектувати рекомендаційну систему для реалізації товарів в інтернет просторі. Рекомендаційна система повинна бути частиною інтернет-магазину, який буде спеціалізуватись на продажі персональних комп'ютерів та комплектуючих. Для

цього необхідно спроектувати інтернет-магазин, який би був більш якісним ніж аналоги, мав зрозумілий інтерфейс та приємний дизайн, а також мав в собі всі функціональні можливості магазину, щоб клієнт зміг придбати товар. Рекомендаційна система повинна надавати користувачу товари, які би змогли його зацікавити на основі його поведінки в системі[2]. Також повинна бути створена додаткова можливість надання рекомендацій товарів, у вигляді можливості швидкого підбору персонального комп'ютера на основі вказаних напрямків використання майбутнього пристрою. Також повинна надаватись інформація, щодо сумісності комплектуючих, аби запобігти виконанню помилок під час конфігурації. Під час обирання комплектуючих, користувач повинен отримувати інформацію, щодо продуктивності майбутнього комп'ютера. Система повинна мати можливість рекомендувати вже готові збірки, на основі побажань користувача.

Для фронтенд частини має бути використані HTML 5, CSS, JavaScript і фреймворки Bootstrap 5 та jQuery.

Бекенд частина буде розроблена за допомогою мови програмування C#[3]. Також необхідно використовувати EntityFramework для доступу та використання бази даних SQL Express LocalDB [4].

2 ФОРМУВАННЯ ВИМОГ ДО ПРОГРАМНОЇ СИСТЕМИ

2.1 Функціональні можливості програмної системи

В ході роботи передбачається створення рекомендаційної системи, яка допоможе створити власну конфігурацію персонального комп'ютера, яка буде відповідати бажанням користувача, а також усуне можливості створення помилок користувачем, під час створення конфігурації. Також місцем для розміщення рекомендаційної системи повинен бути створений інтернет-магазин та його базові функціональні можливості для придбання товару користувачем.

Інтернет-магазин повинен мати в наявності наступні базові функціональні можливості:

- а) зберігання та надання даних про товари;
- б) можливість порівняння товарів між собою, додавання до кошику та додавання до бажаного;
- в) фільтрування та сортування товарів;
- г) можливість зареєструватись та авторизуватись у системі;
- д) можливість перейти до підсумовування коштів за додані до кошику товари, а також можливість оформити придбання товару;
- е) створення зв'язку між банківськими системами для переведення коштів за товари обрані користувачем;
- ж) можливість обрати зручний, для користувача, спосіб отримання товару;
- и) можливість створити власну конфігурацію комп'ютера, або обрати готові варіанти збірок, які будуть запропоновані на основі вказаних критеріїв, які вказують на те, для чого буде використовуватись персональних комп'ютер;
- к) можливість змінювати конфігурацію вже готових збірок персональних комп'ютерів.

Рекомендаційна система повинна мати наступні можливості:

- а) надання інформації, щодо продуктивності тієї чи іншої конфігурації персонального комп'ютера;

- б) створення інтерфейсу обирання певного комплектуючого комп'ютера;
- в) сортування комплектуючих і блокування вибори комплектуючих, які є не сумісними з вже обраними;
- г) наявність лічильника кадрів в секунду для конфігурацій, які спрямовані на кіберспорт та в цілому для ігор;
- д) можливість замовити збірку майстром, яка була власноруч обрана користувачем під час конфігурації комп'ютера, або можливість замовити окремі частини комплектуючих, для самостійного монтування користувачем;
- е) блокування можливості придбання комп'ютера під час його самостійної конфігурації, якщо користувач не обрав всі необхідні для роботи комп'ютера комплектуючі;
- ж) автоматична рекомендація готових конфігурацій на основі поведінки користувача у системі;
- и) можливість швидкого підбору персонального комп'ютера на основі напрямку використання та діапазону коштів, які вкаже користувач.

2.2 Нефункціональні вимоги

Нефункціональні вимоги до програмної системи повинні задати критерії для оцінки того, якою система повинна бути, тому на основі аналізу ринку та встановлених вимог, перша версія програмної системи повинна мати:

- а) повинна бути розроблена навігація для швидкого користування системою;
- б) повинна бути локалізація для української та англійської мови;
- в) розроблений інтерфейс програмної системи повинен бути зрозумілим та зручним у використанні;
- г) сторінки інтернет-магазину повинні завантажувати дані не більше ніж за 5 секунд;

- д) повинна мати можливість адміністрування;
- е) програмна система повинна мати можливість швидкого та безпечного редагування структури та можливість масштабуватись;
- ж) дизайн системи повинен бути розроблений з уживанням сучасних стандартів, а також бути привабливим та зручним для користувача.

2.3 Допущення та залежності

Для успішного користування системою, необхідно дотримуватись наступних допущень:

- а) у користувача має бути пристрій, який має доступ до мережі Інтернет;
- б) користувачі мають використовувати сучасні версії інтернет браузерів.

У якості залежностей можна виділити, те що:

- а) система буде коректно працювати тільки з сучасними версіями інтернет браузерів, таких як: Microsoft Edge, Google Chrome та Safari;
- б) рекомендації на основі поведінки користувача будуть надавати рекомендації товарів тільки, якщо користувач буде мати хоча б одну дію в історії користування системи.

3 ОГЛЯД НАУКОВОЇ ЛІТЕРАТУРИ

3.1 Загальні відомості

Інформація завжди була невід'ємною частиною для життєдіяльності людини, але щоб зрозуміти її сутність нам необхідно було дочекатись 20 віку. Тоді Клод Шенон зрозумів, що інформацію можна перетворити в числа, а якщо точніше, то в два можливих стани. Зміст будь-якої інформації можна було уявити, як 0 або 1, так чи ні, існує або ні і за допомогою цих перетворень з'явилась можливість перетворити будь-яку інформацію в програмний код.

Зробивши інформацію цифровою, людство отримало те, що інформація стала не просто довговічною, але і те, що інформація може зберігатися майже в незчислених обсягах і в той же час люди отримали можливість з дивовижною швидкістю її розповсюджувати.

З часом розповсюдження комп'ютерів серед людей для використання в громадянських цілях, почалось прикладання великої кількості зусиль для того, щоб покращити та спростити робочий процес людини за комп'ютером. Завдяки зусиллям розробників та дослідників, які вивчали потреби людей і намагались зробити більш персоналізовані послуги, які би надавав комп'ютер, почали створюватись більш зручні для користувачів рішення.

Основи рекомендаційних систем були закладені дослідниками в галузі когнітивних наук і галузі інформаційного пошуку. Першим прикладом була комунікаційна система Usenet, яка була створена Університетом Дюка в другій половині 1970 років. Суттю, якої була можливість користувачам обмінюватись текстовою інформацією між собою. Інформація була розділена на новинні групи і підгрупи для спрощення пошуку, але вони не були напряму побудовані на вподобаннях користувачів системи і не були орієнтовані на них.

Першим відомим рішенням був такий собі комп'ютерний бібліотекар Гранді, який спочатку робив опитування користувачів про їхні вподобання, а потім надавав їм рекомендації книг з урахуванням отриманої інформації. На

основі отриманої інформації, система досить примітивним методом розподіляла користувачів в стереотипну групу, рекомендуючи ті ж самі книги усім користувачам в одні такій групі. На сьогоднішній день цей метод може здатися застарілим, але на той час це був прорив в галузі автоматизованих послуг, адже він був персоналізованим. Але навіть нині не всі інтернет-магазини досягли цієї мети.

Все ж таки незважаючи на те, що Гранді було чудовим рішенням, воно отримало дуже багато критики в галузі науки. Науковці зауважували, що людям досить складно з необхідною точністю дослідити та описати власні когнітивні процеси. Згідно дослідженням, люди досить часто підкреслюють тільки ті якості, які виділяють їх серед частини певної групи інших людей, що призводить до складності формування стереотипів поведінки людей. Інколи люди просто прагнуть зробити інший образ самих себе. На цей рахунок, в одному з інтерв'ю, досить різко висловлювалась Хелі Вайнію, яка була менеджером одного з найвеликих торгових центрів Північної Європи, вона повідомляла, що: «Люди заповнюють анкети таким чином, аби вони самі виглядали краще. Мені не потрібна брехня. Мені потрібні факти». Саме тому вона обладнала свій торговельний центр обладнанням Wi-Fi, яке надало змогу відслідковувати клієнтів всередині і в околиці торговельного центру з точністю до 2 метрів. Метою цього рішення було в тому, щоб замість клієнтів за них говорили їх дії.

З часом розвинулись два дуже різних напрямку рекомендаційних систем:

- а) фільтрація на основі контенту;
- б) спільна фільтрація.

Фільтрація на основі контенту полягає в тому, що є необхідність знати параметри об'єктів, які необхідно рекомендувати, наприклад для музикального контенту необхідно враховувати такі параметри, як виконавець, жанр, дата релізу альбому, тощо [5]. Також необхідно знати вподобання користувача згідно цих параметрів або характеристик об'єкту[6]. Таким чином, якщо казати про музику, то кожен раз коли користувачу подобається нова пісня, то ця інформація додається до його профілю.

А спільна фільтрація намагається відобразити смак користувачів, а також пропонує їм контент, який сподобався іншим користувачам з іншими уподобаннями.

Першим прикладом спільної фільтрації, а також початком існування цього терміну була система Tapestry, яка була розроблена Херох PARC. Вона дозволяла користувачам робити нотатки та коментувати документи, які вони читали. На початку це було представлено у вигляді бінарної форми: подобається або не подобається. Таким чином користувачі могли не тільки використовувати контент документів для того, щоб власноручно зменшити межі пошуку, але й за допомогою нотаток та відгуків інших користувачів, які досягали їх певної кількості, змогли досить добре сортувати тематичні документи в рейтингу на основі їх актуальності та корисності. В 1992 році було випущено в реліз GroupLens, вона могла надавати автоматичні рекомендації для статей в Usenet, якщо користувач вже оцінив деякі статі в системі. В наступні роки було створено безліч тематичних сайтів, які могли надавати рекомендації, такі як Ringo, який був створений в Масачусетському технологічному інституті. Пізніше були створені сторінки рекомендації музики Firefly і також система рекомендацій фільмів BellCore. Першим рішенням, яке намагалося охопити не тільки більш вузько направлені теми, але і теми, не менше ніж мережа Інтернет, був Yahoo! тоді він мав іншу назву. Два аспіранти із Стенфорда створили тематичний каталог веб-сайтів, в якому були проіндексовані сторінки, які в свою чергу отримали велику популярність і спростили пошук в Інтернеті для мільйонів людей. На основі рейтингу Alexa він і нині має великий попит і сягає 5 місце серед найвідвідуваних веб-сайтів.

Якщо казати про контентну фільтрацію то свій початок вона бере з галузі інформаційного пошуку звідки були перенесена велика кількість методів. Перше рішення, яке було офіційно задокументовано було від Емануеля Голдберга в 1920 роках. Його рішення відображало собою деяку статистичну машину, яка намагалась автоматично знаходити документи, які зберігались на целулоїдних стрічках, методом пошуку шаблонів. В 1960 роках завдяки групі дослідників з

Університету Корнуола майже за десятиріччя була створена модель автоматичного індексування текстів, яка стала основою відомих на сьогоднішній день процесів аналізу тексту. Цей процес досить простий, документи класифікуються по певним параметрам, які були встановлені заздалегідь, і які збираються у векторну форму як індекси. Чим більше схожі між собою два документи, тим менше кут, який описує їх вектори. Наступним кроком в удосконаленні та поширенні рекомендаційних систем було створення системи онлайн-каталогу CITE, яка була розроблена в 1979 році Тамашем Дошкочем для Національної медичної бібліотеки. Вона дозволяла користувачам шукати не тільки книги по категоріям, але й сортувала їх по релевантності на основі умов пошуку. Контентна фільтрація отримала сенс досить пізно, а саме в 1990 роках у вигляді розділу інформаційного пошуку. Основна причина цьому факту сприяло, те що створення добре функціонуючої системи фільтрації на основі контенту, навіть в конкретній галузі, є дуже великою проблемою, тому що поставлена задача складається з принаймні зрозуміння досліджуваної теми, а також факторів, які мають вплив на взаємозв'язок користувачів до дисципліни.

Одним з перших і досить успішним дослідженням на дану тему був в 1999 році проект Music Genome, його метою було розуміння та сприйняття музики через його властивості. Для цього було виявлено 450 властивостей і був описаний їх взаємозв'язок за допомогою алгоритму. В його основі лежить те, що коли користувачу подобається пісня, то певним її властивостям надаються позитивні значення. Потім пісні зі схожими властивостями просуваються далі в списку вподобань і після рекомендуються користувачу. Дуже великою перевагою є те, що спільній фільтрації при запуску необхідно дуже мало інформації, в той час, як для контентної фільтрації, нажаль потребує більшої кількості користувачів та відгуків для виявлення користувачів із спільними смаками. Проте недоліком системи Music Genome є те, що вона зазвичай з певними складнощами, або зовсім не може надавати рекомендації поза музикальних списків користувача, оскільки алгоритм не базується на схожості між користувачами, а тільки на розумінні властивостей музики, як певного об'єкту. До речі таке інтернет-радіо, як Pandora з

більше ніж 250 мільйонами користувачів засновано саме на цьому проекті і працює зі схожими алгоритмами і нині.

Першим рішенням метою, якого було поєднання двох різних рішень, а саме спільну та контентну фільтрації. Це рішення мало назву Fab, яке було розроблене студентами Стенфорду і представлене в 1994 році. Вони відмічають, що створення гібридної системи була націлена на усунення недоліків процедур обох фільтрацій, які були виявлені з часом. Модель гібридної системи базується на двох основних процесах: спочатку відбувається збір контенту по певним темам, після першого етапу відбувається вибір для кожного користувача тих елементів, які швидше за все їх зацікавить і їх зміст буде порекомендований користувачу.

Поєднання двох рішень може бути реалізуватись декількома способами. Одна процедура може бути вбудована в іншу, як приклад рішення Fab, або можна надати спільну рекомендацію в результаті двох процедур, як наприклад, це робить Netflix. Алгоритм Netflix, CineMatch був дуже успішною рекомендаційною системою для онлайн-продажу фільмів на початку 2000 років. Це стало серйозним катализатором для досліджень і наукових напрямків, які почали стрімко розвиватися. Задачею премії Netflix в 2006 році було те, що потрібно створити рекомендаційний алгоритм на основі 100 мільйонів обзорів фільмів, які вони надали. Алгоритм повинен був надавати рекомендації, як мінімум на 10% краще ніж були результати CineMatch. В 2009 році, премія в 1 мільйон доларів була присуджена за рішення, яке поєднувало 107 різноманітних алгоритмів і змішувало їх рекомендації в залежності від обставин.

Також існує гарний приклад реферальних онлайн-систем Amazon.com, які рекомендують продукти користувачам на основі методу спільної фільтрації, приймаючи до уваги переглянуті і придбані товари раніше, а також те, що вони переглядають на даний момент часу. Цей метод нині використовується багатьма інтернет-магазинами для поліпшення показників продажу. Мета рекомендаційних систем, які працюють в клієнтському онлайн-просторі, полягає в тому, щоб налагодити вітрину магазину згідно з поточними смаками та потребами клієнтів. Такі методи створюють дуже велику конкурентну перевагу для інтернет-

магазинів порівняно з традиційними покупками в звичайних магазинах, єдиною сильною характеристикою, яких полягає в тому, що товар можна подивитись, доторкнутись до нього та приміряти в реальному житті. Але нині досить часто виходить так, що після примірювання товару в фізичному магазині, клієнт робить замовлення онлайн, або ж робить замовлення з використанням можливості повернення товару без примірювання товару в фізичному магазині.

3.2 Огляд наукових публікацій

В ході аналізу предметної галузі і виявлення її проблем було прийнято рішення звернутись до наукових публікацій, які би допомогли обрати необхідний метод вирішення сучасних проблем в галузі і дослідити його. У якості місця пошуку наукових публікацій було обрано портал ResearchGate, так як цей портал є відкритим для майже усіх країн світу і вже більше ніж 20 мільйонів дослідників поділилися своїми науковими дослідженнями, які охоплюють досить велику кількість галузей.

Для пошуку публікацій я використовував запити рекомендаційних систем, а саме створення рекомендаційних систем за допомогою методів спільної фільтрації (Collaborative filtering) та контентної фільтрації (Content-based filtering). Для подальшої розробки програмного забезпечення необхідно було дослідити ці два методи побудови рекомендаційних систем, які досить близько відповідають потребам для рішення проблематики досліджуваної галузі. Для більш детального дослідження було обрано роботу автора Xinyi Wu з Університетського коледжу Лондона під назвою «Порівняння між спільною фільтрацією та контентною фільтрацією»[7]. Метою роботи автора було дослідження двох методів побудови рекомендаційної системи, які вирішують сучасну проблему більш швидкого і ефективнішого отримання необхідною для користувача інформації. Також в роботі виконаний детальний опис ідеї та принципів роботи кожного з методів

побудови рекомендаційних систем.

Слід зауважити на тому, що автор в ході дослідження описав поступові кроки виконання цих двох методів та їх процесів і на основі цього виявив переваги та недоліки кожного з методів побудови рекомендаційних систем, а також в яких галузях ці методи мають великих вплив та більшу популярність. Крім цього автор зауважив, які обмеження та недоліки є нині, а також напрямок для подальшого вдосконалення в майбутньому.

Основною проблемою але і водночас покращенням життя людей є те, що доступ до інформації з часами дуже розвинувся і став більш зручним і різноманітним. Але щоденно в мережу Інтернет додається дуже великий обсяг інформації. З одного боку збільшення інформації насичує життя людей та прискорює зв'язок та навчання в суспільствах. З другого боку надмірний обсяг даних призводить до певних труднощів в фільтрації інформації. Надлишок інформації також має в собі таких характер, що люди не можуть знайти певну інформацію, яка би відповідала їх потребам, цей феномен також називається «інформаційне перевантаження».

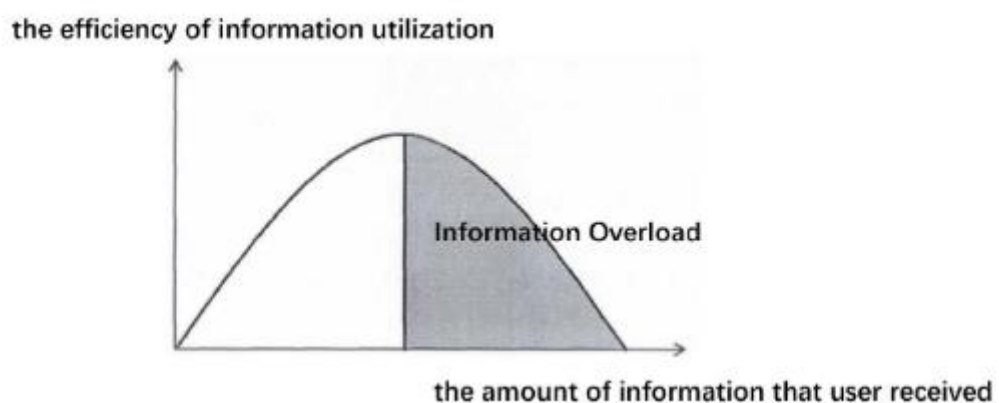


Рисунок 5 – Взаємозв'язок між здатністю користувачів обробляти інформацію і об'ємом інформації

В такому випадку швидкість з якою користувач може використовувати інформацію стрімко знижується. Тому це стало основною проблемою всіх великих інтернет-ресурсів, а також пріоритетною задачею стало рішення, яке

спричиняє обмеження або навіть видалення не потрібної інформації з дуже великого об'єму невпорядкованих даних і те, яким чином можна зробити пошук інформації для людей більш швидким та ефективним.

Тому для вирішення проблеми інформаційного перевантаження було розроблено рекомендаційну систему для фільтрації інформації для поліпшення користувацького досвіду і задоволення потребам користувачів різноманітних сайтів та створення більш персоналізованого обслуговування.

Існує безліч різноманітних алгоритмів для створення рекомендаційних систем, але зазвичай вони поділяються на два основних, це: спільна фільтрація (collaborative filtering), контентна фільтрація (content-based filtering).

Так відбулось, тому що серед усіх класичних рекомендаційних алгоритмів, нині найбільше розповсюдження отримали саме ці два алгоритми.

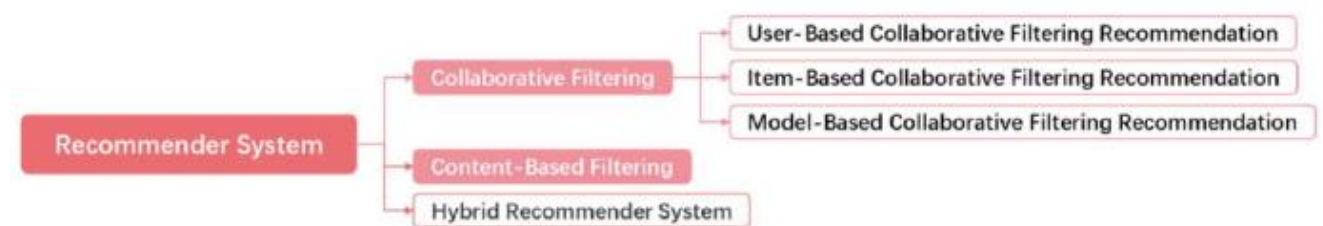


Рисунок 6 – Класифікація рекомендаційних систем

Спільна фільтрація (Collaborative filtering) виділяється серед безлічі видів рекомендаційних систем, тим що досить широко вона використовується серед багатьох соціальних інтернет-платформ та ресурсів, завдяки передовій концепції і повному використанню своїх унікальних переваг для персоналізованої рекомендації інформації користувачу. Ідея алгоритму полягає в тому, що користувачі з однаковими вподобаннями можуть мати інші схожі інтереси. Алгоритм спільної фільтрації використовує декілька методів для пошуку та збору інформації, яку залишають користувачі при взаємодії в мережі Інтернет. Потім він порівнює велику кількість різноманітної користувацької інформації і аналізує подібність між цією інформацією. Таким чином потім алгоритм може легко передбачити потенційні вподобання користувачів певної платформи і

рекомендувати їм певні товари або послуги.

Автор публікації виділяє два принципи роботи спільної фільтрації, це: обробка даних спільною фільтрацією та класифікація даних.

Алгоритм обробки даних за допомогою спільної фільтрації має наступні етапи: збір даних про вподобання користувачів та даних про певні об'єкти, поступове очищення та нормалізація зібраних даних, кількісне порівнювання схожості між користувачем і певним об'єктом.

Алгоритм спільної фільтрації має характеристики, які надають змогу мати безліч каналів для збору різноманітної користувацької інформації. Для збору інформації ефективно допомагає комплексний метод збору необхідної інформації до бази даних. Це сприяє подальшій розробці персоналізованих рекомендацій для користувачів. Під час користування веб-сайтом, поведінка користувача буде зберігатися в базі даних, наприклад натискання, зберігання або додавання позначки. Також алгоритм спільної фільтрації зберігає деякі записи, на які користувач не обертає свою увагу. До таких записів відноситься: кількість та порядок натискань, адреса особистого інтернет-протоколу, а також час знаходження на певній сторінці веб-сайту. Перевагою алгоритму спільної фільтрації полягає в тому, що для його системи записується усі об'єкти, які навіть не цікавлять користувачів, або отримують негативні відгуки або зовсім ігноруються, це все є необхідним для досягнення мети, яка полягає у фільтрації інформації.

Щоб зробити результати фільтрації інформації більш точними і порекомендувати потрібний користувачу об'єкт, необхідно зробити очищення сторінки від «шуму». Коли алгоритм спільної фільтрації збирає дані, є деяка ймовірність того, що частина не основного контенту або інформація яка не належить до контенту сторінки будуть зібрані алгоритмом. Також є ймовірність, що користувач може зробити певні дії під час роботи з сторінкою веб-сайту, які можуть до цього призвести. Все це є потенційною перешкодою для алгоритму, який надає рекомендації користувачам. Для аналізу необхідних даних, програмісти зазвичай віддають перевагу використанню деяких додаткових

класичних і зручних алгоритмів інтелектуального аналізу даних для фільтрації зайвої інформації, які зібрав алгоритм. Окрім очищення від «шуму», алгоритму необхідно нормалізувати дані. Нормалізація полягає в тому, щоб зіставити дані кожної поведінки з заданим діапазоном, а потім обробити їх, щоб загальне вподобання користувача отримане шляхом зваженого сумування, стало більш точним. Нормалізація також дозволяє більш зручніше і швидше оброблювати дані, крім того, вона також дозволяє порівнювати і зважувати об'єкти різних одиниць та розмірів. Вона дозволяє спростувати складний розрахунок і також додатково розширює існуючі дані.

Після попередніх двох етапів алгоритм повинен створити з зібраної інформації, набори ознак користувачів, а після провести кількісне порівняння. В алгоритмі спільної фільтрації існує багато рішень розрахунку схожості. Серед таких рішень, велику популярність має порівняння схожості на основі відстані (1).

$$similarity = \frac{1}{1 + d} \quad (1)$$

Відстань d можна отримати за допомогою евклідової та манхетенської відстані. Також розповсюдженим методом є порівняння схожості набору контенту.

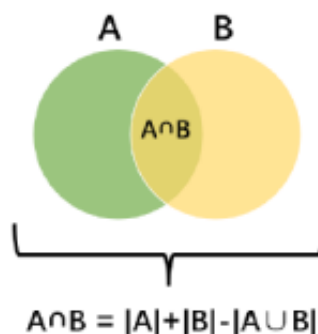


Рисунок 7 – Схожість Жаккара зображене діаграмою Венна

Наприклад схожість Жаккара відображає ступінь збігу тих самих елементів

в A і B через діаграму Венна (2).

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (2)$$

Загалом суть алгоритму спільної фільтрації полягає в використанні спільного розуму в кластері. Коли програма почне роботу, з початку буде створена база даних об'єктів вподобань користувачів, а потім буде розрахована схожість між персонажами користувачів. Це необхідно для того, щоб зіставити користувачів зі схожими інтересами та захопленнями, що є досить корисно для виконання рекомендації.

Існує багато варіацій використання спільної фільтрації і для різних галузей використовуються різні алгоритми. Але незважаючи на їх безліч та складність, їх основні теорії та принципи побудування дуже схожі. Різниця полягає в тому, яку схожість необхідно розрахувати. Тому алгоритм спільної фільтрації поділяється на три підкатегорії:

- а) спільна фільтрація на основі користувачів (User-based collaborative filtering);
- б) спільна фільтрація на основі елементів (Item-based collaborative filtering);
- в) спільна фільтрація на основі моделей (Model-based collaborative filtering).

Спільна фільтрація на основі користувачів підходить для платформ з відносно невеликої кількістю користувачів і індивідуальних відмінностей. Наприклад, в останні роки цей алгоритм був дуже популярним серед сайтів знайомств серед меншин з унікальними вподобаннями. Тому коли він порівнює загальні інтереси і вподобання користувачів, алгоритм вибирає із безлічі користувачів, які мають багато схожостей для надавання один одному взаємні рекомендації.

Алгоритм спільної фільтрації на основі елементів більше пасує для ситуацій, коли кількість елементів менше ніж кількість користувачів. Він також гарно використовується до галузей з сильними персоналізованими вимогами,

такими як рекомендація фільмів та огляди книжок. Також він досить широко використовується на платформах націлених на інтернет-покупки електронної комерції і коли користувач хоче знайти щось схоже на цільовий продукт, система буде надавати рекомендації в базі даних.

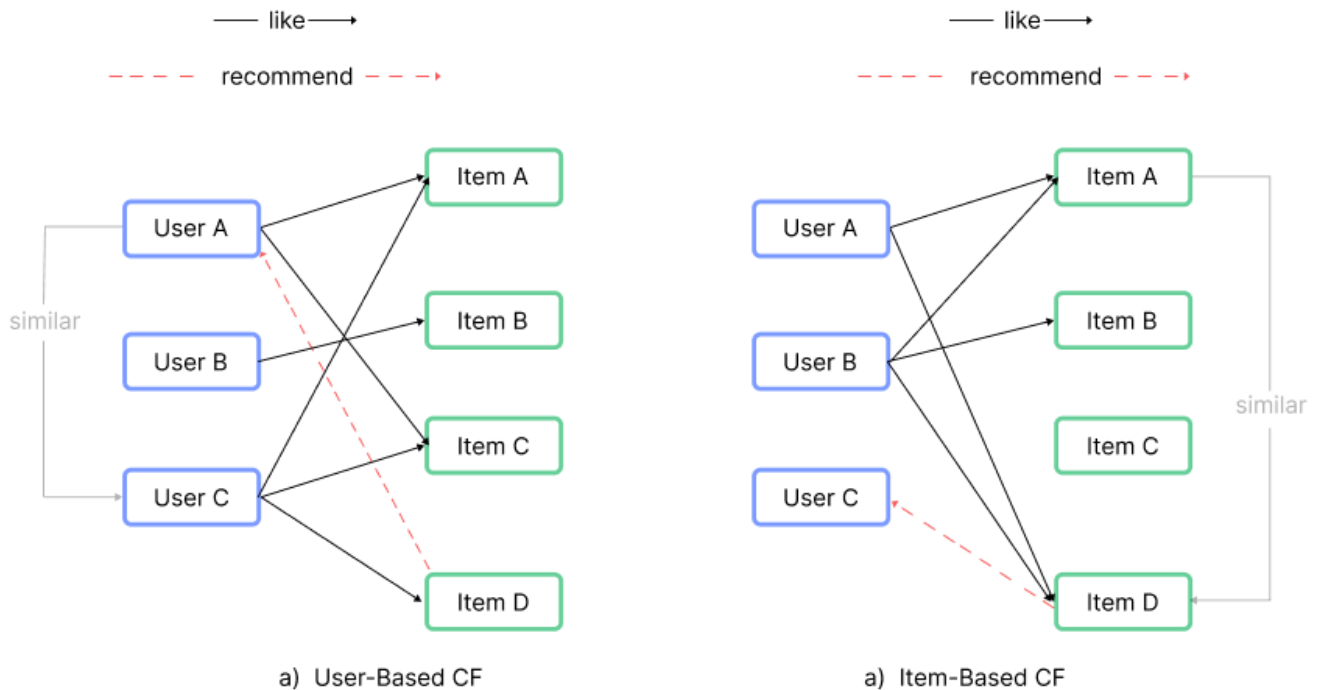


Рисунок 8 – Порівняння між спільною фільтрацією на основі користувача та на основі елементу.

Спільна фільтрація на основі моделей, це система рекомендацій, яка заснована на інформації про поведінку і вподобання користувачів. Він надає рекомендації через аналогічні моделі.

Виділяючи переваги алгоритму спільної фільтрації можна виділити те, що алгоритм досить легко отримує інформацію про поведінку користувачів. Інформація алгоритму заснована на поведінці окремих користувачів, наприклад на покупці і оцінці певних товарів. Таким чином програмістам не потрібно власноручно отримувати і узагальнювати інформацію і тому порівняно з іншими рекомендаційними алгоритмами, за допомогою нього, дуже легко та зручно отримувати характеристику елементів з певних об'єктів. Також він може

аналізувати та класифікувати об'єкти в галузях, які є складні для точного визначення, такі як мистецтво та музика. Також спільна фільтрація може розширювати інформаційне поле отримане від користувачів і це також є одною з важливих причин популярності цього алгоритму.

В порівнянні з іншими схожими алгоритмами, спільна фільтрація має більш високу персоналізовану швидкість навчання. При першому використанні веб-сайту, або іншої певної платформи, користувачам не потрібно вказувати та обирати власні інтереси. Навпаки, платформа з використанням алгоритму буде автоматично оцінювати та доповнювати типи захоплень користувачів відповідно до історії їх поведінки.

Також він може використовувати відгуки та вподобання інших користувачів відносно до аналогічної інформації або об'єкту для автоматичного покращення контенту, який рекомендується. Таким чином алгоритм досягає досить гарного ступеню автоматизації з високим рівнем інтелектуалізації та персоналізації.

До недоліків можна віднести проблему «холодного старту», вона є досить серйозною для алгоритму спільної фільтрації[8]. Це пов'язано з тим, що алгоритм базується на відгуках користувача. Оскільки користувач, щойно зареєструвався на певній платформі, то є деяка проблема в тому, що система не може передбачити інтереси цього користувача, тому що не вистачає записів про історію його поведінки. Тому на цьому етапі є досить частим випадки коли користувач отримує непотрібну інформацію у якості рекомендацій.

Також рекомендаційні системи деяких популярних сайтів зазвичай мають в собі мільярди користувачів та записи відслідковування їх даних. Чим більше розмір бази даних, тим більше буде розрідженість, це шкідливо для алгоритму спільної фільтрації, який спрямований на відслідковування схожості користувачів. Тому що, якщо користувачі нададуть декілька позитивних або негативних оцінок невеликій кількості об'єктів, то системі буде досить складно зробити певні висновки на основі тільки лише відгуків користувачів, а невдала спроба побудувати відношення схожості може призвести до відмови рекомендаційного алгоритму.

По мірі зростання кількості користувачів і розширення області інтересів, кількість елементів в базі даних також зростає. Таким чином, масштаб інформації в базі даних також починає розширюватись і це призводить до експоненціального збільшення об'єму необхідних розрахунків і навантаження на процесор, який оброблює дані на сайті також буде зростати. В результаті при накопиченні досить великого масштабу інформації, яка записана в базу даних, обробка та алгоритм порівняння великого об'єму даних призведе до уповільнення роботи усього веб-сайту.

Переходячи до опису фільтрації на основі контенту треба сказати, що технологія пошуку інформації є дуже важливою частиною для цього методу[9].

Основна ідея алгоритму фільтрації на основі контенту полягає в кількісній оцінці схожості текстової інформації об'єктів, а потім побудування на цій основі конкретної моделі. Потім алгоритм поглиблюється в зв'язані з користувачем дані і потім точно прогнозує можливу їх поведінку для точної рекомендації інформації, яка би могла їх зацікавити. Принцип роботи алгоритму контентної фільтрації майже такий самий, як і для спільної фільтрації. Різниця полягає лиш у тому, що алгоритми спільної фільтрації зазвичай вдосконалюють власну систему рекомендацій за рахунок відгуків користувачів на ту чи іншу інформацію, яку пропонує сайт, в той час, як алгоритм контентної фільтрації потребує, щоб система точно отримувала інформацію про об'єкти та поведінку користувачів. Це робиться для створення певних портретів вподобань користувачів для певних елементів, щоб знайти цільові об'єкти, які відповідають їх характеристикам.

Всього можна виділити три головні етапи обробки даних методом контентної фільтрації:

- а) отримання інформації від користувачів або подій;
- б) перетворення набору ознак в вектори ознак;
- в) імпорт даних у векторний простір і подальше порівняння.

Алгоритм контентної фільтрації полягає в порівнянні ступеню схожості шляхом отримання текстової інформації про певні елементи[10]. Він може отримувати текстову інформацію про користувачів або елементи веб-сайту, таку

як особисту інформацію користувача і маркетинговий текст опису продукту. Ці атрибути текстових об'єктів алгоритм організує в набори ознак користувачів[11].

Після створення наборів ознак, алгоритм буде рівномірно цифрувати їх і тим самим перетворювати в вектори ознак. В алгоритмі контентної фільтрації найбільш розповсюдженим є вибір моделі векторного простору у якості основної форми для векторів ознак. Цей алгоритм відтворює всю важливу інформацію у вигляді тривимірних векторів для подальшого порівняння.

Після перетворення векторів ознак, алгоритм імпортує ті вектори, які необхідно порівняти в один і той же векторний простір. На даному етапі завжди використовується косинусна схожість для кількісної оцінки і розрахунку схожості між векторами[12]. Косинусна схожість є найбільш часто використовуваним методом розрахунку в даному алгоритмі.

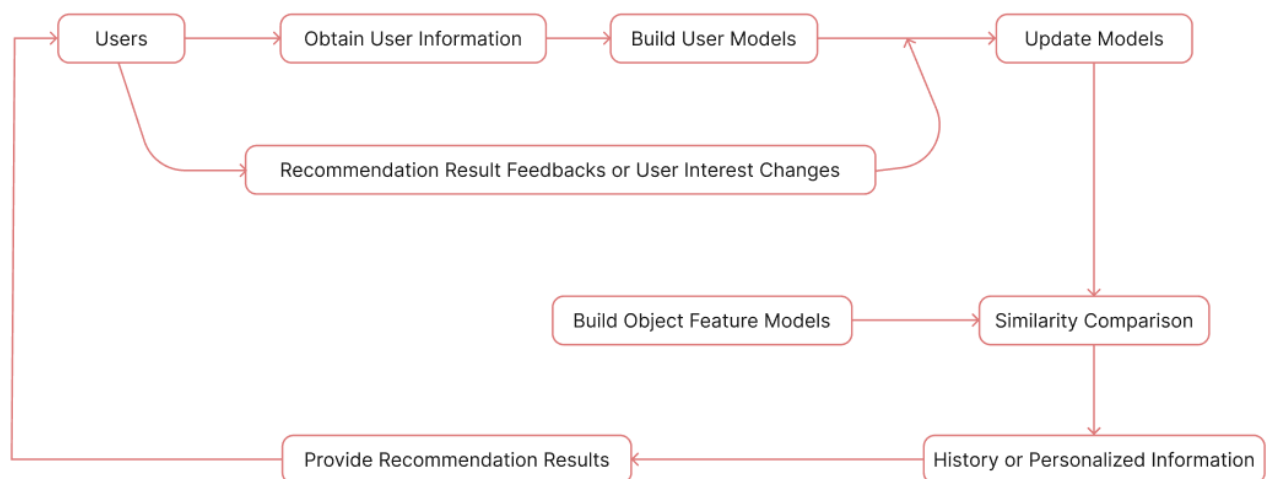


Рисунок 9 – Діаграма алгоритму контентної фільтрації

До переваг алгоритму контентної фільтрації можна віднести те, що він не залежить від інших користувачів і менше має вплив на проблему холодного запуску і розрідженості даних. Алгоритм базується на інформації поведінки користувача для пошуку можливих елементів, які можуть представляти інтерес для подальшої рекомендації користувачу. Таким чином, навіть якщо на сайті є історія тільки про єдину поведінку користувача при виконванні певної функції,

то алгоритм може рекомендувати релевантний контент для користувачів на основі змісту цієї функції. Завдяки цьому легко уникається проблема запуску та легше вирішується проблема розрідженості даних.

Оскільки алгоритм фільтрації на основі контенту повністю пов'язаний з історією поведінки користувача, об'єкти які рекомендуються є тим, що користувач вже знає, а не чимось новим, що може викликати у користувача неприємні почуття. Порівняно з алгоритмами спільної фільтрації у платформи є досить причин, які можуть пояснити, чому рекомендується певний продукт і також користувачі мають досить високий ступінь визнання цих рекомендацій.

До недоліків можна віднести складність отримання інформації. Для того, щоб спростити порівняння різноманітних видів інформації, перед повторним порівнянням алгоритм повинен спочатку перетворити базову інформацію про користувачів у вектори ознак з чіткими значеннями. Однак люди відносяться до різних культур і мають різне походження, що може призвести до того, що інформація о різноманітних зображень, музиці або відео не може бути отримана з високою точністю за допомогою існуючих технологій. Це може призвести до помилок в процесі пошуку необхідної інформації. В порівнянні з алгоритмом спільної фільтрації, отримана інформація має більшу ймовірність мати «шум», що збільшує складність її обробки. Також алгоритм в більшому ступені залежний від професійних знань в галузі обробки інформації ніж алгоритм спільної фільтрації.

Якість текстових ознак отриманих з текстової інформації, часто виявляє рекомендаційний ефект алгоритму. Однак йому іноді не вдається виділити суттєву інформацію на основі змісту тексту і в результаті цього точність рекомендацій зменшується, а аналіз інформації знаходиться на поверхневому рівні, що є очевидним недоліком.

Алгоритм фільтрації на основі контенту використовує поведінку користувачів у якості основного стандарту вимірювання. Тому якщо в базу даних не буде додана певна інформація о діях користувача, алгоритму буде складно рекомендувати різноманітні об'єкти.

В ході огляду наукових публікацій та інформації про досліджувані методи

було отримано велику кількість корисної інформації про основні методи побудови рекомендаційних систем, а також методи спільної фільтрації та фільтрації на основі контенту. Покроковий аналіз процесів цих двох методів дозволяє точно проаналізувати та узагальнити їх переваги та недоліки. В ході огляду літератури було виявлено, що алгоритми спільної фільтрації досягає високого рівня інтелекту, персоналізації та автоматизації за рахунок використання можливостей суспільства, що дозволяє зберігати новизну результатів власних рекомендацій. Однак існують великі проблеми з «холодним запуском» і розрідженістю даних. В той час, як для алгоритму на основі контенту є великі переваги у простоті побудови системи рекомендацій та зручність його роботи. Але аналогічно з попереднім методом питання підвищення точності отримання важливої інформації є досі актуальне.

4 АРХІТЕКТУРА ТА ПРОЕКТУВАННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

4.1 Проектування архітектури програмного забезпечення

В ході встановлення вимог до системи було прийнято рішення створити класичну багаторівневу архітектуру для програмного забезпечення, ця архітектура має три основні шари і містить: шар доступу до даних, логіки та шар клієнту. До переваг цієї архітектури можна віднести:

- а) досить зручну масштабованість;
- б) продуктивність роботи;
- в) доступність та легкість у використанні.

Для реалізації цієї архітектури використовуються можливості ASP. NET [13].

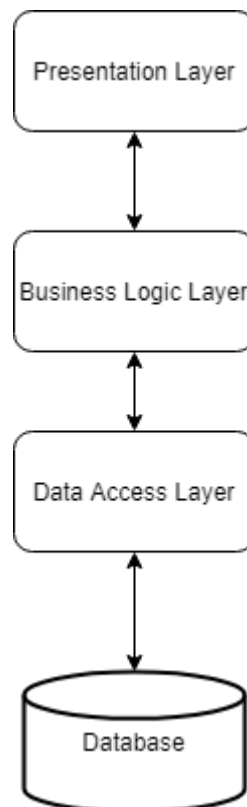


Рисунок 10 – Схема обраної архітектури програмного забезпечення

Presentation Layer – це рівень, на якому виконується зв'язок між

користувачем. Цей рівень поєднує в собі усі компоненти інтерфейсу та можливості взаємодіяти з ним.

Business Logic Layer – це рівень, який має в собі різноманітність класів та методів, що виконують низку обчислювань даних, які були отримані від Presentation Layer. Також цей рівень реалізує зв'язок між базою даних та передає результати обробки знову до рівня представлення.

Data Access Layer – цей рівень зберігає дані моделей та класів, які зберігаються в базі даних і також репозиторії через, які виконується зв'язок між рівнем бізнес логіки та базою даних [14]. Запитувана інформація потім знов повертається на рівень логіки і вже після її отримує користувач.

Кожен з рівнів архітектури є окремим та залежить тільки від рівнів, які знаходяться нижче. Кожен рівень виконує тільки свій набір функціональних можливостей, що спрощує редагування вмісту кожного з рівнів та запобігає створюванню конфліктів між рівнями.

4.2 Проектування UML

В ході проектування програмного забезпечення було вирішено створити діаграму розгортання за допомогою UML [15]. Дану діаграму наведено нижче.

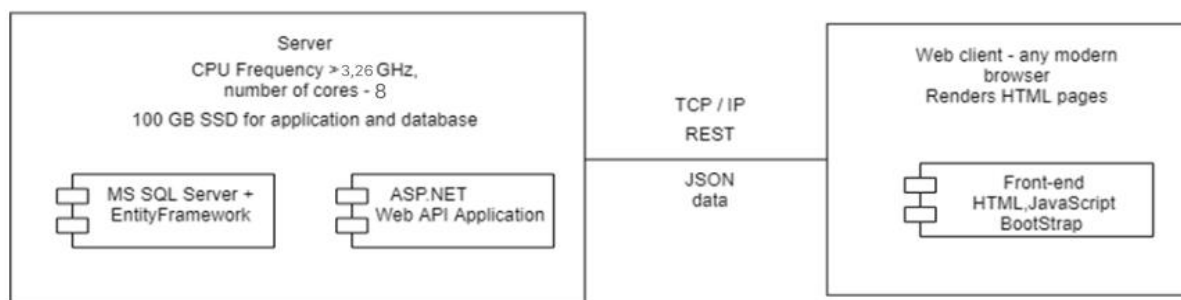


Рисунок 11 – Діаграма розгортання

Діаграма розгортання – це діаграма UML, яка відображає компоненти,

об'єкти, які виконуються на обчислювальних вузлах, та власне обчислювальні вузли під час роботи програми. Веб-клієнт буде являтися для користувача інтерфейсом, з яким він буде працювати та щоб переглянути інформацію з сервера буде здійснюватися зв'язок REST[16].

Також була створена діаграма поведінки користувача у рамках використання частини системи з можливістю створіння власної конфігурації персонального комп'ютера, а також отримання рекомендацій, щодо вже готових конфігурацій, які були куплені іншими користувачами та оцінені, і які відповідають потребам користувача(див. рис. 9).

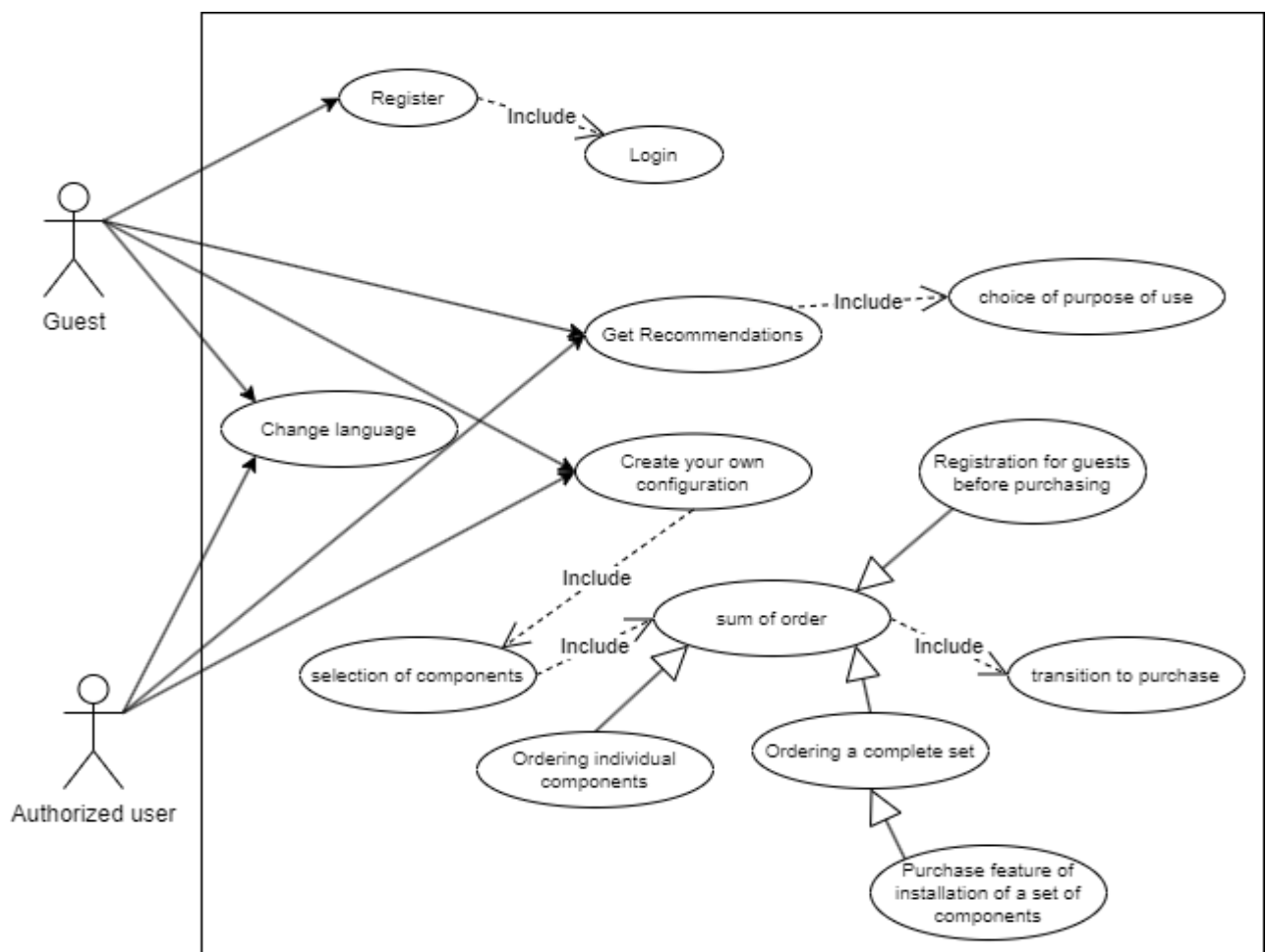


Рисунок 12 – Діаграма варіантів використання можливостей створення власної конфігурації та отримання рекомендацій

Створення діаграми варіантів використання, надає розуміння поведінки системи, а також описує різноманітність запитів користувача до системи. Так як

загальний проект складається з двох частин, це частина класичного інтернет-магазину та частина побудови рекомендаційної системи для кращою реалізації товару. Головною метою роботи є дослідження методів побудови рекомендаційної системи, тому в діаграмі використання було наведено саме цю частину.

При взаємодії з цією частиною системи користувач має можливість створити власну конфігурацію персонального комп'ютеру, а також отримати готові варіанти конфігурацій, які вже були придбані іншими користувачами і мають відгук.

В рамках цієї частини проекту, користувачі поділяються на гостя та авторизованого в системі користувача. Функціонал системи з якою може працювати кожен з типів користувачів майже не відрізняється. Гостю надається можливість зареєструватись, а потім авторизуватись у системі.

Також згідно діаграми можна побачити, що користувачі можуть змінити мову інтерфейсу системи.

Користувачі системи можуть створити власну конфігурацію комп'ютера для цього вони повинні перейти до сторінки конфігуратора і почати створювання. Для придбання повної збірки необхідно обрати всі необхідні комплектуючі, які необхідні для роботи персонального комп'ютера. Також користувачу надається можливість зробити замовлення окремих комплектуючих. Якщо користувач у якості оформлення замовлення обрав придбання повного набору комплектуючих, то система надає можливість придбати додаткову послугу монтування цих комплектуючих, тобто користувачу не потрібно власноруч збирати усі комплектуючі для роботи комп'ютера, цим будуть займатись співробітники магазину, а користувач вже отримає готовий до роботи комп'ютер. Після вказування усіх необхідних даних, авторизований користувач переадресується до сторінки оформлення замовлення та остаточної купівлі продукту, а гостю перед придбанням товару необхідно буде зареєструватись у системі.

4.3 Проектування структури бази даних

У якості СУБД було обрано MS SQL Server Express, це система управління базами даних, яка використовує для запитів мову SQL. У якості переваг цієї СУБД можна виділити:

- а) дуже швидкий доступ до даних;
- б) легкість у використанні та адміністрування бази даних;
- в) безпека зберігання даних, завдяки їх шифрування та резервного копіювання.

У якості вбудованої бази даних було використано LocalDb, це легковажний двигун, який є спрощеною версією SQL Server Express. Він ідеально підходить для експериментів дослідження, тому що має швидке встановлення та спрощену роботу з сервером бази даних. Після встановлення зв'язок з базою даних відбувається за допомогою рядка підключення.

Для взаємодії з даними бази даних використовується EntityFramework, який підтримує підхід Code first, суттю якого є написання коду моделей на мові програмування C#, по яким генеруються таблиці в базі даних[17].

5 ОПИС ПРИЙНЯТИХ ПРОГРАМНИХ РІШЕНЬ

5.1 Опис засобів розробки

Для розробки програмного забезпечення використовується середовище розробки Visual Studio та мова програмування C#. Середовище Visual Studio має досить велику кількість підтримуючих технологій та інструментів, які допомагають в розробці програмного забезпечення.

Також використовується ASP.NET, який призначений для створення веб-застосунків, його зручним рішенням є IIS сервер, на якому можна запускати програмне забезпечення, це дуже спрощує проведення експериментів розроблюваного ПЗ. Також він базується на Common Language Runtime, що є основою усіх додатків Microsoft.NET і дозволяє використовувати мову програмування C# та інші, які входять в набір .NET Framework.

Для зберігання даних застосовується Local db та EntityFramework для зв'язку з базою даних.

Також для розширення можливостей додатку повинна використовуватись технологія Web Api 2. Вона надає можливість додаткового отримання даних з серверу, а також можливість підключення до інших додатків за умовою налагодження зв'язку між ними.

5.2 Опис прийнятих програмних рішень

В ході дослідження було розглянуто основні методи побудови рекомендаційних систем, це: спільна фільтрація та фільтрація на основі контенту. Провівши дослідження наукової літератури було прийнято рішення використовувати фільтрацію на основі контенту.

Основна ідея алгоритму фільтрації на основі контенту полягає в кількісній

оцінці схожості текстової інформації об'єктів, а потім побудування на цій основі конкретної моделі. Це дуже гарно підходить для обробки інформації про товари для подальшого знаходження схожості між вподобаннями користувача та усіма товарами у системі і на основі цього аналізу пропонування рекомендації інших товарів, які би могли зацікавити користувачів.

Для контентної фільтрації існує 3 основних етапи обробки даних, це:

- а) отримання та обробка інформації про поведінку користувача в системі;
- б) побудова набору ознак та перетворення даних у векторну форму;
- в) порівняння даних у векторній формі та використання методів для знаходження схожості.

Перевагами контентної фільтрації від спільної фільтрації є те, що вона не залежить від інших користувачів. Він базується на поведінці окремих користувачів та пошуку інформації в товарах, яка цікавить того чи іншого користувача. Тому це вирішує проблему холодного старту, коли для рекомендації певного товару, алгоритму потрібно хоча б єдиний запис про поведінку в історії користувача.

Для повноцінної роботи алгоритму автоматичної рекомендації певних товарів, потрібно створити необхідні таблиці в базі даних для розміщення потрібних для розрахунків даних. Система повинна мати дані про: загальні дані зареєстрованих користувачів, дані про поведінку користувачів у системі, інформація про окремі комплектуючі, інформація про готові конфігурації комп'ютерів, дані про всі товари в магазині.

Тож робота алгоритму в рамках проекту полягає в тому, що необхідно створити словник за допомогою, якого буде розраховуватись TF-IDF для отримання ваги певних слів у товарі та подальшому знаходженню схожості. Кожна повна конфігурація комп'ютерів має в собі набір комплектуючих для роботи цього пристрою. Усього кожна конфігурація може мати не менше ніж 7 комплектуючих, це: процесор, графічна карта, материнська плата, оперативна пам'ять, накопичувач, блок живлення та корпус. Кожне комплектуюче зберігає в собі інформацію про напрямок використання під назвою «позначки». Для

початкової версії будуть доступні наступні позначки: для кіберспорту, для офісу, гібрид (для ігор та для роботи в складних програмах).

Для алгоритму використовується словник, який має в собі ці позначки. Далі він збирає інформацію про поведінку користувачів, важливими даними будуть ті, коли користувач переглядав певний товар, додавав його в бажане або в кошик. На основі цього обирається остання дія користувача по заданому критерію та відповідно обирається той товар з яким він взаємодіяв. Далі товар, з яким взаємодіяв користувач порівнюється з усіма готовими конфігураціями в базі даних. Для кожної з конфігурації розраховується TF-IDF за наступною формулою:

$$tf - idf(t, d, D) = tf(t, d) \times idf(t, D) \quad (3)$$

де tf – відношення кількості входжень деякого слова до загальної кількості слів в документі,

idf – зворотна частота, з якою слово зустрічається в усіх документах.

Tf розраховується за наступною формулою:

$$tf(t, d) = \frac{n_t}{\sum_k n_k} \quad (4)$$

де n_t – кількість входжень слова t в документ,

n_k – загальна кількість слів в документі.

Idf розраховується за наступною формулою:

$$idf(t, D) = \log \frac{|D|}{|\{d_i \in D | t \in d_i\}|} \quad (5)$$

де $|D|$ – загальна кількість документів,

$|\{d_i \in D | t \in d_i\}|$ – кількість документів з загальної кількості, в яких зустрічається t .

Після розрахунків TF-IDF для товару, з яким взаємодіяв користувач та іншими товарами у базі даних. Отримані значення представляються у вигляді векторів. Далі ці вектори розраховуються за допомогою методу косинусної схожості метою, якого є визначення схожості між двома порівнюваними товарами. Косинусна схожість розраховується за наступною формулою (6).

$$similarity = \cos(\theta) = \frac{A * B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (6)$$

де A та B - вектори ознак.

Після розрахунків порівнюється отриманий коефіцієнт схожості між товарами, де значення ближче до 0 означає, що товари не схожі, а значення ближче до 1, навпаки має високу схожість. Алгоритм виконує розрахунки і після обирає 5 товарів, які мають дуже велику схожість з порівнюваним товаром, тобто тим, з яким взаємодіяв користувач і вже безпосередньо рекомендує ці товари користувачу, які виводяться на головну сторінку з усіма товарами.

Також проект має додаткову можливість рекомендацій. На сторінці конфігуратора персонального комп'ютера, система пропонує можливість перейти безпосередньо до створення власної конфігурації, або користувач може скористатися функцією швидкого підбору готової конфігурації. Для цього перед отриманням рекомендації, користувач повинен обрати мету використання пристрою, а також ціновий діапазон. Для вибору позначок початкова версія системи буде мати 3 варіанти: для кіберспорту, для офісу, гібрид (для ігор та роботи в складних програмах).

На основі вибору виконується пошук по базі даних серед готових конфігурацій. Пошук виконується серед тих конфігурацій, які входять до заданого цінового діапазону. Користувачу надається топ 5 готових до використання комп'ютерів, комплектуючі яких в загальному обсязі мають більшість тих позначок, які напередодні обрав користувач. Також обираються ті комп'ютери, які більше разів були придбані іншими користувачами та мають кращий рейтинг

по відгукам.

Під час створювання власною конфігурації, система реагує на вибір користувача та блокує можливість обирання товару, який є не сумісним з поточною конфігурацією. Наприклад при виборі материнської плати від AMD користувачу заблокується можливість обирання процесорів на базі Intel, або блок живлення повинен мати той же стандарт, як і корпус та ін. Також, якщо користувач вибере не повний комплект комплектуючих для власної конфігурації, система заблокує можливість придбати товар, як повну конфігурацію, тому для користувача буде тільки наявна можливість придбання комплектуючих, як окремих частин.

6 АНАЛІЗ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ

Для впевненості в обраному методі побудови рекомендаційної системи було проведено експеримент роботи алгоритму надання рекомендації враховуючи усі необхідні вхідні дані, які необхідні для розрахунків.

Мета експерименту полягає в переконанні, щодо обраного методу побудови рекомендаційної системи, а також перевірка його працездатності в умовах даного проекту і виявлення його переваг та недоліків[18].

Перед виконанням алгоритму необхідно створити словник позначок, які відповідають за напрямок використання того чи іншого комплектуючого. За цими позначками буде відбуватись розрахування TF-IDF, для надання ваги кожній із позначок, це буде необхідно для подальшого обчислення. Змоделюємо ситуацію, що користувач зробив взаємодію з товаром, який наявний в базі даних. За приклад створимо таблицю, в якій запишемо Id готової конфігурації, яка містить в собі набір комплектуючих з кількістю певних позначок напрямку використання.

Таблиця товарів та позначок

Id товару	Для киберспорту	Для офіса	Гібрид
1	1	7	0
2	8	0	9
3	2	4	6
4	0	3	8

Рисунок 13 – Приклад таблиці з товарами та кількістю позначок різного напрямку, які вони містять.

Припустимо, що користувач переглянув перший товар, після цього

алгоритм починає розрахунок TF-IDF цього товару та інших товарів в системі. Отримані дані перетворюються в векторну форму, де записано кожну вагу тієї чи іншої позначки. Згідно таблиці вище, отримуємо наступні дані TF-IDF товарів у векторній формі.

$$\begin{aligned} \text{TF-IDF}_1 &= (0,01, 0,10, 0) \\ \text{TF-IDF}_2 &= (0,05, 0, 0,06) \\ \text{TF-IDF}_3 &= (0,01, 0,03, 0,06) \\ \text{TF-IDF}_4 &= (0, 0,03, 0,08) \end{aligned}$$

Рисунок 14 – TF-IDF товарів у векторній формі

Після отримання ваги кожної позначки і отримання їх у векторній формі виконується розрахунок схожості товарів між собою за методом косинусної схожості. Цей метод допоможе знайти коефіцієнт схожості між товаром, який був переглянутий користувачем під час роботи з сайтом та іншими товарами наявними в базі даних.

$$\begin{aligned} \text{Similarity}_{1/2} &= \frac{0,0005}{0,007} = 0,07 \\ \text{Similarity}_{1/3} &= \frac{0,0031}{0,006} = 0,51 \\ \text{Similarity}_{1/4} &= \frac{0,003}{0,008} = 0,35 \end{aligned}$$

Рисунок 15 – Отриманні коефіцієнти схожості за методом косинусної схожості

Згідно розрахунків ми маємо те, що товар №1 має коефіцієнт схожості 0,07 з товаром № 2 і це означає, що товари зовсім не схожі. Тому як оцінка виконується в межах від 0 до 1, де 0 зовсім не схожі, а 1 ідентичні. З товарами №3 і №4 оцінка набагато вища, і якщо обирати 1 товар для рекомендації, то звісно користувачу буде надана рекомендація про товар №3, тому як він має найвищий показник схожості.

Отже провівши експеримент роботи алгоритму в рамках цього проекту, можна зробити висновок, що алгоритм добре підходить для рекомендації товару в інтернет-магазині для продажу комп'ютерів. Метод контентної фільтрації добре підходить для даної роботи, тому що рекомендаційний алгоритм базується на інформації про товари. Метод спільної фільтрації орієнтований на інших користувачів і тому в даному випадку буде надавати менш релевантні рекомендації.

Головною перевагою контентної фільтрації в рамках цього дослідження є те, що користувачі маючи навіть єдину дію в системі можуть отримати досить точну рекомендацію, яка в той же час буде релевантною для користувача.

Недоліками даної реалізації алгоритму є те, що він має складність отримання інформації. Тому, як для кожного повторного порівняння алгоритму знову необхідно робити розрахунки, це буде навантажувати систему і тому продуктивність роботи буде знижена. Також алгоритм контентної фільтрації в більшому ступені залежить від професійних знань в галузі обробки інформації на відміну від спільної фільтрації.

Для удосконалення ефективності роботи алгоритму, можна створити більш складну систему пошуку інформації про товари, яка буде більш оптимізованою та зменшить кількість розрахунків, що збільшить продуктивність, як даного алгоритму так і систему, в яку він вбудований.

7 ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ

Досліджувані методи побудови рекомендаційної системи в рамках цієї роботи можуть мати практичне використання у сфері інтернет-продажу і буде гарним рішенням для використання у галузі персональних комп'ютерів. Ця галузь є досить складною для звичайного користувача, щоб вільно орієнтуватись серед безлічі різноманітних характеристик комплектуючих. Також попит на придбання персональних комп'ютерів буде залишатися на високому рівні ще досить довго, тому як майже кожна сфера діяльності людини вимагає наявності комп'ютера.

Також це дослідження має досить багато напрямків для розвитку і має високий потенціал для створення великої системи рекомендацій, які вирішать майже усі проблеми, з якими зіштовхується користувач при придбанні тієї чи іншої конфігурації персонального комп'ютера, або окремого комплектуючого до власного пристрою.

Одним із напрямків удосконалення досліджуваної рекомендаційної системи, є вирішення питання оптимізації розрахунків для зниження навантаження на систему, а також прискорення рекомендаційного алгоритму. Також у якості розширення можливостей системи, є створення універсальної системи, яка буде рекомендувати не тільки готові конфігурації комп'ютерів, але і окремих комплектуючих. Ще одною з опцій використання цього методу можна знайти у налаштуванні можливостей машинного навчання, наприклад для передбачення певним користувачем придбання того чи іншого товару. Це допоможе більше зрозуміти попит певних товарів, а також збільшити кількість продажів.

ВИСНОВОК

У результаті роботи були проаналізовані та досліджені методи побудови рекомендаційної системи для продажу товарів в інтернет-просторі. В ході дослідження були розглянуті основні методи побудови рекомендаційних систем, це спільна фільтрація та контентна фільтрація. Також були досліджені особливості кожного з методів і на основі аналізу вирішено використовувати саме метод контентної фільтрації, який базується на пошуку схожих товарів для рекомендації їх користувачам, аналізуючи вміст кожного з товарів. Також було описано додаткові можливості рекомендацій готових конфігурацій комп'ютерів на основі напрямків використання зазначених користувачем.

Було проведено аналіз предметної галузі та виявлено проблемні елементи в існуючих аналогах. На основі аналізу було прийнято рішення створити рекомендаційну систему з використанням методу контентної фільтрації, яка зможе надавати релевантні рекомендації, щодо існуючих в інтернет-магазині готових до роботи конфігурацій. допоможе користувачу придбати комп'ютер, який буде підходити до бажань клієнта, а також створити за її допомогою власну конфігурацію комп'ютера, яка буде більш точно відповідати до критеріїв користувача, щодо використання персонального комп'ютера також місцем для розташування рекомендаційної системи було рішення створити інтернет-магазин, який буде спеціалізуватись на продажі персональних комп'ютерів та комплектуючих.

Було поставлено низку задач, які повинні бути реалізовані для вирішення сучасних проблем в галузі. Також на основі аналізу предметної галузі було сформовано низку функціональних та нефункціональних вимог.

Було досліджено наукову літературу та існуючі наукові публікації для більшого розуміння принципів роботи алгоритмів рекомендаційних систем. В ході роботи з науковими матеріалами було зрозуміло, що використання методу контентної фільтрації більше підходить для досліджуваної галузі.

Було спроектовано архітектуру програмного забезпечення в ході, якого було вирішено використати класичну багаторівневу архітектуру програмного забезпечення для більш зручного та безпечного користування та редагування окремих його рівнів. І також було спроектовано за допомогою UML, діаграму розгортання системи, яка дає змогу зрозуміти, як буде виконуватись зв'язок між частинами програмного забезпечення і діаграму варіантів використання, яка описує можливість поведінки користувача у системі.

У роботі було досліджено алгоритм роботи контентної фільтрації, а також проведено експеримент метою, якого було впевненість у вибраному алгоритмі, а також доцільність та загалом можливість використання для рекомендації персональних комп'ютерів. Були виявлені переваги алгоритму, а також недоліки, які будуть спостерігатись у перших версіях програмної системи. Також були описані програмні рішення, які необхідні для створення автоматичної рекомендаційної системи на основі поведінки користувача у ній, а також додаткові можливості рекомендацій, які необхідні для вирішення можливих проблем, які виникають у користувачів при виборі комп'ютера або створення власної конфігурації системи.

Були описані можливості подальшого удосконалення рекомендаційної системи шляхом оптимізації та розширювання функціональних можливостей, зокрема використання можливостей машинного навчання. А також зазначено, що запропонований алгоритм буде більш ефективним для рекомендації товарів в галузі персональних комп'ютерів.

ПЕРЕЛІК ДжЕРЕЛ ПОСИЛАННЯ

1. Продаж персональних комп'ютерів на світовому ринку – Gartner [Електронний ресурс] – URL: <https://www.gartner.com> (дата звернення: 10.02.2023).
2. Recommender System [Електронний ресурс] – URL: <https://learn.microsoft.com/en-us/sql/database-engine/configure-windows/sql-server-express-localdb?view=sql-server-ver16> (дата звернення: 02.04.2023).
3. Троелсен Е. С# і платформа .NET. Бібліотека програміста [Текст] / Е. Троелсен – Київ. Діалектика, 2014. – 1310 с.
4. Sql Server Express LocalDB [Електронний ресурс] – URL: <https://learn.microsoft.com/en-us/sql/database-engine/configure-windows/sql-server-express-localdb?view=sql-server-ver16> (дата звернення: 10.04.2023).
5. Prem Melville, Vikas Sindhvani. Recommender Systems. Encyclopedia of Machine Learning Chapter No: 00338 Page Proof Page 1 22-4-2010. URL: <https://www.prem-melville.com/publications/recommender-systems-eml2010.pdf> (дата звернення: 8.04.2023).
6. Chalyi, S., Leshchynskiy, V. “Temporal modeling of user preferences in recommender system”. CEUR Workshop Proceedings, 2020, 2711, с. 518-528
7. Xinyi Wu. Comparison Between Collaborative Filtering and Content-Based Filtering. URL : https://www.researchgate.net/publication/365494125_Comparison_Between_Collaborative_Filtering_and_Content-Based_Filtering (дата звернення: 4.04.2023).
8. Chalyi, S., Leshchynskiy, V., Leshchynska, I. Method of forming recommendations using temporal constraints in a situation of cyclic cold start of the recommender system. EUREKA, Physics and Engineering, 2019, 2019(4), с. 34-40
9. Bendito Freitas Riberiro. Content based filtering [Електронний ресурс] - URL: <https://www.slideshare.net/BenditoFreitasRibeir/content-based-filtering> (дата звернення: 03.04.2023).

10. A Guide to Content-Based Filtering In Recommender Systems [Електронний ресурс] – URL: <https://www.turing.com/kb/content-based-filtering-in-recommender-systems> (дата звернення: 03.04.2023).
11. Document-term matrix [Електронний ресурс] – URL: https://en.wikipedia.org/wiki/Document-term_matrix (дата звернення 14.04.2023).
12. Cosine similarity [Електронний ресурс] – URL: https://en.wikipedia.org/wiki/Cosine_similarity (дата звернення: 05.04.2023).
13. Carl-Hugo Marcotte. An Atypical ASP.NET Core 6 Design Patterns Guide: A SOLID adventure into architectural principles and design patterns using .NET 6 and C# 10, 2nd Edition. Packt Publishing, 2022 – 678 с.
14. Діго С.М. Бази даних: проектування та використання [Текст]: навч. / С. М. Діго. – М.: «Фінанси і статистика», 2005 – 518 с.
15. Unified Modeling Language [Електронний ресурс] – URL: <https://www.uml.org/> (дата звернення: 21.03.2023).
16. Фаулер, М. UML. Основы: пер. з англ. – СПб.: Символ, 2006. – 184 с.
17. C# documentation [Електронний ресурс] – URL: <https://learn.microsoft.com/uk-ua/dotnet/csharp/> (дата звернення: 15.04.2023).
18. Tan Pengshi Alvin. A Content-Based Recommender for E-Commerce Web Store. Toward Data Science, 2020 [Електронний ресурс] – URL: <https://towardsdatascience.com/a-content-based-recommender-for-e-commerce-web-store-7554b5b73eac> (дата звернення: 08.04.2023)