

Міністерство освіти та науки України
Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____

Кафедра _____ Системотехніки _____

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

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

«Дослідження методів підбору кандидатів в системах надання послуг із
урахуванням уподобань клієнтів»
(тема)

Виконала:

студентка 6 курсу групи _____ ІТПм21-1 _____

_____ Щетініна А.В. _____

Спеціальності _____ 122 Комп'ютерні науки _____
(код і повна назва спеціальності)

Тип програми _____ освітньо-професійна _____

Освітня програма Інформаційні технології проектування
(повна назва освітньої програми)

Керівник _____ проф. Колесник Л.В. _____
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ Гребінник І.В. _____
(підпис) (прізвище, ініціали)

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

14.12.2022




Щетініна А.В.

Кваліфікаційна робота не містить відомостей заборонених до відкритого опублікування.

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

Попередній захист проведено 14 грудня 2022 р.

Керівник кваліфікаційної роботи  проф. Колесник Л.В.

Харківський національний університет радіоелектроніки

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

ЗАТВЕРДЖУЮ:

Зав. Кафедри _____
(підпис)

«21» листопада 2022 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

студенці _____ Щетініній Анастасії Володимирівні _____
 (прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів підбору кандидатів в системах надання послуг із урахуванням уподобань клієнтів
затверджена наказом по університету від 21 листопада 2022 р. № 1263СТ
2. Термін подання студентом роботи до екзаменаційної комісії 20 грудня 2022р
3. Вихідні дані до роботи Дослідження методів підбору в рекомендаційних системах
4. Перелік питань, що потрібно опрацювати в роботі 4.1 Вступ. 4.2 Аналіз переметної області. 4.3 Огляд методів та технологій, які застосовуються в предметній області. 4.4 Математичний опис задачі. 4.5 Визначення вимог до розроблюваної системи. 4.6 Логічне і фізичне моделювання даних. 4.7 Розробка алгоритму підбору кандидатів. 4.8 Висновки. 4.9 Перелік джерел посилання.
5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) 5.1 Організаційна структура агенції знайомств. 5.2 Алгоритм організації побачення зі сторони користувача. 5.3 Контекстна діаграма основного бізнес-процесу. 5.4 Діаграма декомпозиції основного бізнес-процесу. 5.5 Діаграма дерева вузлів. 5.6 Діаграма прецедентів. 5.7 Логічна схема даних. 5.8 Фізична схема даних. 5.1 Алгоритм послідовної поведінки користувача. 5.9 Схема роботи рекомендаційної системи на основі

колабораційної фільтрації. 5.10 Схема роботи рекомендаційного системи при «холодному старті». 5.11-13 Модернізований алгоритм колаборативної фільтрації з використання середньозваженого значення

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

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

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів	Примітка
1.	Отримання завдання кваліфікаційної роботи	03.10.2022	<i>Виконано</i>
2	Аналіз предметної області та постановка задачі	11.10.2022	<i>Виконано</i>
3	Огляд методів та технологій, які застосовуються в предметній області	25.10.2022	<i>Виконано</i>
4	Аналіз існуючих аналогів	03.11.2022	<i>Виконано</i>
5	Огляд різних методів для підбору кандидатів	08.11.2022	<i>Виконано</i>
6	Математичний опис задачі	16.11.2022	<i>Виконано</i>
7	Визначення вимог до розроблювальної системи	20.11.2022	<i>Виконано</i>
8	Розробка проектних рішень по програмному та інформаційному забезпеченню системи	20.11.2022	<i>Виконано</i>
9	Оформлення пояснювальної записки	10.12.2022	<i>Виконано</i>
10	Подача кваліфікаційної роботи на допуск до захисту	18.12.2022	<i>Виконано</i>
11	Підготовка доповіді до захисту роботи	19.12.2022	<i>Виконано</i>
12	Подання кваліфікаційної роботи	20.12.2022	<i>Виконано</i>

Дата видачі завдання 3 жовтня 2022 р.

Студентка _____ *Щетініна А.В.* _____
 (підпис) (посада, прізвище, ініціали)

Керівник роботи _____ *Колесник Л.В.* _____
 (підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи: 116 с., 1 табл., 18 рис., 4 додатка, 26 джерел інформації. Графічний матеріал кваліфікаційної роботи містить 11 арк.

АВТОМАТИЗАЦІЯ, АНКЕТУВАННЯ, БАЗА ДАНИХ, ІНФОРМАЦІЙНА СИСТЕМА, КОРИСТУВАЦЬКИЙ ІНТЕРФЕЙС, КОРИСТУВАЧ, РЕКОМЕНДАЦІЙНА СИСТЕМА

Об'єктом дослідження кваліфікаційної роботи є один з основних бізнес-процесів агенції знайомств під назвою «Blind date», який полягає в автоматизації процесу підбору пари для потенційного знайомства та організація побачення на основі анкетування (психологічного тестування).

Предметом досліджень є альтернативні методи підбору кандидатів в системах надання послуг (рекомендаційних системах) з урахуванням уподобань клієнтів, що дозволяє автоматизувати основний бізнес-процес агенції знайомств.

Метою дослідження є підвищення ефективності процедури підбору кандидатів в системах надання послуг за рахунок урахування уподобань клієнтів.

Методами дослідження є наступні обрані методи: алгоритми рекомендаційних систем, колаборативна фільтрація, власна модифікаційна колаборативна фільтрація для обраної предметної області.

У ході написання кваліфікаційної роботи були проведені наступні етапи: детальний аналіз обраної предметної області, дослідження альтернативних методів підбору кандидатів в системах надання послуг з урахуванням уподобань клієнтів, розробка модифікованого методу колаборативної фільтрації, функціональне моделювання логічне і фізичне моделювання даних, практична реалізація інформаційної системи.

Область застосування – автоматизація роботи агенції знайомств під назвою «Blind date».

Результатами работ є: проаналізовані наявні аналоги дейтінгових застосунків та їх недоліки, проаналізовані сучасні методи підбору кандидатів, що використовуються в рекомендаційних системах, модифікація вже існуючого методу, а саме – колаборативної фільтрації, та його практична реалізація.

ABSTRACT

Explanatory note to the certification work: 116 pages, 1 table, 18 figures, 4 appendices, 26 sources of information. The graphic material of the attestation work contains 11 sheets.

AUTOMATION, QUESTIONNAIRE, DATA BASE, INFORMATION SYSTEM, USER INTERFACE, USER, RECOMMENDATION SYSTEM

The research object of the qualification work is one of the main business processes of the dating agency called "Blind date", which consists in automating the process of selecting a couple for a potential acquaintance and organizing a meeting based on a questionnaire (psychological testing).

The subject of research is alternative methods of selecting candidates in service provision systems (recommendation systems) taking into account customer preferences, which allows automating the main business process of a dating agency.

The goal is to research and analyze alternative methods of selecting candidates in service delivery systems, taking into account customer preferences and choosing the best of the alternatives for the chosen subject area.

The research methods are the following selected methods: recommender system algorithms, collaborative filtering, own modification collaborative filtering for the selected subject area.

During the writing of the qualification work, the following stages were carried out: detailed analysis of the selected subject area, research of alternative methods of candidate selection in service provision systems taking into account customer preferences, development of modified collaborative filtering, functional modeling, logical and physical modeling of data, practical implementation of the information system.

The field of application is the automation of the work of the dating agency called "Blind date".

The results of the work are: analyzed existing analogues of dating applications and their shortcomings, analyzed modern candidate selection algorithms used in recommendation systems, modification of an already existing algorithm, namely collaborative filtering, practical implementation of the chosen algorithm.

ЗМІСТ

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ.....	10
ГЛОСАРІЙ.....	11
ВСТУП.....	12
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ.....	13
1.1 Організаційна структура обраного підприємства.....	14
1.2 Аналіз аналогічних систем.....	16
1.3 Визначення сфери застосування розроблювальної системи.....	19
1.4 Визначення основних бізнес-процесів, що вимагають автоматизації.....	21
1.5 Постановка задачі для автоматизування.....	23
2 ОГЛЯД МЕТОДІВ ТА ТЕХНОЛОГІЙ, ЯКІ ЗАСТОСОВУЮТЬСЯ В ПРЕДМЕТНІЙ ОБЛАСТІ.....	26
2.1 Огляд різних методів для підбору кандидатів.....	26
2.1.1 Алгоритм сортування.....	27
2.1.2 Рекомендаційні системи. Колаборативна фільтрація.....	28
2.2 Огляд технологій для практичної реалізації проекту.....	31
2.2.1 Обґрунтування вибору СКБД.....	32
2.2.2 Обґрунтування вибору мови програмування.....	33
3 МАТЕМАТИЧНИЙ ОПИС ЗАДАЧІ.....	35
3.1 Колаборативна фільтрація.....	35
3.2 Модифікація методу колаборативної фільтрації з використанням середньозваженого значення.....	38
4 ВИЗНАЧЕННЯ ВИМОГ ДО РОЗРОБЛЮВАНОЇ СИСТЕМИ.....	41
4.1 Визначення системних вимог до розроблювальної інформаційної системи.....	41
4.2 Визначення функціональних вимог до розроблювальної системи.....	41
4.3 Визначення вимог до інтерфейсу розроблювальної системи.....	45
4.4 Розробка діаграми варіантів використання системи.....	48
5 ЛОГІЧНЕ І ФІЗИЧНЕ МОДЕЛЮВАННЯ ДАНИХ.....	50

5.1 Логічне моделювання даних.....	50
5.2 Фізичне моделювання даних.....	54
6 РОЗРОБКА АЛГОРИТМУ ПІДБОРУ КАНДИДАТІВ.....	59
6.1 Загальний алгоритм роботи застосунку.....	59
6.2 Алгоритм підбору кандидатів.....	61
ВИСНОВКИ.....	65
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....	66
ДОДАТОК А Графічний матеріал кваліфікаційної роботи.....	69
ДОДАТОК Б «Текст програми».....	84
ДОДАТОК В «Інструкція користувача».....	100
1. ПРИЗНАЧЕННЯ ТА УМОВИ ЗАСТОСУВАННЯ.....	104
2. ПІДГОТОВКА ДО РОБОТИ.....	105
3. ОПИС ОПЕРАЦІЙ.....	108
4. АВАРІЙНІ СИТУАЦІЇ.....	114
ДОДАТОК Г «Відомість кваліфікаційної роботи».....	115

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

БД – база даних.

ІС – інформаційна система.

ПК – персональний комп'ютер.

СКБД – система керування базами даних.

ЦА – цільова аудиторія.

SQL – Structured query language — мова структурованих запитів.

ER-діаграма – entity-relationship diagram – графічна реалізація відношення «сутність-зв'язок».

CEO – chief executive officer.

ГЛОСАРІЙ

Атрибут – невід’ємна властивість або частина, яка є необхідною для забезпечення цілісності об’єкта або суб’єкта.

Анкетування – психологічний тест, який проходить кожен користувач системи під час реєстрації, на результатах аналізу якого підбирається пара та організується побачення.

Згода – функція системи, за якою користувач має можливість дати згоду на організацію побачення.

Інтерфейс користувача – сукупність графічних елементів (іконки, меню та інші) та методів для взаємодії користувача та системи.

Інформаційна система – сукупність технічних засобів збереження та обробки інформації для забезпечення потреб користувачів.

Рекомендаційна система – підклас системи фільтрації, що будує рейтинговий перелік об’єктів (в даному випадку профілів кандидатів), яким користувач може надати перевагу.

Тип даних – це характеристика об’єкту, яку надано у явному чи неявному вигляді.

«Blind date» – назва агенції, що надає послуги організації побачень.

ВСТУП

З початком глобального міжнародного карантину у лютому 2020, сучасна молодь зрозуміла, наскільки ж для нас важливий соціум. Коли зустрічі в Zoom та Skype вже набридлі, та карантин стали послабляти, багато з нас намагалися компенсувати витрачений вдома час та повернутися до попереднього життя.

Але тут люди зустрілися з двома проблемами: по-перше, більша частина студентів та працівників вирішили не виходити з дистанційного режиму. А, по-друге, ми забули як знайомитися у реальному житті. Це означає, що шукати нових знайомих, а тим паче потенційного партнера для романтичних стосунків стало ще складніше.

Багатогадинні пошуки не приносять бажаного результату і тільки ще більше демотивують користувачів дейтінгових застосунків. Щоб мінімізувати час, який витрачається на пошук гідного партнера та організацію побачення, та максимізувати «реальне» якісне спілкування, набагато зручніше звернутися до застосунку, який підбере тобі партнера, що має схожі інтереси, та, базуючись на цих інтересах, організує справжню незабутню зустріч.

Основною ідеєю є надання сервісу користувачам, що дозволить їм отримати вибірку кандидатів, які відповідають заданим критеріям та допоможе організувати побачення за допомогою веб застосунку.

Актуальним завданням є створення унікального застосунку для автоматизації роботи агенції знайомств «Blind date», основним бізнес-процесом якого є організація побачення для клієнтів.

Метою дослідження є підвищення ефективності процедури підбору кандидатів в системах надання послуг за рахунок урахування уподобань клієнтів.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

В якості предметної області для дослідження була обрана агенція знайомств під назвою «Blind date». Основна особливість цієї агенції, що відрізняє її на фоні інших дейтінгових застосунків, полягає в тому, що користувач або користувачка можуть не тільки знайти вдалий «match», а і в результаті отримати повністю організоване побачення.

Основний бізнес-флоу складається з таких етапів: спочатку користувач або користувачка реєструється в веб-застосунку, проходить психологічне тестування, на основі якого буде відбуватися підбір партнерів для потенційного знайомства. Після цього, автоматично за допомогою системи узгоджується дата майбутнього побачення, далі агенція займається організацією побачення і надсилає через застосунок обом партнера повідомлення з детальною інформацією про побачення.

Головна ідея даного стартапу – допомогти людям знайти партнера за допомогою застосування психологічного тестування. Усі відомі аналоги сортують кандидатів лише за локацією або за уподобанням фотографій. Але нерідко відбувається так, що наші очікування не відповідають реальності. Також усі ми всі з дитинства знаємо, що не варто судити людей за зовнішнім видом, тому що найважливіше знаходиться «всередині».

Прибуток від цієї ідеї планується отримувати двома шляхами. Перший – отримання прибутку шляхом продажу різних підписок. Тобто, кожен користувач має оплачувати місячну підписку за використання веб застосунку та послуг агенції (звісно ж має сенс зробити декілька тарифних планів за різною вартістю, які будуть включати в себе різний доступний функціонал). А другий – продаж франшизи, який буде містити допомогу з відкриттям нових офісів, маркетингом, пошук локальних партнерів і т.д.

Логічніше буде розпочати свою роботу спочатку в великих містах України, а потім продовжити масштабуватися та відкривати нові філіали в інших містах країни та Європи шляхом продажу франшиз.

За законодавством України, агенція «Blind date» буде обслуговувати лише повнолітніх, тобто тих користувачів, яким вже є 18 років. Для перевірки реального віку користувачів, планується інтегрувати можливість авторизації за допомогою будь-якого застосунку онлайн банкінгу або через усім відомий застосунок «Дія». Застосунок «Дія» вже давно використовується на законодавчому рівні як офіційний е-документ та є альтернативою фізичних паперових документів.

Такий варіант авторизації допоможе користувачам веб-застосунку та працівникам самої агенції не хвилюватися про безпеку. Цей факт 100-відсотково допоможе швидко популяризувати застосунок серед молоді, особливо серед жіночої половини населення.

Також, такий варіант «безпечного використання» стане додатковою перевагою, що відпилятиме «Blind date» від існуючих сучасних аналогів.

1.1 Організаційна структура обраного підприємства

Ви знаєте багато історій про підроблених профілів та фейкові сайти знайомств, які використовуються для проведення недобросовісної практики. Щоб уникнути цього, агенція «Blind date» вибрала не тільки сучасний веб-застосунок, де кожен користувач має можливості для вибору своєї «половинки», але й гарну команду. Це допомагає мінімізувати ризики, та зробити шанси на успішне знайомство з майбутнім партнером значно вищими.

Агенція знайомств базується на принципі підбору кандидатів за допомогою «тестування» і містить все, що необхідно для надання якісного та надійного сервісу. Якщо ця послуга буде популяризуватися у всьому світі – це дозволить заощадити багато часу та енергії людям, які хочуть знайти своє кохання.

Завдяки великій команді, процес знайомств є максимально простим, швидким, якісним, але найголовніше – безпечним для обох партнерів та команди.

Далі наведена організаційна структура агенції знайомств (рис. 1.1).



Рисунок 1.1 – Організаційна структура агенції знайомств «Blind date»

Верхівкою даної організаційної структури є саме власник агенції «Blind date». Власнику підпорядковуються HR-менеджер, фінансовий відділ та CEO, тобто генеральний директор агенції. Власник не бере участь в основній діяльності агенції, тільки в операційних задачах, є відповідальним за стратегію розвитку та маркетингу і т.д.

Фінансовий відділ на чолі з головним бухгалтером є відповідальними за бухгалтерську діяльність агенції в цілому. Розрахунковий відділ є відповідальним за нарахування заробітної плати усім співробітникам агенції. Комерційний відділ відповідає за розрахунок клієнтів, партнерів і основні витрати агенції;

HR-менеджер є відповідальним за процес пошуку та найму працівників агенції. В зону відповідальності HR-менеджера також входить підтримка корпоративної культури агенції та налагодження стосунків між співробітниками, згладжування конфліктних ситуацій в колективі.

CEO агенції є відповідальним за роботу та результат наступних відділів:

- 1) Відділ маркетингу – відділ, що відповідальний за наступні функції:
 - просування бренду у соціальних мережах;
 - вивчення попиту та тенденцій розвитку ринку;
 - вивчення різних каналів та методів продажу;

- планування іміджу агенції.
- 2) Організаційний відділ є відповідальним за координації для створення ідеального побачення для усіх клієнтів агенції. Сюди входить:
- робота з туроператорами, які займаються бронюванням та купівлею квитків для іноземних клієнтів;
 - роботам з перекладачами, які можуть супроводжувати клієнтів під час побачення та які допомагають створювати контент для відділу маркетингу агенції;
 - робота з організаторами побачень, які займаються організацією самих побачень, включно з підбором концепції, локації, виду проведення дозвілля і т.д.
- 3) Відділ комунікації – відділ, який відповідальний за комунікацію з партнерами. Спершу call-центр займається так званими «холодними» дзвінками та встановлюють контакт з партнерами, які потенційно заінтересовані в спільній співпраці з агенцією. Після первинного контакту, call-центр передає ці контакти чат-менеджерам, які продовжують спілкування щодо потенційного партнерства через корпоративну пошту агенції.
- 4) Юридичний відділ – відділ, який відповідальний за дотриманням законності оформлення документів, також слідкує за укладенням договорів з новими партнерами, складанням усіх юридичних документів і т.д. Також, за вимогою, можуть зайнятися роботою з претензіями або подачею судових позовів.

1.2 Аналіз аналогічних систем

Популярні схожі застосунки для знайомств, такі як Tinder, Badoo або Bumble, хоча і є дуже відомими, але ж все таки мають свої недоліки. По-перше, вони зовсім не підходять для автоматизації роботи нашої агенції, та можуть використовуватися тільки «соло». Тобто, цього вистачить для знайомства с кимось, але зовсім не вирішує проблему зі скороченням часу, проведених за пошуком, вдалого «match-у». По-друге, ці застосунки ніяк не допоможуть користувачам з організацією самого побачення. Далі будуть наведені недоліки кожної з аналогічних дейтінгових систем [1].

Дуже відомий та популярний на території країн СНГ, Badoo – веб- та мобільний стосунок для знайомств, що базується на підборі кандидатів лише за допомогою локації. Користувач може лише ознайомитися зі стислим описом у профілі та парі фотографій. Головний недолік даного аналогу – слабкий алгоритм підбору кандидатів, оснований лише на геотегах та відсутність будь-якого анкетування для подальшої фільтрації.



Рисунок 1.2 – Графічний інтерфейс веб-застосунку «Badoo»

Tinder є дуже схожим застосунком до попереднього Badoo. Відмінною є те, що розпочати спілкування можна лише за умовою взаємної симпатії. Розробники все частіше додають різні тематичні анкетування, що дозволяють «якісніше» підбирати вибірку з кандидатів. Головний недолік – наявність великої кількості реклами, що може дуже дратівливою для користувачів, все же слабкий алгоритм підбору кандидатів.



Рисунок 1.3 – Графічний інтерфейс мобільного застосунку «Tinder»

Bumble є відносно новим та менш відомим аналогом Tinder та Badoo. Тут вже додається функція хештегування, які відображають основні інтереси та хобі користувачів. Ці хештеги використовуються при підборі кандидатів, що роблять алгоритм більш унікальним і в результаті діє більш «якісну» вибірку кандидатів.

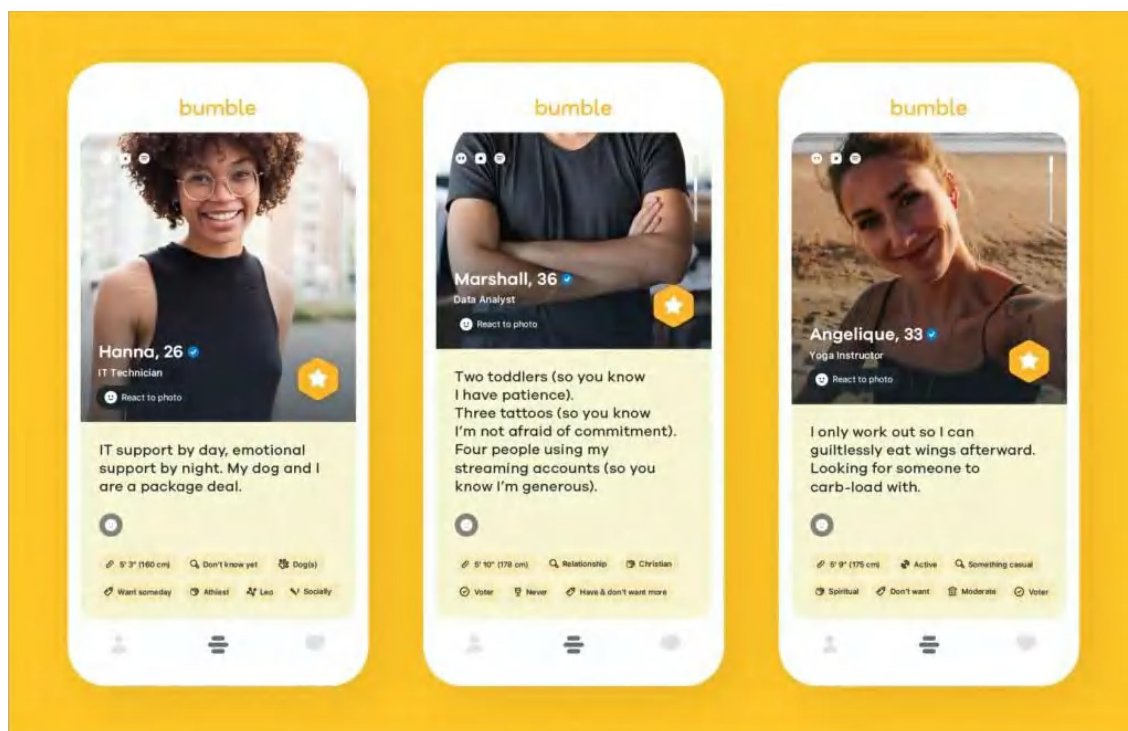


Рисунок 1.4 – Графічний інтерфейс мобільного застосунку «Teamo»

Після аналізу існуючих дейтінгових аналогів можна побачити, що навіть дуже відомі застосунки не зовсім задовольняють вимоги сучасної молоді. Це люди, які справді хочуть знайти гідного партнера для побудови взаємовідносин, але не хочуть витратити тисячі годин, на пошук в соціальних мережах. До списку вже наведених вище недоліків наявних застосунків можна додати ще той факт, що дуже популярною є тема шахрування, яку можна зустріти «онлайн» під час знайомства, причому у різному вигляді: фішинг, несправжні профілі, шантажування, тощо.

Саме для усунення цих «слабких місць», і була створена концепція агенції «Blind date». Як результат цієї кваліфікаційної роботи повинні отримати повністю спроектований та розроблений веб-застосунок, який автоматизує процес організації справжнього побачення. Пошук та підбір відбувається автоматично та базується на результатах психологічного тестування. Для вибору найефективнішого методу підбору кандидатів в системах, далі, в другому розділі, будуть проаналізовані декілька альтернативних методів підбору в рекомендаційних системах.

1.3 Визначення сфери застосування розроблювальної системи

Система, що була обрана для проектування, є унікальною в своєму роді, і зовсім не підходить для використання іншими агенціями або для самостійного використання. Майбутній застосунок проектується на базі унікального бізнес-флоу саме обраної агенції під назвою «Blind-date».

В результаті аналізу аналогічних інформаційних систем або агенцій, було виявлено, що 100 відсоткових аналогів не існує. Наявні лише застосунки, з частково схожим функціоналом.

Допускається використання застосунку іншими агенціями, але за умови покупки вже згаданої вище франшизи. У цьому випадку, у нового філіалу буде така ж назва та такі ж самі бізнес-процеси, але буде зовсім інший власник.

Цей підхід називається «франчайзинг». Він є доволі популярною практикою серед великих брендів та корпорацій. І навіть в Україні з кожним роком стає все більш використовуваним. Наприклад, усім відомі українські франшизи як «Арома кава», «Нова пошта» та інші.

До світового аналогу саме у сфера знайомств можна віднести міжнародну шлюбну агенцію «MORDINSON». Дана агенція фокусується на пошуку так званих «наречених» для іноземців. Це ще раз доводить, що проблема пошуку «другої половинки» є досі дуже актуальною та потребує більш сучасних методів для вирішення.

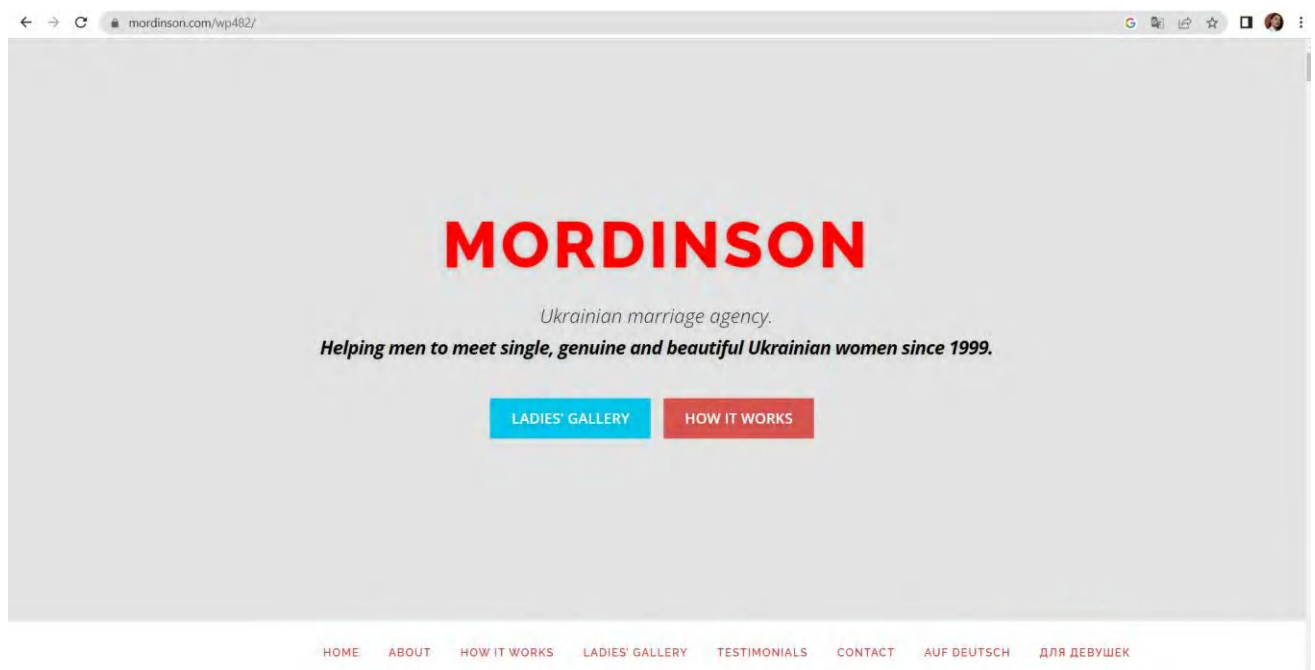


Рисунок 1.5 – Веб-сайт міжнародної шлюбної агенції «MORDINSON»

На відміну від «MORDINSON», агенція «Blind date» не займається пошуком дружин, а лише покриває процес підбору кандидатів для потенційного знайомства та повністю відповідає за організацію побачень. Для автоматизації цих двох основних процесів і створюється застосунок. За допомогою нього планується скоротити час, що люди звичайно витрачають на безглузді переписки у соціальних мережах та альтернативних застосунках, та час, що витрачається для організації справжнього побачення.

1.4 Визначення основних бізнес-процесів, що вимагають автоматизації

Основною задачею для автоматизації є наступні 2 бізнес-функції роботи агенції знайомств: підбір кандидатів та організація побачення.

Основна ідея полягає в тому, що алгоритм підбору кандидатів оснований на результатах аналізу психологічного тестування. Це тестування, яке проходить кожен користувач або користувачка системи під час процесу реєстрації. Даний застосунок виключає усі так звані «випадкові зв'язки» та надає користувачам вибірку з найбільш потенційними для подальшого знайомства профілями.

Після перегляду списку «профілів», що автоматично підбрала система, користувач має можливість відправити запит тим кандидатам, що здаються його найбільш цікавими. Далі наступним етапом є отримання згоди від обох користувачів на організацію побачення. Коли адміністратор отримує згоду обох партнерів, переходимо безпосередньо до процесу організації побачення. Усією організацією займається виключно агенція, а партнери лише отримують СМС-повідомлення з детальної інформацією про побачення.

Для практичної реалізації було виділено 3 режими роботи інформаційної системи:

а) незареєстрований користувач – система спроектована таким чином, що незареєстрований юзер взагалі немає жодного доступу до даних та функцій системи. Єдина функція, яка доступна незареєстрованим користувачам це реєстрація та/або авторизація в системі;

б) зареєстрований користувач – користувач системи, який має доступ для перегляду усіх кандидатів для знайомства, які були автоматично, на основі результатів психологічного анкетування. Також, зареєстрований користувач може переглянути список надісланих та отриманих запитів на побачення. Окрім цього, має доступ до перегляду та редагування даних, що відображаються в особистому кабінеті. Це дуже зручно, коли користувачі вирішують змінити номер телефону або ж місце проживання.

в) адміністратор – має права на перегляд, редагування та видалення інформації з усіх необхідних для його роботи таблиць. Такими таблицями, що містять в собі дані про усі запити на побачення та інші.

Таблиця 1.1 – Перелік статусів можливих типів користувачів

Користувач	Статус користувача	Доступ до даних
Незареєстрований (або неавторизований) користувач	Анонім	Доступ до бази даних повністю відсутній
Зареєстрований (або авторизований) користувач	User	Має доступ до перегляду таблиці Users, Requests
Адміністратор	Admin	Має доступ для перегляду, додавання, редагування та видалення таблиць

Таким чином ми бачимо, що різниця в наведених вище режимах користувачів полягає у різному рівні доступу до даних, що використовуються для роботи застосунку.

Далі наведений список основних бізнес-функцій, доступних для кожного ж типу користувачів:

- а) для незареєстрованого користувача будуть реалізовані такі функції:
 - реєстрація, що включає в себе процес психологічного тестування;
 - авторизація, тобто вхід в систему;
- б) для зареєстрованого користувача будуть реалізовані такі функції:
 - редагування даних в особистому профілі;
 - перегляд списку кандидатів для потенційного знайомства;
 - перегляд особистого профілю кандидата;
 - формування запиту на побачення;

- можливість скасування або підтвердити свою згоду побачення;
- в) для адміністратора будуть реалізовані такі функції:
 - перегляд запитів на побачення;
 - перегляд результатів психологічного тестування;
 - оформлення побачення.

1.5 Постановка задачі для автоматизування

Як вже описано раніше, наявні популярні системи не задовольняють основним потребам обраної цільової аудиторії. Тут зовсім не підходять застосунки, що відтворюють пошук кандидатів лише за геолокацією. Аналогічно не підходять і звичайні дейтінгові додатки для спільної роботи з реальними агенціями з організації побачення.

Звичайні дейтінгові або шлюбні агенції працюють у режимі офлайн. Тобто, мені, як клієнту, потрібно їхати до офісу, різні анкети і т.д. Потім, зазвичай, адміністратор займається підбором потенційних кандидатів мануально.

Сортування усіх анкет займає багато часу та сил, тому для автоматизації обрано саме цей основний бізнес-процес. Додатково автоматизуємо суміжний процес оформлення запиту на побачення та його обробку.

Сформована база кандидатів (тобто усіх користувачів), яка зберігає їх результати психологічного тестування, використовується для автоматизації. Адміністратори системи працюють з тією частиною БД, що зберігає дані про запити на побачення та наявні варіанти для його проведення. Отримуємо, що адміністратора не шукати варіанти самостійно, достатньо лише обрати варіант з переліку усіх можливих варіантів.

Час, збережений за допомогою автоматизації цього процесу, адміністратори витратять на обробку більшого об'єму запитів.

Вхідною інформацією даного алгоритму будуть особисті дані нового користувача та результати тестування. Вихідною – детальна інформація з деталями вже організованого побачення.

Далі наведений алгоритм, що відображає організацію побачення зі сторони користувачів системи (рис. 1.6).

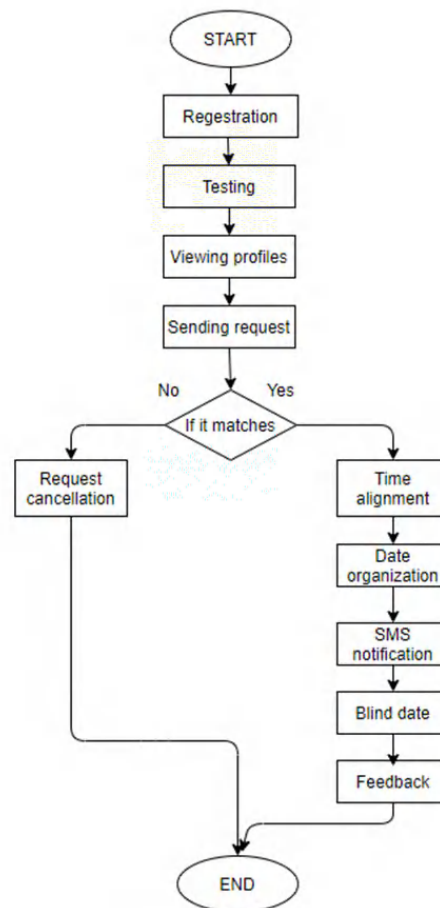


Рисунок 1.6 – Алгоритм організації побачення зі сторони користувача

Наведений алгоритм стосується тільки зареєстрованих та авторизованих користувачів. Для цього обов'язково попередньо необхідно пройти наступні два етапи:

а) registration – стандартна процедура кожної сучасної інформаційної системи, що вимагає від користувача введення особистої інформації, такої як ім'я, прізвище, e-mail і т.д.;

б) testing – головною особливістю застосунку є те, що підбір потенційних кандидатів відбувається на основі психологічного тестування, яке відбувається на попередньому етапі. Тестування базується на концепції «36 питань, щоб закохатися». Деталі цієї концепції будуть розглянуті у наступних розділах;

в) viewing profiles – користувач має можливість переглянути список з 10 кандидатів, які автоматично були обрані системою, базуючись на результатах попереднього тестування;

г) sending request – користувач має можливість відправити запит на побачення тим кандидатам, які сподобалися найбільше;

г) if it matches – перевіряємо, якщо обидва партнери дають згоду на організацію побачення, то переходимо до самої організації. Якщо хоча б один з двох партнерів відмовляється – запит на побачення скасовується;

д) процес організації побачення, що включає в себе time alignment, date organization та sms-notification. Адміністратор час та дату проведення побачення з обома партнерами, після чого обирає варіант для їх побачення. Після чого відправляє обом користувачам смс-повідомлення з деталями побачення;

е) blind date – партнери проходять на «побачення усліпу» в зазначений час, де проводять спільно час, залежно від обраного побачення;

є) feedback – кожен з партнерів має можливість залишити фідбек після свого побачення, для подальшого розвитку агенції.

Для автоматизування обираємо наступні бізнес-функції зі сторони незареєстрованого та зареєстрованого користувачів, а також адміністратора:

а) реєстрація (що включає в себе процес тестування);

б) авторизація;

в) редагування профілю;

г) перегляд сторінки з кандидатами;

г) перегляд персональної сторінки кандидата;

д) перегляд історії запитів на побачення;

е) оформлення запиту на побачення.

2 ОГЛЯД МЕТОДІВ ТА ТЕХНОЛОГІЙ, ЯКІ ЗАСТОСОВУЮТЬСЯ В ПРЕДМЕТНІЙ ОБЛАСТІ

У цьому розділі будуть описані методи підбору кандидатів, а також технології, що обрані для практичної реалізації даної кваліфікаційної роботи.

2.1 Огляд різних методів для підбору кандидатів

Особлива увага була приділена алгоритму підбору потенційних кандидатів для знайомств. За основу була взята стаття відомого журналу «The New York Times» під назвою «The 36 Questions That Lead to Love» [2].

У своїй роботі про сучасне кохання Менді Лен Катрон під назвою «Щоб закохатися в будь-кого, зробіть це» посилається на дослідження психолога Артура Арона (та інших), яке досліджує, чи можна прискорити інтимність між двома незнайомцями, якщо вони запитують один одного конкретний набір особистих питань.

Ідея полягає в тому, що взаємна вразливість сприяє близькості. Цитуючи авторів дослідження, «одним із ключових шаблонів, пов'язаних із розвитком близьких стосунків між однолітками, є стійке, ескалаційне, взаємне особисте саморозкриття». Дозволити собі бути вразливим перед іншою людиною може бути надзвичайно важко, тому ця вправа посилює проблему [2].

На основі цієї статті проведено багато експериментів, частину з них можна знайти навіть на Youtube. Десятки незнайомих людей були поділені на пари. Під час знімального процесу два партнери відповідали на питання, що опубліковані The New York Times, а на при кінці фільмування – вони повинні дивитися в очі один одного протягом 4 хвилин. Як результат – більшість з цих пар дійсно зацікавлені в подальшому спілкуванні зі своїм «незнайомцем».

2.1.1 Алгоритм сортування

Найпростішою ідеєю для підбору кандидатів було використати звичайний найпростіший алгоритм сортування. Саме цей варіант був реалізований в ході написання бакалаврської дипломної роботи у 2020 році.

Алгоритмом сортування називають алгоритм, що розв'язує задачу сортування, іншими словами – здійснює впорядкування масиву елементів [3].

Тобто, у цьому випадку алгоритм підбору списку потенційних партнерів відбувався б наступним чином:

а) під час реєстрація кожен користувач проходить психологічне тестування, яке містить в собі 36 питань. Кожне з цих питань має однаковий набір відповідей, а само «Завжди», «Часто», «Іноді», «Рідко» та «Ніколи». Рішення замінити розгорнуті відповіді та обмежений набір варіантів, викликано це тим, що аналізувати та автоматизувати порівняння розгорнутих відповідей досить складно;

б) для підбору кандидатів, виконується попарне порівняння відповідей, що надали обидва партнери. Кожне з питань має однакову вагу. Кожен варіант відповіді дорівнює певній кількості балів: «Завжди»=1, «Часто»=2, «Іноді»=3, «Рідко»=4 та «Ніколи»=5;

Наприклад, на питання «Я проводжу час зі своїми друзями...» партнер №1 відповідає «Ніколи», що дорівнює 5 балам. А партнер №2 - «Часто», що дорівнює 2 балам. Щоб підрахувати розбіжність наших відповідей, виконує наступні розрахунки:

$$R_{ij}^k = |Q_i^k - Q_j^k|,$$

де R_{ij}^k – розбіжність i -го та j -го кандидата для k питання;

Q_i^k , Q_j^k – відповідь в балах i -го, j -го кандидата для k питання, відповідно.

Тобто, чим менший бал розбіжності R_{ij}^k – тим більше партнери підходять один одному.

в) аналогічно проходимо через усі 36 питань та сумуємо усі бали розбіжності. Отримаємо деяке число N.

г) Після цього, за допомогою сортування бульбашковим методом (насправді можливе використання будь-кого з алгоритмів сортування), знаходимо 10 кандидатів, з найменшим числом розбіжності

$$\sum_{k=1}^n \alpha^k R_{ij}^k \rightarrow \min .$$

Бульбашковим сортуванням, називають простий алгоритм сортування, який багаторазово перебирає вхідний список елемент за елементом, порівнюючи поточний елемент із наступним за ним, міняючи їхні значення, якщо потрібно. Ці проходи крізь масив даних повторюються доти, поки під час проходу не буде виконано жодних заміні, що означає, що список стане повністю відсортованим [4].

Цей алгоритм підбору є доволі простим для розуміння та практичної реалізації, але не дуже ефективний з точки зору якісного підбору кандидатів для знайомств.

2.1.2 Рекомендаційні системи. Колаборативна фільтрація

Далі розглянемо поняття «рекомендаційні системи» та алгоритм «колаборативної фільтрації».

Системою рекомендацій, або рекомендаційною системою (іноді термін «систему» замінюють синонімами «платформа» або «механізм»), називають підкласом системи фільтрації інформації, яка надає пропозиції щодо елементів, які найбільше схожі з уподобаннями конкретного користувача [5].

Рекомендаційну систему зручно використовувати для прискорення процесу прийняття рішень в різних предметних областях. Наприклад, покупки в онлайн-

магазинах, підбір аналогічного аудіо- та відео-контенту (Youtube, Netflix і т.д.). Такі системи особливо корисні, коли людині потрібно вибрати товар чи послугу із величезної кількості аналогів.

Окрім цього, рекомендаційні системи також популярні для певних тем, таких як ресторани та онлайн-знайомств, вивчення наукових статей і експертів, співавторів і фінансових послуг [6].

Далі переходимо до визначення «колаборативна фільтрація». Колаборативною фільтрацією називають один з методів побудови прогнозу в рекомендаційних системах, що використовує відомі уподобання групи користувачів для прогнозування невідомих уподобань іншого користувача.

Цей метод будується на основному припущенні: ті користувачі, що однаково оцінювали будь-які предмети/послуги в минулому, схильні давати схожі оцінки інших предметів/послуг і в майбутньому. Тобто, ці прогнози складаються індивідуально для кожного користувача, хоча інформація, що використовується для аналізу, зібрана від групи користувачів [7].

Значення оцінок, які користувач u дає елементу i , обчислюється як сукупність оцінок деяких схожих користувачів щодо елемента:

$$r_{u,i} = \text{aggr}_{u' \in U} r_{u',i},$$

де U є набором перших N користувачів, що найбільше схожі на користувача u , якому сподобався елемент i .

У вигляді формули цей алгоритм може бути представлений як

$$r_{u,i} = k \sum_{u' \in U} \text{sim}(u, u') r_{u',i},$$

де функція sim – обрана міра схожості двох користувачів, U – безліч користувачів, r – виставлена оцінка, k – нормувальний коефіцієнт

$$k = 1 / \sum_{u' \in U} |sim(u, u')|.$$

Найвідомішим прикладом використання колаборативної фільтрації є такі компанії-гіганти як Amazon та Netflix [8]. Amazon розпочав використовувати метод колаборативної фільтрації в перші роки свого існування. Це допомогло корпорації досягнути підвищення прибутку на 10% тільки внаслідок використання цього алгоритму.

Аналогічно, Netflix збільшує кількість контенту, що переглядається на 40% внаслідок використання алгоритму рекомендаційної системи. Мабуть, зараз простіше знайти компанію, яка не використовує подібний підхід, ніж перераховувати всіх, хто використовує. Цей факт доказує нам, що цей метод є ефективним та дуже популярним серед сучасних застосунків.

Здавалось би, ось він наш ідеальний алгоритм підбору кандидатів. Але, що робити з користувачами, що тільки зареєструвалися та ще не мають ніяких вподобань? Це поняття називається «Холодний запуск».

Нові користувачі представляють велику проблему для систем рекомендацій. Частково цю проблему можна вирішити за допомогою підходу, що заснований на аналізі вмісту.

Аналіз вмісту покладається не на оцінки, а на атрибути. Це допомагає включати нові варіанти в рекомендації для користувачів. Але проблему з наданням рекомендації для нового користувача вирішити складніше.

Як ще один варіант для розв'язання цієї проблеми — можна намагатися отримати інформацію з інших джерел. Наприклад, це можуть бути дані про користувача з інших аналогічних або не аналогічних сервісів або опитувальник при реєстрації.

Як ідея, під час реєстрації нових користувачів, можна запропонувати їм обрати 5 питань, що є пріоритетними для них. Партнери, відповіді яких будуть 100 відсотково збігатися на ці питання, будуть у топі рекомендацій для нових юзерів.

2.2 Огляд технологій для практичної реалізації проєкту

Вибір СКБД, що було обрано для практичної реалізації веб-застосунка буд частково оснований на щорічному рейтингу кращих СКБД, який був опублікований у 2022. Як можна побачити на рисунку 2.1, MySQL є найпопулярнішою серед усіх аналогів. Окрім популярності, одною з ключових переваг був той факт, що СКБД є безкоштовною, що дозволяє реалізувати даний проєкт без додаткових витрат. Додатково можна відмітити, що завдяки такій популярності, серед користувачів, дана система є гарно задокументована, що надає можливість швидко знаходити відповіді на питання, які можуть виникати під час настройки СКБД [9].

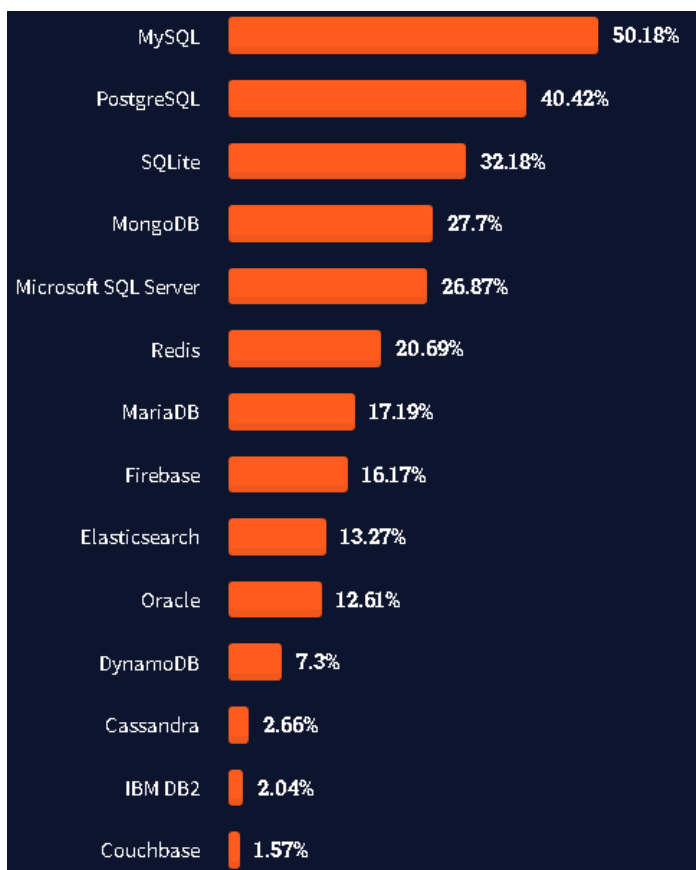


Рисунок 2.1 – Рейтинг найпопулярніших СКБД у 2022 році.

Щодо мов програмування, була обрана технологія ASP.NET, що застосовується для створення масштабних продуктів, які здатні витримувати великі навантаження.

Найбільшою перевагою при виборі мови програмування був той факт, що при розробці застосунку можна використовувати будь-яку мову з платформи .Net – C# або Visual Basic.Net та доступність великої кількості бібліотек, яка надає можливість зберігати час під час написання коду, просто використовуючи вже існуючі елементи. Також, потрібно пам'ятати, що ASP.NET є однією з самих швидких мов програмування (рис.2.2) [10].

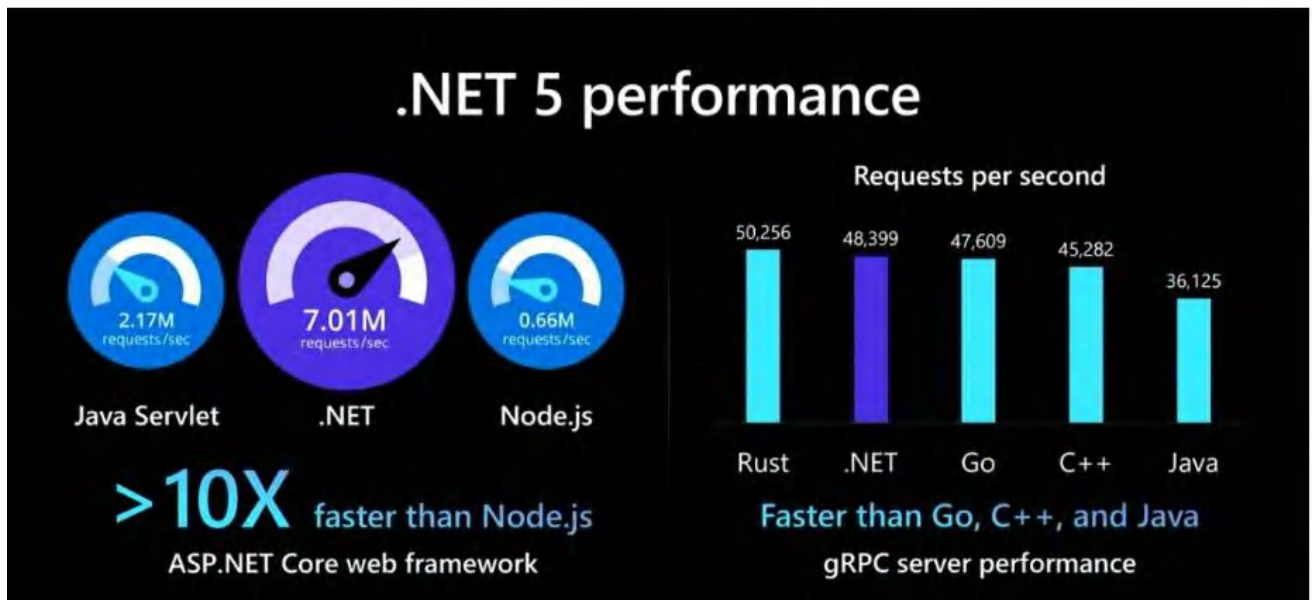


Рисунок 2.2 – Порівняння швидкодії різних мов програмування у 2022 році.

2.2.1 Обґрунтування вибору СКБД

Перш ніж обирати мову програмування, що будемо використовувати для практичної реалізації проєкту, визначимося з СКБД, в якій ми плануємо зберігати дані.

СКБД (система керування базами даних) - це програмне забезпечення, яке взаємодіє з кінцевими користувачами, програмами та самою базою даних для збору й аналізу даних. Програмне забезпечення СКБД додатково містить додаткові засоби, надані для адміністрування бази даних [11].

У нашому випадку, ми розраховуємо на велику кількість користувачів, це означає те, що зберігати даних доведеться багато. У СКБД плануються лише

зберігаються дані, а інтерфейс користувача реалізовано за допомогою мови програмування.

Для зберігання великого об'єму даних використовується СКБД під назвою «MySQL» [12].

MySQL є вільним і відкритим програмним забезпеченням, що дозволяє користувачам взаємодіяти безпосередньо з MySQL Database, використовуючи лише SQL. Але додатково, MySQL може використовуватися з іншими програмами.

Зараз MySQL є одною з найпоширеніших СКБД. Найбільше, вона використовується для створення динамічних веб-сторінок, оскільки має якісну підтримку зі сторони багатьох мов програмування. Також, MySQL надає широкий набір функцій, що підтримують безпечне середовище для зберігання, обробки та отримання даних [13].

Ця система характеризується своєю швидкістю, стійкістю і простотою у використанні. Вона була розроблена для підвищення швидкості обробки великих БД [14].

Переваги використання MySQL сервера:

- а) простота налаштування та використання;
- б) підтримка необмеженої кількості користувачів, що одночасно працюють із БД;
- в) зберігання до 50 мільйонів рядків у таблицях;
- г) висока швидкість виконання команд;
- г) MySQL працює на різних платформах, серед яких Unix, Linux, Windows, OS/2, Solaris, Mac OS;
- д) наявність простої та ефективної системи безпеки.

2.2.2 Обґрунтування вибору мови програмування

Для практичної реалізації серверної та клієнтської частин проєкту була обрана мова програмування ASP.NET. Ця мова програмування є одною з лідерів

по використанню для розробки серверної частини веб застосунків.

ASP.NET – це платформа для розробки веб застосунків від всесвітньо відомої компанії Microsoft. Ця платформа включає різні веб сервіси, програмну інфраструктуру та інші компоненти.

Головною перевагою обраної технології є то, що розробники додатково можуть використовувати інші мови програмування. А саме, C#, Visual Basic.NET, J# та JScript. Саме це суттєво розширює можливості розробки та дозволяє створювати більш високоякісні продукти для будь-якої предметної області [15].

ASP.NET підходить для розробки веб-сайтів, застосунків з реалізацією технологій реального часу, створення мобільних застосунків, та для роботи з WebHooks. Ця платформа є професійним інструментом для створення класних сучасних комерційних продуктів.

Попри те, що для вивчення ASP.NET потрібно витратити досить багато часу, перейдемо для переліку переваг цієї технології:

а) код, написаний на ASP.NET, виглядає лаконічним. Для розв'язання одної задачі потрібно менше часу та ресурсів, ніж при використанні інших мов програмування;

б) ASP.NET добре підходить для проєктів, де очікуються значні навантаження. Якщо застосунок взаємодіє з великою кількістю клієнтів, то ця мова програмування добре з цим впорається;

в) великий набір технологій та інструментів, які дозволяють вирішувати завдання будь-якої складності;

г) ASP.NET – кросплатформова технологія)

г) обслуговування готового проєкту коштує недорого;

д) високі можливості щодо безпеки. Підтвердженням цього є факт, що ASP.NET часто використовується для створення урядових та корпоративних сайтів;

е) велика кількість документації та різної літератури.

3 МАТЕМАТИЧНИЙ ОПИС ЗАДАЧІ

Якщо описати задачу як математичну модель у цілому, отримуємо, що наша головна задача — знаходження найбільш релевантного об'єкта. Якщо повернутися до обраної предметної області, наша задача знайти найбільш підходящий «профіль», тобто партнера для потенційного знайомства

$$\forall u \in U, s'_u = \arg \max_{s \in S} h(u, s),$$

де U – множина зареєстрованих користувачів системи,

u – обраний користувач, тобто той користувач, для якого ми підбираємо кандидатів для знайомств,

S – множина елементів системи (тобто користувачів), що потенційно можуть бути показані (рекомендовані) користувачеві для знайомства,

s – вибраний елемент (профіль кандидату для знайомств),

h – функція, яка показує ступінь сходження деякого елемента s з деяким користувачем u .

Тобто, наша задача зводиться до знаходження такого елемента (кандидату) $s' \in S$ з списку зареєстрованих користувачів, для якого ступінь зіставлення з користувачем $u \in U$ є максимальним.

3.1 Колаборативна фільтрація

Рекомендаційні системи, які базуються на використанні колаборативної фільтрації аналізують дані саме про користувачів системи.

Кожен юзер платформи попадає у так звану «групу» зі схожими смаками. Саме тут з'являється гіпотеза: користувачі, які однаково оцінили деякі об'єкти в минулому, швидше за все, однаково оцінять інші об'єкти в майбутньому [16].

Далі наведена схема, яка точно відображає як саме працює цей принцип (рис 3.1).

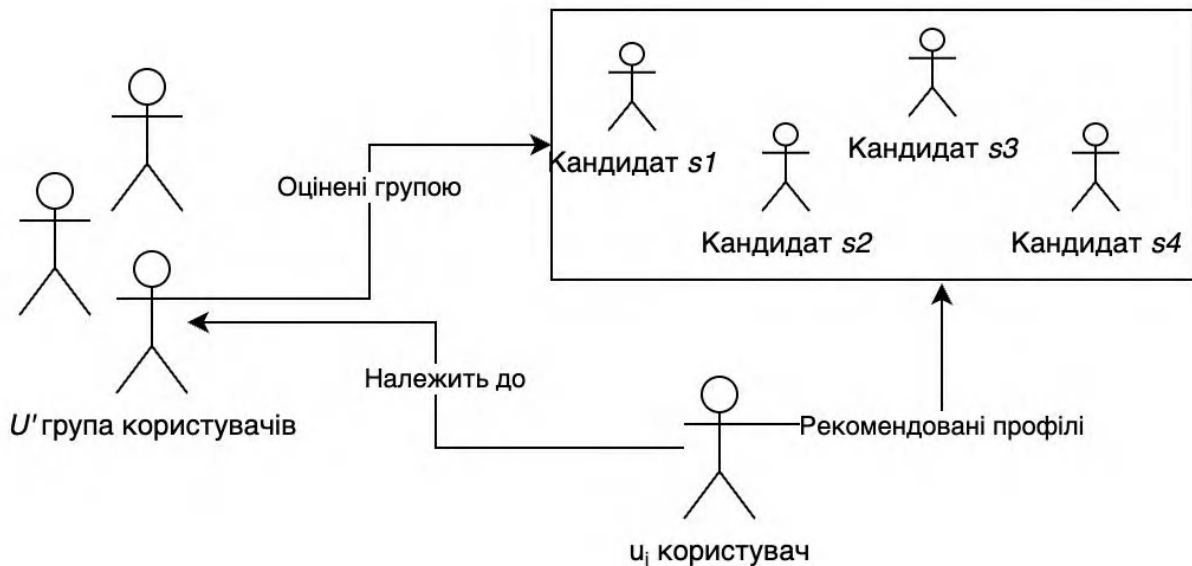


Рисунок 3.1 – Принцип роботи рекомендаційної системи на основі колаборативної фільтрації

Як ми вже знаємо, для побудови списку рекомендацій методи колаборативної фільтрації використовують інформацію про оцінки інших користувачів, які належать до одної групи. Тобто, функція $h'(u, s)$ задає ступінь зіставлення елемента (тобто профілю кандидата) s з користувачем u , та обчислюється в залежності від зіставлення $h'(u_i, s)$ для профілю кандидата s для користувачів $u_i \in U'$, де U' – деяка група користувачів [16].

Щоб віднести користувача до групи зі схожими інтересами, спочатку треба розпочати в створення профілю користувача. Так знаний «профіль» може ґрунтуватися на конкретні дані про його вподобання, або ж на загальній інформації. Наприклад, стать, вік та ін.

Ця інформація є дуже корисною при «холодному старті». Тобто, коли користувач тільки зареєструвався в системі, та ще не встиг оцінити достатньої кількості профілів, для формування його вподобань. Система ще нічого не знає

про користувацькі вподобання, тому і не може збудувати коректні рекомендації [17].

Саме для розв'язання цієї проблеми використовується замість загальних даних – результати психологічного тестування користувачів, що проходить кожен з них під час реєстрації. Щоб віднести нового користувача до конкретної групи користувачів зі схожими інтересами, треба прорахувати коефіцієнт розбіжності з кожним зареєстрованим користувачем. Якщо коефіцієнт розбіжності < 25 – вважаємо, що користувачі належать до спільної групи інтересів. Коефіцієнт розбіжності розраховуємо за формулою:

$$R_{ij}^k = |Q_i^k - Q_j^k|,$$

де R_{ij}^k – розбіжність i -го та j -го користувача, що потенційно належать до однієї групи для k питання;

Q_i^k , Q_j^k – відповідь в балах i -го, j -го користувача для k питання, відповідно.

Отримуємо, якщо бал розбіжності $R_{ij}^k < 25$ користувачі належать до однієї групи. Далі, наша задача зводиться формування списку так званих кандидатів, що ґрунтується на вже існуючій уподобаннях групи.

Наприклад, порівнюємо результати психологічного тестування нового користувача з 5 вже існуючими користувачами системи. Аналогічно п. 2.1.1, розраховуємо коефіцієнт розбіжності для нового користувача та вже існуючих. Отримані результати наведено в таблиці 3.1.

Таблиця 3.1 – Результати розподілення користувачів на групи зі схожими інтересами/уподобаннями

	Коефіцієнт розбіжності	Співпадіння
Юзер_1	80	Не належать до 1 групи
Юзер_2	4	Належать до 1 групи
Юзер_3	15	Належать до 1 групи

Юзер_4	25	Не належать до 1 групи
Юзер_5	26	Не належать до 1 групи

Для цього прикладу бачимо, що у нового користувача схожі уподобання з Юзер_2 та Юзер_3. Тобто, ми можемо вважати, що ті профілі/кандидати, що сподобалися цій «групі» користувачів – потенційно можуть сподобався нашому новому юзеру. Що вирішує проблему «холодного старту» для нових користувачів.

Колаборативна фільтрація гарно працює, коли профіль користувача містить достатньо даних, завдяки чому є можливість достатньо точно передбачити його вподобання. В нашому випадку використовуються результати психологічного тестування для отримання базової інформації для «старту» нового користувача. Система спирається на історію вподобань вже існуючих користувачів зі схожими інтересами.

Великою перевагою цього підходу є той факт, що платформі не потрібно обробляти величезні масиви даних, що належать до інформації про самого користувача, або набору так званих кандидатів.

3.2 Модифікація методу колаборативної фільтрації з використанням середньозваженого значення

Під час аналізу предметної області постало питання, чи можна задавати вагові коефіцієнти тим запитанням, що найкраще відображають вподобання користувача та є найбільш пріоритетними для нього.

Для реалізації цієї можливості, вирішено використовувати колаборативну фільтрацію з додаванням середньозваженого значення. У середньозваженому значенні кілька точок даних роблять більший внесок у результуюче середнє значення. Тут кожна точка даних пов'язана з певною вагою. Залежно від ваг спостережень внесок у підсумкове середнє варіюється.

Цими точками як раз стануть так звані «пріоритетні» питання для кожного користувача.

Дане значення розраховується шляхом множення ваги на пов'язаний з ним кількісний результат, а потім додавання всіх продуктів разом. Після цього, отриманий результат ділиться у сумі всіх ваг, що дає середньозважене значення.

Опис алгоритму модифікованого методу наведено нижче:

- Етап 1. Аналогічно класичній колаборативній фільтрації, віднесемо нового користувача до конкретної групи зі схожими інтересами.
- Етап 2. Після визначення групи зі схожими інтересами, отримуємо список «потенційних кандидатів для знайомств» на основі спільних інтересів групи.
- Етап 3. Після отримання списку ««потенційних кандидатів для знайомств» на основі спільних інтересів однієї групи, виключаємо с цього списку профілі, що повторюються використовуючи перетин множин. Як результат - отримуємо список унікальних кандидатів для потенційного знайомства.
- Етап 4. Переходимо до формування рейтингу кандидатів. Аналогічно пункту 2.1.1, розраховуємо коефіцієнт розбіжності. Після цього замінюємо його на коефіцієнт співпадіння відповідно до значень: 0 – 100% співпадіння, 1 – 75%, 2 – 50%, 3 – 25%, 4 – 0%.
- Етап 5. Далі розраховуємо середньозважене значення, яке будемо використовувати для формування рейтингу за формулою

$$\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i},$$

де \bar{x} – середньозважене значення для формування рейтингу кандидатів,

w_i – ваговий коефіцієнт i -го питання,

x_i – процент співпадіння двох користувачів на i -те питання.

Наша задача зводиться до формування списку с кандидатами, які є потенційно цікавими для користувача. Найбільш вдалимими «співпадіннями» є профілі, у яких середньозважене значення є більшим.

Якщо порівняти формування рейтингу за допомогою звичайного алгоритму сортування та колаборативну фільтрацію, то другий метод звичайно є більш складним, але на той самій час більш ефективним и результативним.

Не дивлячись на головний недолік цього методу, а саме проблему «холодного старту», знайдено рішення, яке задовольняє обрану предметну область. Категоризація нових користувачів буде виконуватися за допомогою даних, які система отримує під час обов'язкового психологічного тестування, яке є одним з етапів реєстрації.

Модернізація колаборативного методу за допомогою використання середньозваженого значення для побудови рейтингу кандидатів, допомогла швидко та просто додати функцію пріорітезації, яка є дуже важливою для даного типу тестування. Таким чином вона надає користувачам можливість робити акценти на тих речах, які є для них важливими та принциповими.

4 ВИЗНАЧЕННЯ ВИМОГ ДО РОЗРОБЛЮВАНОЇ СИСТЕМИ

4.1 Визначення системних вимог до розроблювальної інформаційної системи

Як вже згадувалося у попередніх розділах, для автоматизація основних бізнес-процесів агенції знайомств, було обрано реалізувати веб-застосунок. Цей веб-застосунок допоможе прискорити процес підбору кандидатів для знайомств та процес обробки запитів на побачення.

Система матиме три рівні доступу: незареєстровані користувачі (не має доступу до даних системи, має можливість зареєструватися та/або авторизуватися), зареєстровані користувачі (має доступ до перегляду та редагуванню інформації) та адміністратор інформаційної системи (має права на перегляд, редагування та видалення інформації).

Виходячи з того, що стартап планують активно масштабуватися, майбутня система повинна витримувати високі навантаження. Агенція очікує великий потік користувачів, які одночасно використовуватимуть застосунок. Так як основні запити до системи будуть на редагування та зчитування інформації з одного ж і того джерела, важливо враховувати, що система повинна швидко обробляти ці запити та повертати відповідь від сервера до «клієнта».

Далі наведений перелік системних вимог для майбутнього веб-застосунку:

- а) однакове коректне відображення в наступних сучасних браузерях: Chrome, Safari, Opera;
- б) використання версії СКБД: MySQL ≥ 5.1 ;
- в) швидкість інтернет-з'єднання: 256 Кбіт/с;
- г) процесор (ЦП): з двома або більше фізичними ядрами, що підтримують технологію SSE2.

1.2 Визначення функціональних вимог до розроблювальної системи

Функціональними вимогами називаються такі вимоги ІС, що описують її внутрішню роботу та поведінку. Тобто, точно визначається яким чином буде виконатися робота з даними (їх обробка). Кожна функціональна вимога описує кожну специфічну функцію майбутньої системи. Іншими словами, цей вид вимог допомагає визначити, що саме і як наша система повинна робити [18].

До функціональних вимог можна віднести перевірку коректності даних, що вводить користувач. Наш застосунок повинен містити перевірку для коректного вводу даних користувача для таких полів як пароль, номеру телефону, email, дата народження (так як системою можуть користуватися лише повнолітні користувачі).

Для визначення функціональних вимог будь-якої інформаційної системи звичайно використовують побудування функціональної моделі з використанням стандарту IDEF0. Також її ще називають контекстною діаграмою.

Контекстна діаграма зображується у вигляді прямокутника з вхідними й вихідними даними. На цьому прямокутнику основна функція (ціль системи). В нашому випадку це процес створення побачення.

Входи й виходи діаграми по чотирьох сторонах прямокутника:

Стрілки зліва відповідають входам системи, тобто тим даним, що поступають до ІС і переробляються нею у вихідні дані.

Стрілки зверху відповідають входам по керуванню. Це набір різних керуючих дій, команд, процедур і т.д, що регламентують процеси виконання роботи. Ці дані служать тільки для керування.

Стрілки справа відповідають виходам системи, тобто продуктам діяльності ІС.

Стрілки знизу відповідають механізмам, тобто засобам та ресурсам, за допомогою яких виконуються вказані в прямокутнику функції.

На наступній концептуальній діаграмі (рис 4.1), яка була створена з точки зору адміністратора, можна побачити основний бізнес-процес та інтерфейси майбутньої ІС.

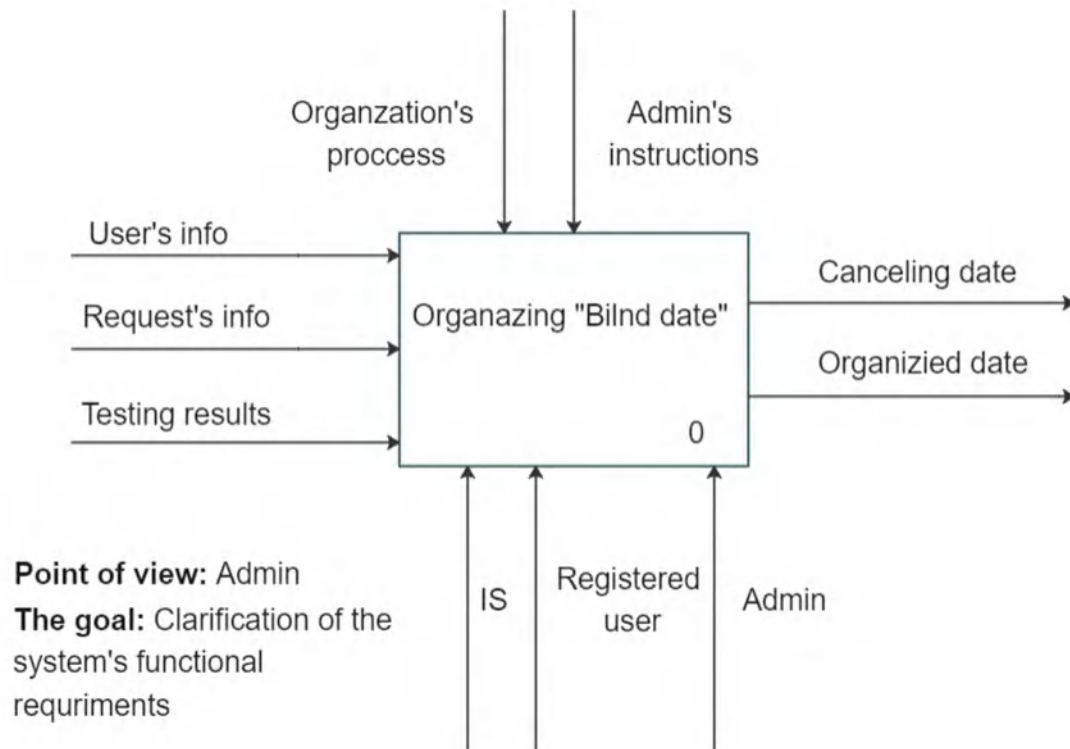


Рисунок 4.1 – Контекстна діаграма основного бізнес-процесу інформаційної системи агенції знайомств «Blind date»

На цій контекстній діаграмі чітко бачимо перелік вхідних даних, що необхідна для функціонування системи:

- результати психологічного тестування;
- персональні дані про користувача;
- запит на організацію побачення.

Зверху бачимо механізми керування: правила організації побачень в агенції знайомств (порядок робіт, будь-що) та інструкції адміністратора. Знизу бачимо, що у якості механізмів присутні користувач системи, ІС та адміністратор. Як результат, справа на виході отримуємо наступну інформацію:

- а) якщо обидва користувача дали свою згоду на побачення – отримуємо детальну інформацію про організацію;
- б) у разі відмови хоча б одного з користувачів, отримуємо інформацію о причині відмови та інформацію про скасоване побачення.

Далі, декомпозуємо нашу контекстну діаграму (рис 4.2).

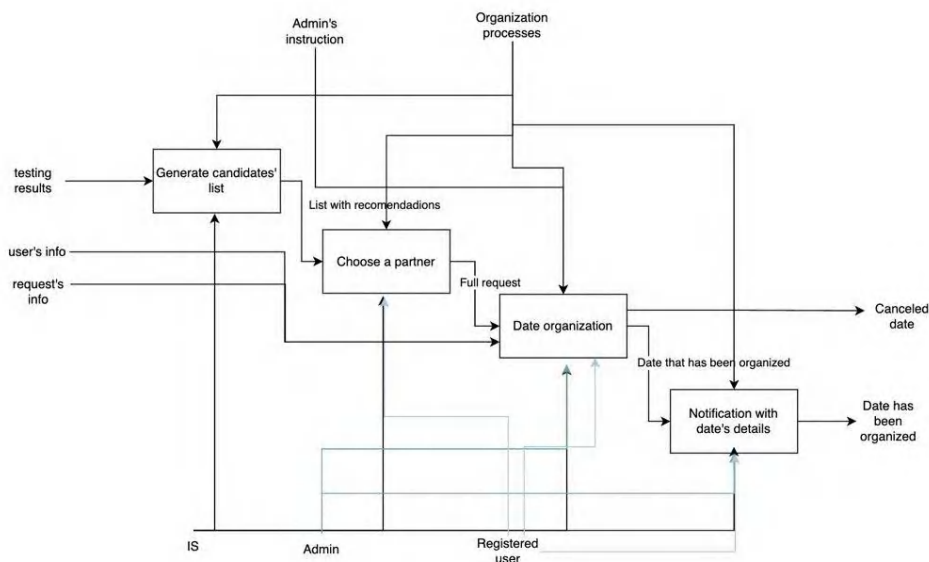


Рисунок 4.2 – Діаграма декомпозиції основного бізнес-процесу ІС агенції знайомств «Blind date»

Діаграма декомпозиції завжди відображає послідовність етапів, з яких складається основний бізнес-процес.

- першим етапом є генерація списку-рекомендації профілів користувачів, що можуть бути потенційно цікаві нашому юзеру;
- далі іде процес вибору партнера для майбутнього побачення;
- наступним етапом організація самого побачення, до якого додається детальна інформація про майбутнє побачення;
- заключачим етапом є надіслання нотифікації з детальної інформацією про організоване побачення.

Далі наведена діаграма дерева вузлів (рис 4.3), що визначає точну структуру моделі. Деревом вузлів називають представлення відношень між батьківськими і дочірніми вузлами моделі IDEF0 у формі деревовидного графа.

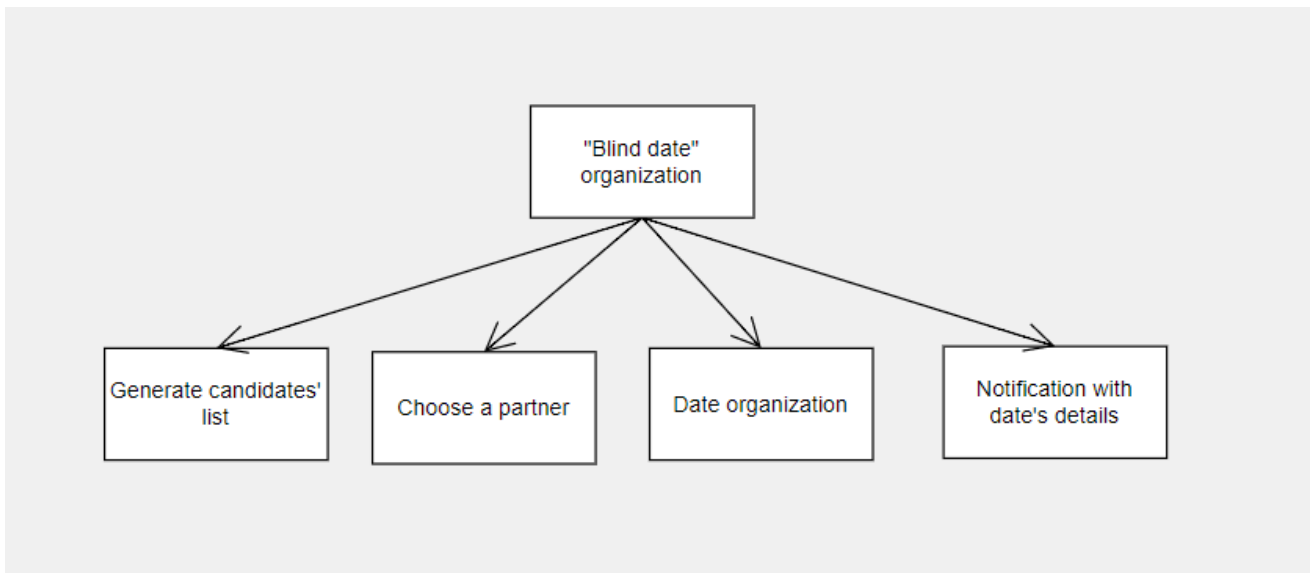


Рисунок 4.3 – Діаграма дерева вузлів основного бізнес-процесу ІС агенції знайомств «Blind date»

1.3 Визначення вимог до інтерфейсу розроблювальної системи

Одною з головних вимог до розроблюваного застосунку, як і до усіх сучасних веб-сервісів, є наявність інтуїтивно зрозумілого інтерфейсу. Користувальницький інтерфейс називається інтуїтивним у тому випадку, коли усі користувачі розуміють його поведінку та функціонал без використання розуму, чиеїсь допомоги або спеціальної підготовки.

Якщо сервіс має інтуїтивний інтерфейс, який легко зрозуміти і використати, тоді відвідувачі, з більшою ймовірністю, перейдуть до використання цього продукту.

Інтерфейс майбутнього застосунку не повинен використовувати складні конструкції, а також нетиповий для дейтінгових сервісів функціоналу.

Зі сторони користувача, найважливішим є зрозуміле відображення даних про кандидатів, побачень та їх статуси, інформації в особистому кабінеті а також можливість їх редагування.

Зі сторони ж адміністратора, інтерфейс користувача буде складатися в основному з звичайних таблиць, які будуть відображати необхідні для роботи дані, кнопок на полів для редагування.

Далі буде наведений перелік графічних елементів, що будуть використовуватися під час практичної реалізації.

а) Сторінка «Вхід» – є першої сторінкою, з якою зустрічається користувач системи. Ця сторінка відображає наступні графічні елементи: текстове поле «Логін», текстове поле «Пароль», кнопка «Вхід». Додатково, створюємо ще одно текстове поле, що буде перенаправляти користувача на сторінку «Реєстрації» у випадку, якщо користувач ще не має аккаунту. При натисканні на кнопку «Вхід» відбувається вхід до системи.

б) Сторінка «Реєстрація» – є сторінкою, що використовується для реєстрації нового користувача в системі. Ця сторінка відображає наступні графічні елементи: текстове поле «Ім'я», текстове поле «Фамілія», текстове поле «Країна», текстове поле «Місто», текстове поле «email», текстове поле «Номер телефона», текстове поле «Гендер», текстове поле «Дата народження», текстове поле «Пароль», текстове поле «Додаткова інформація», кнопка «Продовжити реєстрацію». Після натискання на кнопку «Продовжити реєстрацію», користувач переходить до наступного етапу, тобто до психологічного тестування, на основі якого далі буде проходити процес підбору кандидатів.

в) Сторінка «Тестування» – сторінка, що відобразатиме перелік з 36 питань та варіантів відповідей для проведення психологічно тестування. Ці питання автоматично відображають дані із системи, у вигляді наступних графічних елементів: текстове поле «Питання_n», що відображає саме питання, а для виведення відповідей використовується такий графічний елемент як checkbox. До кожного питання надається 5 стандартних відповідей: завжди, часто, іноді, зазвичай, завжди. Після останнього запитання на екрані з'являється кнопка «Закінчити тестування» і користувач автоматично потрапляє до особистого кабінету.

г) Сторінка «Особистий кабінет» – сторінка, яка відображає дані про користувача. На сторінці особистого кабінету кожен користувач має доступ до редагування своїх персональних даних. Ця сторінка відображає наступні графічні елементи: текстові поля «місто», «email» та «номер телефона» є доступні для

редагування. А поля «гендер» та «дата народження» не доступні для редагування. Усі дані автоматично підтягуються з БД. Для оновлення даних у системі необхідно натиснути кнопку «Зберегти». Також, в окремій секції користувач має переліт сторінок, які доступні для перегляду: «Мої партнери», «Побачення».

г) Сторінка «Мої партнери» – сторінка, що відображає список кандидатів, що були автоматично підібрані системою для потенційного знайомства. Користувач має можливість як переглянути список кандидатів у вигляді списку, а також детально розглянути кожну анкету більш детально, лише за допомогою кліку на ім'я кандидата.

д) Сторінка «Побачення» – сторінка, що відображає 2 списки: перший містить запити на побачення, що відправив користувач, другий – запити, що адресовані нашому користувачеві. Для роботи з другим списком, користувач має кнопки для скасування побачення, або навпаки – для надання згоди на організацію побачення з n-им кандидатом. Зробити це можна за допомогою наступних графічних елементів: кнопка «Погодитися» та кнопка «Відмінити». Для роботи з власними запитами, у користувача є також кнопка «Відмінити», для тих випадків, коли користувач передумав або відправив запит випадково.

е) Сторінка «Профіль користувача» – сторінка, що відображає анкету кандидата. Ця сторінка відображає наступні графічні елементи: текстові поля, що відображають дані про партнера з нашої БД, кнопка «Відправити запит», яка, після натискання, формує новий запит на побачення, додаючи його в базу даних.

є) Додатково, для користувачів з типом доступу «Admin» будуть реалізовані додаткові сторінки, які будуть відображати дані за допомогою графічного елементу `dataGridView`. За його допомогою, легко та зручно не тільки відображати дані, а й редагувати та видаляти. Кожен «Admin» буде мати доступ до наступних «робочих» таблиць:

- міста;
- місця для побачення;
- типи побачень;
- питання;

- відповіді;
- результати тестування;
- запити на побачення;
- рахунки.

4.4 Розробка діаграми варіантів використання системи

Для того, щоб точніше зрозуміти як саме повинна працювати система у майбутньому, частіше всього використовується опис через варіанти використання.

Діаграма варіантів використання системи (ще її також називають «Діаграма прецедентів»), зазвичай будується для відображення проектованою систем у вигляді набору сутностей та акторів, які взаємодіють із системою за допомогою так званих варіантів використання.

Варіант використання відображає перелік послуг, що надаються кожному актору системою. Тобто, кожен варіант використання визначає деякий набір дій, який виконує система при діалозі з актором.

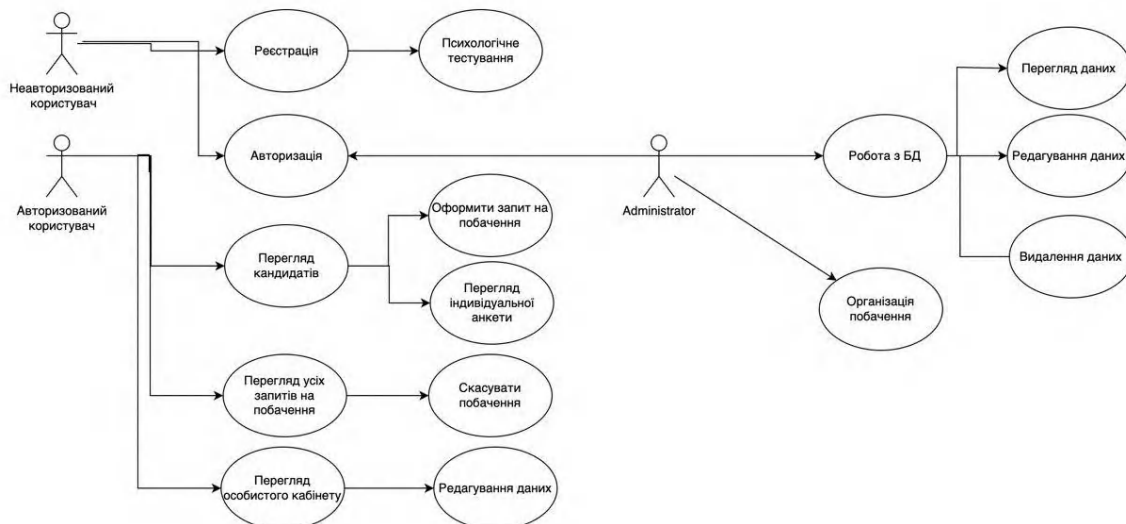


Рисунок 4.4 – Діаграма прецедентів (Use Case Diagram) ІС агенції знайомств «Blind date»

Як можна побачити на діаграмі прецедентів (рис.4.4) Головними акторами даної інформаційної системи є: адміністратор, авторизований користувач та неавторизований користувач.

Прецедентами для неавторизованого користувача є: можливість реєстрації або авторизації в системі. Даний актор не має доступу до внутрішнього функціоналу ІС, а також даних.

Прецедентами для авторизованого користувача є: функція перегляду особистого кабінету, перегляд усіх запитів на побачення, перелік кандидатів.

Для ролі адміністратора можна виділити наступні прецеденти: робота з даними та можливість організації побачення.

Додатково, хочеться приділити увагу наступному прецеденту – «Перегляд кандидатів». Саме тут буде реалізований функціонал автоматичного підбору списку кандидатів, який надається кожному користувачеві, на основі психологічного тестування. Даний функціонал базується на модифікованому алгоритмі колаборативної фільтрації, який буде розглянуто у наступних розділах.

5 ЛОГІЧНЕ І ФІЗИЧНЕ МОДЕЛЮВАННЯ ДАНИХ

У ході аналізу предметної області визначено, що для майбутнього дейтінг-застосунку планується реалізувати 3 режими користувачів: Адміністратор, авторизований та не авторизований користувач.

Щоб описати структуру даних, які ми використовуємо, далі буде проведений процес моделювання даних та наведена логічна та фізичні модель даних.

5.1 Логічне моделювання даних

Моделюванням даних називають процес створення моделі даних у вигляді діаграми (яка показує сутності системи, їх атрибути та зв'язки між ними) для проектування інформаційної системи [19].

Логічною схемою даних називають структуру, що описує певну предметну область. Вона складається з описів, які представлені у вигляді таблиць. Для її побудови використовуємо інформацію, що була отримана на етапі аналізу предметної області.

Основними сутностями для дейтінг-застосунку «Blind date» будуть:

а) User (користувач) – один з найголовніших об'єктів системи, що містить основну інформацію про користувачів застосунку. Сутність «User» містить в собі такі атрибути:

- первинний ключ;
- ім'я;
- прізвище;
- гендер (зовнішній ключ);
- e-mail;
- номер телефона;
- дата народження;
- місто проживання (зовнішній ключ);

- пароль;
- додаткова інформація;
- дата останнього відвідування;

б) Request (запит на побачення) – один з основних об'єктів системи, що зберігає необхідну інформацію про майбутнє побачення. Сутність «Request» містить у собі такі атрибути:

- первинний ключ;
- користувач, що відправив запит (зовнішній ключ);
- користувач, кому відправили запит (зовнішній ключ);
- статус заявки (зовнішній ключ);
- тип побачення для даного запиту (зовнішній ключ);
- дата;
- час;
- коментар;
- відгук;

в) City (місто) – додаткова таблиця, що зберігає в собі назви усіх міст, які обслуговуються агенцією в конкретний момент часу. Сутність «City» містить у собі такі атрибути:

- первинний ключ;
- назва міста;

г) Gender (гендер) – у 2022 році офіційно країнами ЄС признається 58 гендерів. Агенція «Blind date» також підтримує цю ідеологію, тому спеціально була створена додаткова таблиця, що зберігає у собі усі види гендерів. Сутність «Gender» містить у собі такі атрибути:

- первинний ключ;
- назва гендеру;

г) Results (результати тестування) – проміжна таблиця, що зберігає усі відповіді на тестування кожного користувача застосунку. Сутність «Results» містить у собі такі атрибути:

- користувач, що проходить тестування (перша частина складового ключа);

- питання, на яке користувач дає відповідь (друга частина складового ключа);

- відповідь (зовнішній ключ);

- опис;

д) Question (питання) – таблиця, що зберігає перелік питань для проходження психологічного тестування. На основі його результатів відбувається пошук партнера. Сутність «Question» містить у собі такі атрибути:

- первинний ключ;

- питання;

е) Answer (відповідь) — таблиця, що зберігає перелік відповідей для проходження психологічного тестування. На основі його результатів відбувається пошук партнера. Сутність «Answer» містить у собі такі атрибути:

- первинний ключ;

- питання;

є) Place (місце для побачення) – таблиця, що зберігає в собі інформацію про усі заклади-партнерів, з якими агенція має договір, на проведення побачень. Сутність «Place» містить у собі такі атрибути:

- первинний ключ;

- місто, в якому знаходиться заклад (зовнішній ключ);

- назва закладу;

- адреса;

ж) Date type (тип побачення) – таблиця, що зберігає інформацію про майбутнє побачення. Сутність «Date type» містить у собі такі атрибути:

- первинний ключ;

- місто, в якому проходитиме побачення (зовнішній ключ);

- дата та час побачення;

- тривалість побачення;

- детальний опис;

- додатковий коментар;

- вартість;

з) Date status (статус запиту) — таблиця, що показує поточний стан запиту на побачення (наприклад, заявка тільки створена та має статус «нова», або побачення вже відбулося, та статус запиту є «завершено» і т.д.). Сутність «Date status» містить у собі такі атрибути:

- первинний ключ;
- питання.

Після виділення усіх сутностей та атрибутів майбутнього застосунку, отриманні дані та їх зв'язки можна показати у вигляді логічної схеми даних, яка наведена далі на рисунку 3.1.

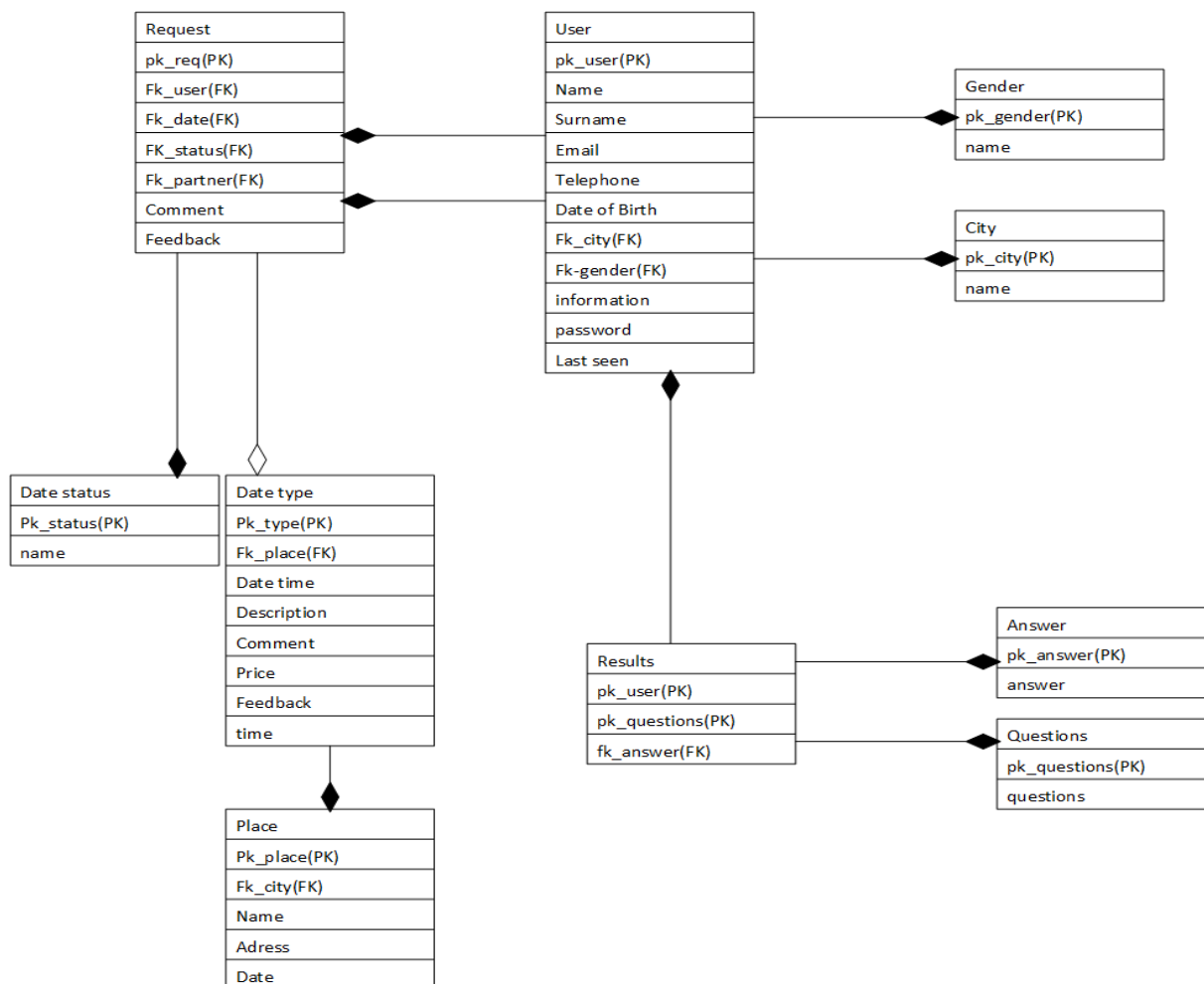


Рисунок 5.1 – Логічна схема даних ІС агенції знайомств «Blind date»

Як можна побачити на логічній схемі даних, дані можна розділити на 2 секції:

- дані, що відносяться до самих користувачів системи (User, Results). Цими даними в основному будуть користуватися самі юзери системи, рідше — адміністратори;
- дані, що належать до майбутніх побачень (Request, Place, Date type). Дані, що будуть використовувати виключно адміністратори.

Виходячи з вже формалізованого набору даних, яким повинен оперувати наш майбутній застосунок, можна сформулювати задачу дослідження наступним чином: дослідити методи підбору кандидатів в системах надання послуг з урахуванням уподобань клієнтів на прикладі агенції знайомств «Blind date».

Для практичної реалізації задачі підбору кандидатів, розглянуті наступні методи:

- колаборативна фільтрація (або ще називають user-based);
- модернізована колаборативна фільтрація з використанням середньозваженого значення.

Далі, у четвертому розділі буде описана математична модель самої задачі, а також детально порівняні обрані підходи для реалізації задачі пошуку кандидатів.

5.2 Фізичне моделювання даних

Наступним етапом є перетворення логічної моделі даних (рис 5.2) у фізичну. Процес перетворення відбувається на базі правил цілісності, дають змогу уникнути введення некоректних даних у БД, а також забезпечити можливість зв'язування 2 та більше таблиць [4].

Для її створення була використана платформа сервера MySQL, яка була обрана під час написання розділу 2. Типом таблиць при створенні БД був обраний InnoDB. Обрана нотація – IDEF1X, яка використовує спеціальний синтаксис зручної побудови концептуальної схеми [5].

Під час аналізу предметної області на логічного моделювання даних, були обрані наступні основні сутності: «Users», «Requests», «Gender», «City», «Results», «Question», «Answer», «Place», «Date type», «Date status».

Таблиця «Users» зберігає інформацію про кожного користувача:

- поле: id-користувача. Тип integer, обов'язкове для заповнення поле, є первинним ключем;
- поле: прізвище. Тип varchar, обов'язкове для заповнення поле;
- поле ім'я. Тип varchar, обов'язкове для заповнення поле;
- поле номер телефона. Тип varchar, обов'язкове для заповнення поле;
- поле email. . Унікальне поле типу varchar, обов'язкове для заповнення;
- поле пароль. Тип varchar, обов'язкове для заповнення поле;
- поле дата народження. Тип date, обов'язкове для заповнення поле;
- поле місто. Тип integer, обов'язкове для заповнення поле, є зовнішнім ключем;
- поле гендер. Тип integer, обов'язкове для заповнення поле, є зовнішнім ключем;
- поле додаткова інформація. Тип varchar, не обов'язкове для заповнення поле;
- поле дата останнього відвідування. Тип date, не обов'язкове для заповнення поле.

Таблиця «Gender» зберігає інформацію про різні типи гендерів:

- поле id-гендера. Тип integer, обов'язкове для заповнення поле, є первинним ключем;
- поле ім'я. Тип varchar, обов'язкове для заповнення поле.

Таблиця «City» зберігає інформацію про міста, в яких є філіал агенції знайомств «Blind Date»:

- Поле id-міста. Тип integer, обов'язкове для заповнення поле, є первинним ключем;
- поле назва. Тип varchar, обов'язкове для заповнення поле.

Таблиця «Results» зберігає інформацію про результати психологічного тестування користувачів системи:

- поле id-користувача. Тип integer, обов'язкове для заповнення поле, є частиною первинного ключа;
- поле id-питання. Тип integer, обов'язкове для заповнення поле, є частиною первинного ключа;
- поле відповідь. Тип integer, обов'язкове для заповнення поле, є зовнішнім ключем.

Таблиця «Answer» зберігає варіанти відповідей на питання психологічного тестування:

- поле id-відповіді. Тип integer, обов'язкове для заповнення поле, є первинним ключем;
- поле відповідь. Тип varchar, обов'язкове для заповнення поле.

Таблиця «Question» зберігає питання для психологічного тестування користувачів:

- поле id-питання. Тип integer, обов'язкове для заповнення поле, є первинним ключем;
- поле відповідь. Поле назва. Тип varchar, обов'язкове для заповнення поле.

Таблиця «Date status» зберігає варіанти статусів побачень:

- поле id-статусу. Тип integer, обов'язкове для заповнення поле, є первинним ключем;
- поле назва. Поле назва. Тип varchar, обов'язкове для заповнення поле.

Таблиця «Date type» зберігає варіанти для проведення побачень:

- поле id-типу. Тип integer, обов'язкове для заповнення поле, є первинним ключем;
- поле місце проведення. Тип integer, обов'язкове для заповнення поле, є зовнішнім ключем;
- поле дата та час. Тип datetime, обов'язкове для заповнення поле;
- поле опис. Тип text, обов'язкове для заповнення поле;

- поле коментар. Тип text, обов'язкове для заповнення поле;
- поле відгук. Тип varchar, обов'язкове для заповнення поле;
- поле тривалість. Тип float, обов'язкове для заповнення поле.

Таблиця «Place» зберігає варіанти місць для проведення побачення:

- поле id-місця. Тип integer, обов'язкове для заповнення поле, є зовнішнім ключем;
- поле місто. Тип integer, обов'язкове для заповнення поле, є зовнішнім ключем;
- поле назва. Тип text, обов'язкове для заповнення поле;
- поле адреса. Тип varchar, обов'язкове для заповнення поле;
- поле вартість. Тип float, обов'язкове для заповнення поле;
- поле дата додавання. Тип date, обов'язкове для заповнення поле.

Таблиця «Request» зберігає усі запити на побачення:

- поле id-запиту. Тип integer, обов'язкове для заповнення поле, є частиною первинного ключа;
- поле id-користувача. Тип integer, обов'язкове для заповнення поле, є частиною первинного ключа;
- поле партнер. Тип integer, обов'язкове для заповнення поле, є зовнішнім ключем;
- поле статус. Тип integer, обов'язкове для заповнення поле, є зовнішнім ключем;
- поле побачення. Тип integer, не обов'язкове для заповнення поле, є зовнішнім ключем;
- поле коментар. Тип text, не обов'язкове для заповнення поле;
- поле відгук. Тип varchar, не обов'язкове для заповнення поле;
- поле час. Тип float, обов'язкове для заповнення поле;
- поле дата. Тип date, обов'язкове для заповнення поле.

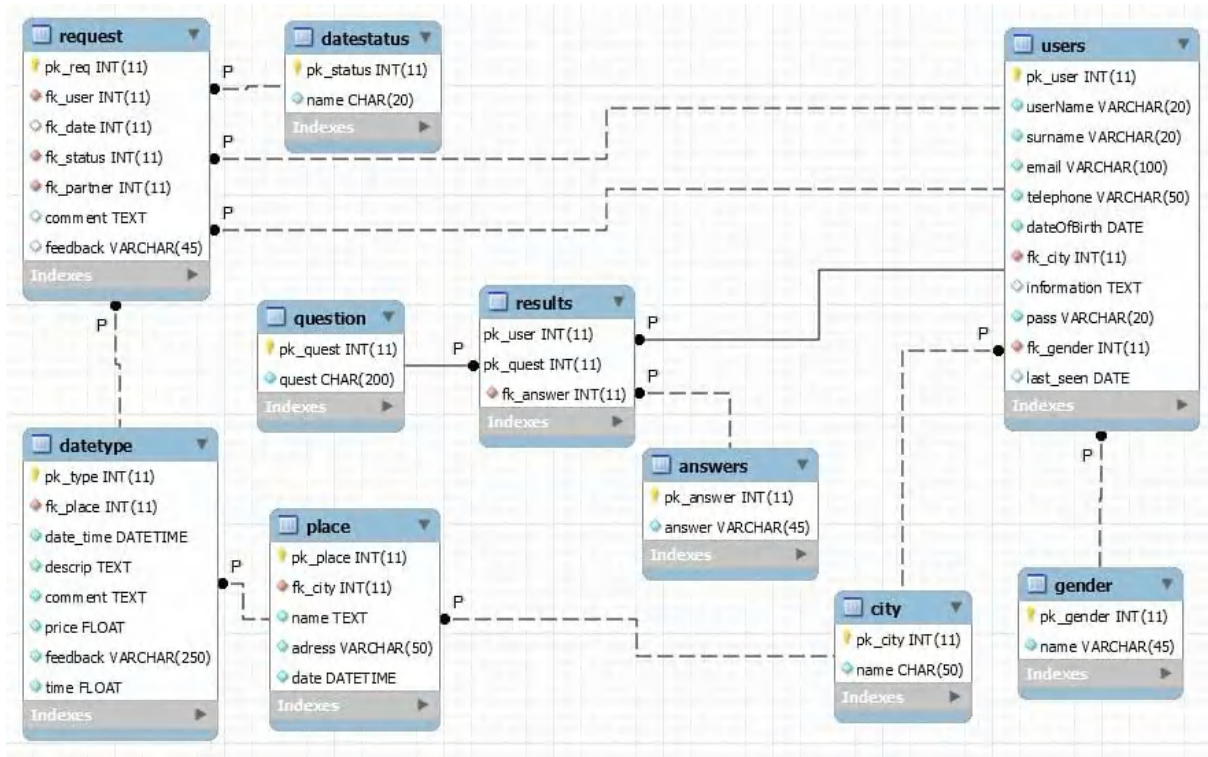


Рисунок 5.2 – Схема даних ІС агенції знайомств «Blind date»

6 РОЗРОБКА АЛГОРИТМУ ПІДБОРУ КАНДИДАТІВ

6.1 Загальний алгоритм роботи застосунку

З результатів розділу «Визначення вимог до розроблювальної системи» можна побачити, що майбутній веб-застосунок повинен містити великий функціонал. Необхідний для роботи агенції знайомств функціонал буде побудований на основі чіткої логіки послідовності використання даної системи. На рисунку 6.1 можна побачити алгоритм користувача при використанні веб-застосунку.

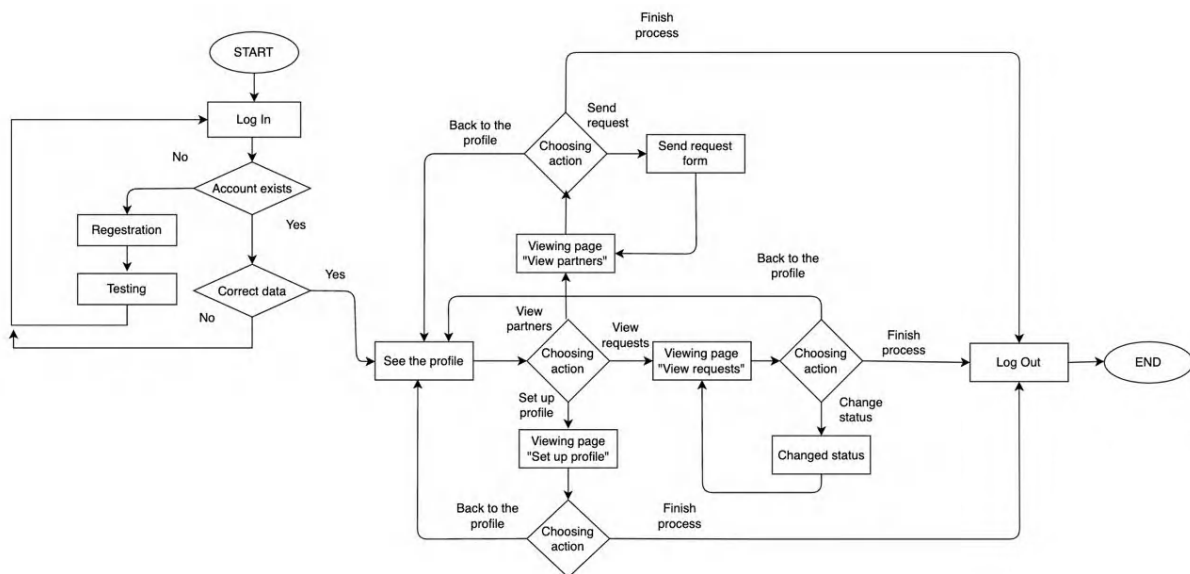


Рисунок 6.1 – Алгоритм використання веб-застосунку в ролі користувача

Далі приведемо розгорнутий опис кожного етапу даного алгоритму, що відображає послідовність використання компонентів ІС зі сторони користувача:

– 1й етап – «Перевірка наявності аккаунту». Система побудована таким чином, що неавторизований користувач не має доступу до жодного функціоналу системи, окрім авторизації та/або реєстрації. Саме через це обмеження, першим етапом є перевірка, чи існує даний користувач в системі чи ні. Якщо такий аккаунт вже існує – далі переходимо до перевірки коректності даних, що вводить користувач. Якщо дані коректні, користувач отримує доступ до свого профілю. Якщо ні – користувачеві надається можливість повторної авторизації.

Якщо даного аккаунту не існує – користувач перенаправляється до форми реєстрації, а потім форми тестування. Після завершення тестування, система направляє користувача до форми авторизації;

– 2й етап – «Перегляд власного профілю». Кожен користувач системи має доступ до особистого профілю, що зберігає персональні інформації користувача (ПІБ, email, номер телефону і т.д.). Додатково, користувач має доступ до редагування деяких полів. Наприклад, місто або номер телефону, для випадків, коли користувач місце проживання або номер телефону.

Із власного профілю, користувачеві надається можливість перейти до наступних розділів:

- а) перегляд партнерів;
- б) перегляд запитів на побачення;
- в) редагування профілю;

– 3й етап – «Перегляд партнерів». Щоб створити новий запит на побачення, користувачеві для початку необхідно переглянути список кандидатів, які автоматично підбрала ІС. Для цього користувач повинен перейти на сторінку «View partners», де він має доступ до повного списку з кандидатами. Тут же, користувач може відправити запит тим кандидатам, з якими він би хотів організувати побачення;

– 4й етап – «Перегляд запитів на побачення». Щоб переглянути список запитів на побачення, користувач повинен перейти на сторінку «View requests». На цій сторінці користувач може знайти 2 списку запитів: ті, що надіслав даний користувач, та ті, що були надіслані нашому користувачеві іншими користувачами системи. Також, додатково з цієї сторінки користувач може відмінити будь-який з відправлених запитів на побачення;

– 5й етап – «Log Out». Користувач має можливість завершити робочу сесію в застосунку за допомогою лише одного кліку на кнопці «Log Out», яка відображається з будь-якої сторінки.

6.2 Алгоритм підбору кандидатів

Для практичної реалізації було обрано модифікований алгоритм колаборативної фільтрації, що використовує середньозважене значення.

Для вирішення проблеми «холодного старту», ми будемо обирати до якої ж групи віднести нового користувача, за допомогою психологічного анкетування. По дефолту ми вважаємо, що до однієї групи зі схожими інтересами ми відносимо тих користувачів, у яких коефіцієнт розбіжності менше 25%.

При так званому «холодному старті» система працює наступним чином як на рис. 6.2. Наведемо приклад: Павло новий користувач нашої системи. Після проходження психологічного тестування, він був віднесений до групи зі схожими уподобаннями N.

Виявляється що користувачам Марку, Алексу, і Джону сподобалися принаймні один і той же профіль. Так як вони відносяться до однієї групи і розділяють схожі смаки, швидше за все, Павлу сподобаються ці ж самі профілі. Отримаємо ті профілі, що ми збираємося використовувати, щоб дати рекомендації Павлу.

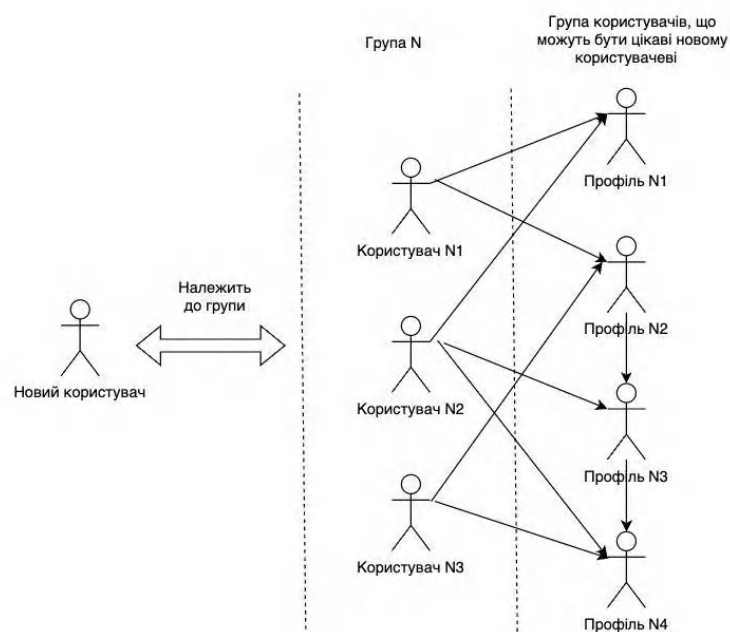


Рисунок 6.2 – Приклад роботи рекомендаційної системи при «холодному старті»

Для того, щоб зробити «match» найбільш влучним, було вирішено додати вагові коефіцієнту кожному питанню під час тестування. По-перше, це допоможе виявити ті аспекти життя, які є ключовими для даного користувача. А по-друге, за допомогою використання середньозваженого значення, ми зможемо зробити підбір кандидатів максимально якісним.

Далі, на рисунках 6.3-6.5 буде наведений алгоритм, зі сторони реалізованої системи.

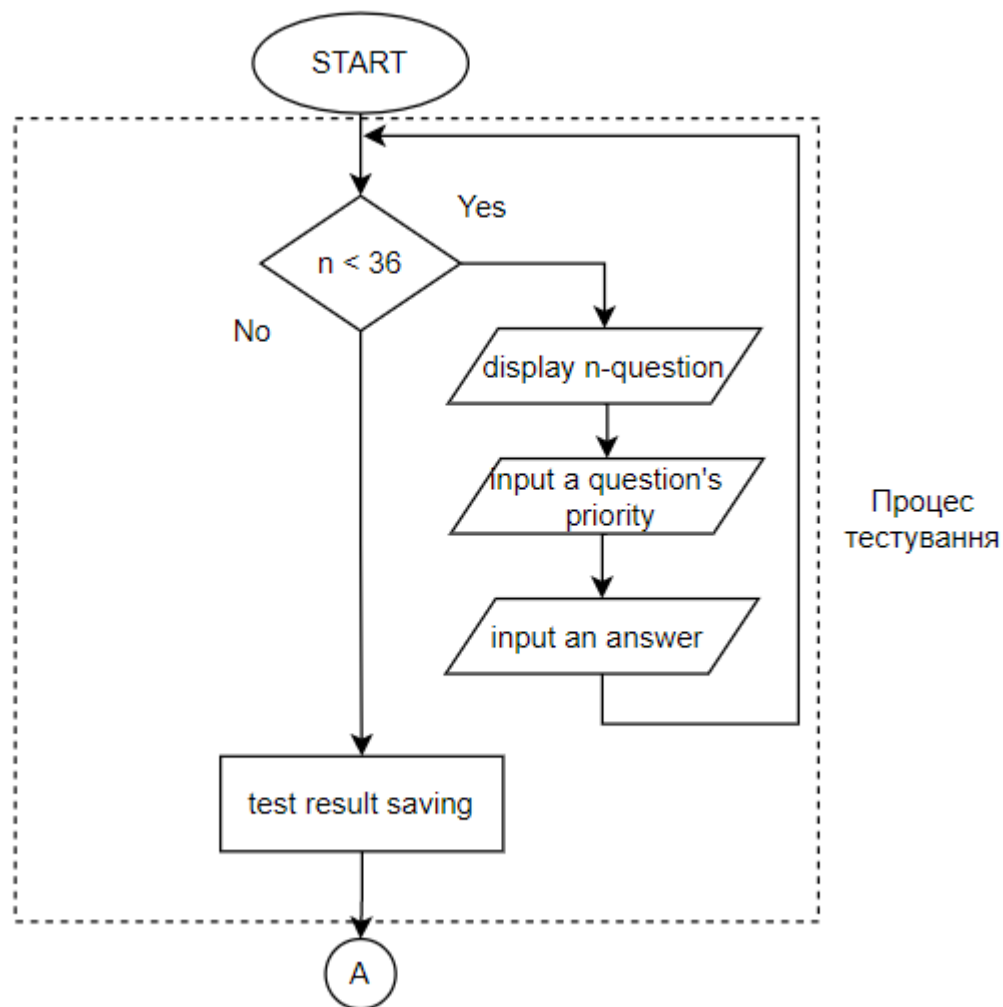


Рисунок 6.3 – Модернізований алгоритм колаборативної фільтрації з використанням середньозваженого значення (част.1)

На першій частині алгоритму зображений процес, що відповідає за психологічне тестування нових користувачів. Тестування складається з 36 питань,

які по черзі відображаються на дисплеї користувача. Також, кожному питанню надається можливість встановити пріоритет.

Ці пріоритети будуть використовуватися при формуванні рейтингу кандидатів, що надаються користувачеві у якості «рекомендації» для знайомства.

Після цього переходимо до процесу, де виділяємо до якої групи зі схожими інтересами відносити нового користувача.

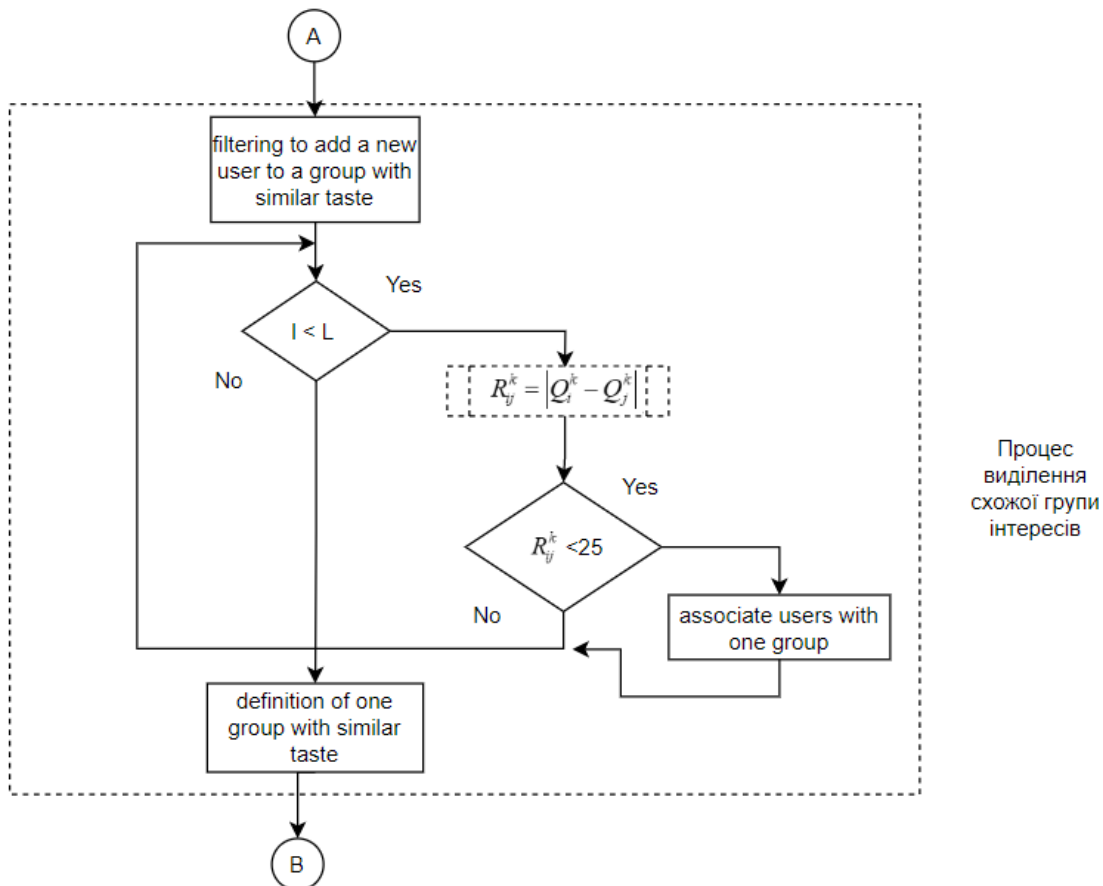


Рисунок 6.4 – Модернізований алгоритм колаборативної фільтрації з використанням середньозваженого значення (част.2)

Щоб віднести нового користувача до тієї чи іншої групи зі схожими інтересами, попарно порівнюємо результати психологічного тестування користувачів та обчислюємо для них коефіцієнт розбіжності. Якщо цей коефіцієнт менше 25 – відносимо користувачів до одної групи інтересів.

Отримуємо, що для нового користувача потенційними профілями для рекомендацій будуть ті, що вже вподобані іншими членами групи.

Останнім етапом є процес формування вибірки профілів для рекомендації користувачеві.

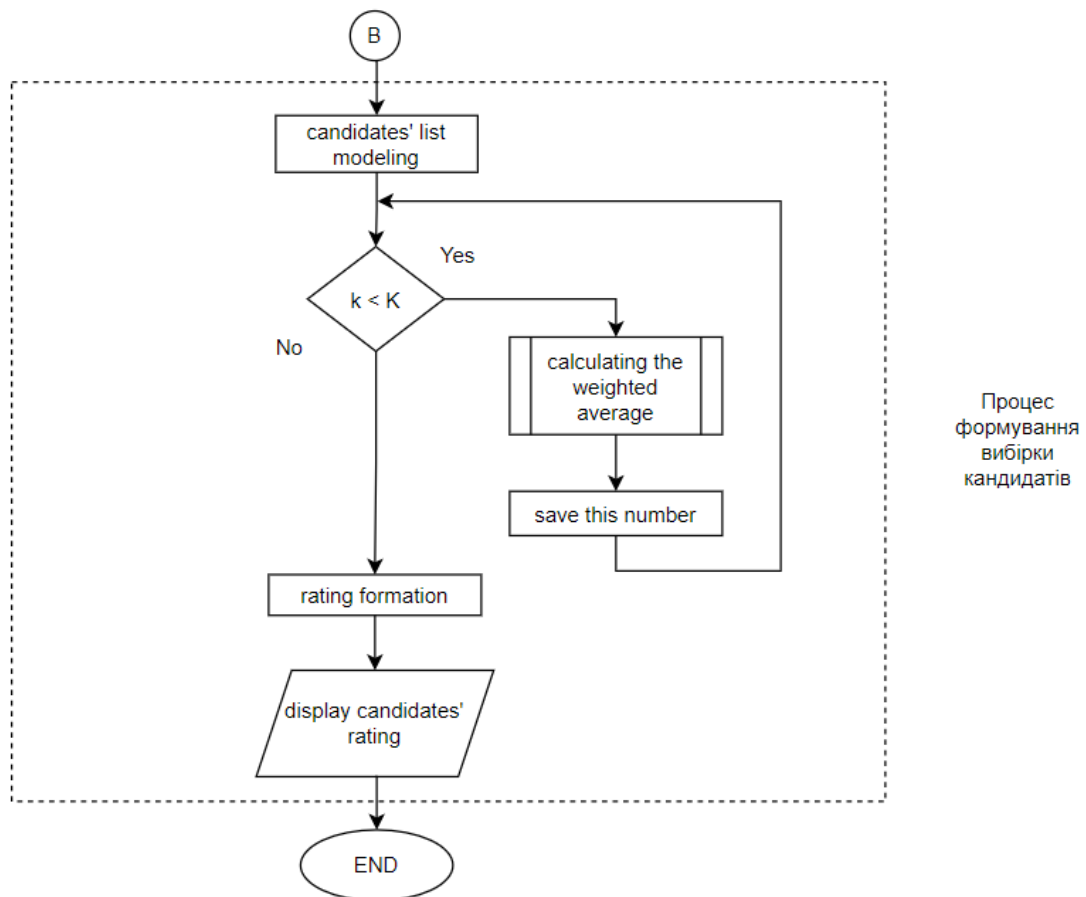


Рисунок 6.5 – Модернізований алгоритм колаборативної фільтрації з використанням середньозваженого значення (част.3)

Використовуючи середньозважене значення, формуємо рейтинг кандидатів/профілів для нашого нового користувача.

Після обчислення середньозваженого значення, яке базується на використанні тих самих пріоритетів, що задавалися користувачем під час проходження психологічного тестування, сортуємо отриманий список кандидатів від більшого до меншого значення. Чим більше середньозважене значення – тим більше кандидати підходять один одному.

Отриманий відсортований список профілів буде показаний користувачеві, після чого в нього буде можливість відправити запит на організацію побачення з тими кандидатами, які будуть йому найбільш цікавими.

ВИСНОВКИ

Кваліфікаційна робота присвячена вивченню та аналізу альтернативних методів підбору кандидатів в системах надання послуг з урахуванням уподобань клієнтів та вибір кращого з альтернатив для обраної предметної області, а саме – агенції знайомств.

Були проаналізовані існуючі методи, що використовуються в сучасних рекомендаційних системах, для формування списку так званих «можливих уподобань».

Детально було розглянуто алгоритм колаборативної фільтрації та недоліки цього методу. Щоб збільшити точність надання рекомендацій для нових користувачів системи, алгоритм було модифіковано за допомогою використання середньозваженого значення при формування рейтингу кандидатів.

Була проаналізована обрана предметна область, визначено поняття та сучасні методи для її реалізації. Також, були сформовані вимоги до майбутньої системи, на основі яких потім відбулося проектування системи.

Основна увага даної роботи приділена методам, що використовуються в сучасних рекомендаційних системах та розробці алгоритму підбору кандидатів, що використовується агенцією знайомств.

Результати дослідження в ході магістерської роботи представлено на міжнародній науково-практичній конференції “MODERN RESEARCH IN WORLD SCIENCE” 2022р. у вигляді опублікованих тезисів за темою «Використання колаборативної фільтрації в системах надання послуг з урахуванням уподобань клієнтів».

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Best Dating Apps of 2022. 2022. URL: <https://www.brides.com/best-dating-apps-5105328>
2. Daniel Johes. The 36 questions that leads to love 2015. URL: <https://www.nytimes.com/2015/01/09/style/no-37-big-wedding-or-small.html>
3. Алгоритм сортування. URL: <https://uk.wikipedia.org/wiki/sorting-algorythm>
4. Bubble sort. URL: https://en.wikipedia.org/wiki/Bubble_sort
5. Recommender system. URL: https://en.wikipedia.org/wiki/Recommender_system
6. G. Adomavicius and A. Tuzhilin. Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering 17. 2005. С. - 634–749.
7. Jones, M. T. (2013, December 12). Recommender systems. Introduction to approaches and algorithms. Retrieved November 25. 2017.
8. J. Bennet and S. Lanning. The Netflix Prize. KDD Cup and Workshop. 2007. URL: www.netflixprize.com.
9. Анализ популярных реляционных систем управления базами данных. 2022. URL: <https://drach.pro/blog/hi-tech/item/196-popular-relational-dbms-2022>
10. Переваги та недоліки .NET: швидкий розвиток, велика поширеність. 2022. URL: <https://dou.ua/lenta/articles/pros-and-cons-of-dotnet/>
11. Коннолли Т., Бегг К., Страчан А. База даних. Проектування, реалізація та супровід: Теорія і практика : Видавничий дім «Вільямс», 2003. 1436
12. Imangulova Z. Kolesnyk, L. An algorithm for building a project team considering interpersonal relations of employees // Eastern-European Journal of Enterprise Technologies. 2016. Vol. 6. P. 19-25.
13. СУБД: які бувають, як вибрати. 2022. URL: <https://highload.today/uk/subd-yaki-buvayut-yak-vibrati/>

14. Какую СУБД выбрать и почему? 2021. URL: <https://habr.com/ru/post/579248/>
15. Advantages and Disadvantages of ASP.NET. 2021. URL: <https://www.software-developer-india.com/advantages-and-disadvantages-of-asp-net/>
16. Recommender system. URL: https://en.wikipedia.org/wiki/Recommender_system
17. Введення в рекомендаційні системи. 2019. URL: <https://habr.com/ru/post/476222/>
18. Коннолли Т., Бегг К., Страчан А. База даних. Проектування, реалізація та супровід: Теорія і практика : Видавничий дім «Вільямс», 2003. 1436.
19. Логічна модель даних. URL: https://uk.wikipedia.org/wiki/%D0%9B%D0%BE%D0%B3%D1%96%D1%87%D0%BD%D0%B0_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C_%D0%B4%D0%B0%D0%BD%D0%B8%D1%85
20. Методичні вказівки щодо розробки та оформлення магістерської атестаційної роботи за спеціальністю 8.05010101 – Інформаційні управляючі системи та технології. Освітньо-кваліфікаційний рівень – магістр / Упоряд.: Левикін В.М., Міхнов Д.К., Саєнко В.І., Євланов М.В., Міхнова А.В., Керносів М.А. – Харків: ХНУРЕ, 2012. – 28 С.
21. 2. ДСТУ 3008:2015. «Звіти у сфері науки і техніки. Структура та правила оформлення» - Український інститут науково-технічної і економічної інформації; Технічний комітет стандартизації «Інформація і документація» (ТК 144) В. Земцева; Ю. Поліщук, канд. фіз.-мат. наук; Р. Санченко, канд. техн. наук; Л. Шрамко; А. Ямчук (науковий керівник) від 22 червня 2015 р. 20 с.
22. Колесник Л.В., Щетініна А.В. Використання колаборативної фільтрації в системах надання послуг з урахуванням уподобань клієнтів. IX Міжнародна науково-практична конференція “MODERN RESEARCH IN WORLD SCIENCE”, 28-30.11.2022 Львів, Україна
23. Колесник Л.В., Лещенко Д.Д. Актуальність рекомендаційної системи «Аптека» // Сучасні напрями розвитку інформаційно-комунікаційних технологій

та засобів управління : тези доповідей дванадцятої міжнародної науково-технічної конференції. Баку – Харків – Жиліна, 2022. Т.2. С.54.

24. Морозова А.І., Щетініна А.В. UML для бізнес-моделювання функціонування малого та середнього бізнесу. VI Міжнародна науково-практична конференція TOPICAL ISSUES OF MODERN SCIENCE, SOCIETY AND EDUCATION, 26-28.12.2021 Харків, Україна.

25. ISO 690:2010 Information and documentation — Guidelines for bibliographic references and citations to information resources. – 3th ed 2010-06-