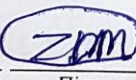
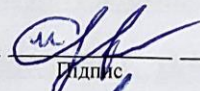
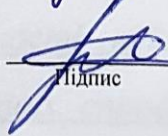


КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

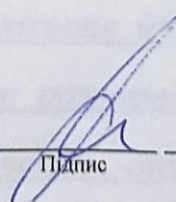
на тему Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних

Галузь знань 12 – Інформаційні технології
Шифр і назва галузі знань
Спеціальність 122 – Комп'ютерні науки
Шифр і назва спеціальності
Освітня програма Комп'ютерні науки
Назва освітньої програми

Виконав: студент групи КНм-23-1  Дмитро ЖУК
Група виконавця Підпис Ім'я, ПРІЗВИЩЕ
Керівник: к.т.н., доц. каф. КН  Олександр МАЗУРЕЦЬ
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ
Нормоконтроль: к.т.н., доц. каф. КН  Руслан БАГРІЙ
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

До захисту допускаю:

Зав. кафедри КН, д.т.н., професор

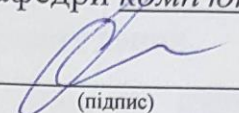


Олександр БАРМАК
Ім'я, ПРІЗВИЩЕ

16 грудня 2024 р.

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
Факультет інформаційних технологій
Кафедра комп'ютерних наук
Освітній ступінь магістр
Галузь знань 12 – Інформаційні технології
Спеціальність 122 – Комп'ютерні науки

ЗАТВЕРДЖУЮ
Завідувач кафедри комп'ютерних наук


(підпис)
д.т.н., професор Олександр БАРМАК
«02» вересня 2024 року

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ МАГІСТРА**

1. Тема кваліфікаційної роботи магістра: «Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних»
2. Завдання видано студенту Дмитру Жуку
(ім'я, прізвище)
3. Керівник роботи доцент кафедри КН Олександр МАЗУРЕЦЬ
(посада, ім'я, прізвище)
4. Затверджені наказом університету від «26» серпня 2024 р. № 60.
5. Зміст пояснювальної записки (перелік задач) та вихідні дані:
Мета кваліфікаційної роботи магістра – покращення визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних методами інтелектуальним аналізом анкетних даних, для чого слід розробити відповідний метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, спроектувати компоненти метода, фреймову модель подання анкетних даних та архітектуру відповідної нейронної мережі для ідентифікації сумісності клієнтів шлюбного агентства за їхніми текстовими описами, що використовує розроблений метод; після чого слід програмно реалізувати метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних у вигляді спроектованої програмної системи й дослідити його ефективність.

Реферат

Кваліфікаційна робота магістра розв'язує задачу покращення визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Результатом роботи є розроблений метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, що дозволяє визначати відсоток сумісності між парами клієнтів на основі їхніх анкетних даних.

Актуальність теми. У сучасному світі процес пошуку партнера для побудови міцних та довготривалих стосунків значно ускладнився через напружений темп життя, зростаючі соціальні вимоги та високий рівень міграції. У таких умовах, шлюбні агентства відіграють велику роль, допомагаючи людям шукати бажаних партнерів відповідно до їхніх вимог. Проте, традиційні методи підбору сумісних партнерів втрачають свою актуальність, через величезну кількість клієнтів, анкети яких неможливо проаналізувати вручну. Тому, автоматизація процесу підбору сумісних партнерів з використанням методів інтелектуального аналізу даних дає змогу обробляти запити від великої кількості клієнтів та знаходити підходящих партнерів з мінімальними часовими затратами та виключаючи суб'єктивні оцінки працівників традиційного шлюбного агентства.

Таким чином, в епоху цифровізації та постійного збільшення обсягів інформації та цифрових даних, покращення визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних є актуальним завданням. Це дасть змогу шлюбним агентствам обслуговувати більшу кількість клієнтів та значно розширити базу кандидатів для пошуку сумісностей.

Мета і задачі роботи. Метою роботи є покращення визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Для досягнення поставленої мети необхідно вирішити такі задачі:

- провести аналіз предметної області, пов'язаної із дослідженням діяльності шлюбних агентств та методами інтелектуального аналізу анкетних даних клієнтів шлюбних агентств;

- розробити метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних;

- обрати датасет для формування клієнтської бази;
- створити набір промптів для автоматизованої генерації текстових описів клієнтів через ChatGPT;
- спроектувати структуру веб-застосунку для пошуку партнера з високою сумісністю;
- спроектувати структуру даталогічної моделі бази даних веб-застосунку для пошуку партнера з високою сумісністю;
- виконати вибір засобів розробки для реалізації веб-застосунку для пошуку партнера з високою сумісністю;
- виконати програмну реалізацію веб-застосунку для пошуку партнера з високою сумісністю, який використовуватиме запропонований метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних;
- провести тестування розробленого застосунку для пошуку партнера з високою сумісністю;
- виконати дослідження ефективності запропонованого методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних використовуючи реалізований веб-застосунок.

Об’єкт дослідження. Процес визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

Предмет дослідження. Моделі, методи та засоби визначення сумісності суб’єктів.

Методи дослідження. Засоби інтелектуального аналізу даних, фреймове порівняння даних.

Наукова новизна одержаних результатів. Результати виконання кваліфікаційної роботи магістра містять наукову новизну, зокрема було якісно покращено визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Розроблений метод працює шляхом паралельного використання методу порівняння анкет за фреймовою моделлю та нейромережевого визначення оцінок схожості текстових описів клієнтів, було отримано відсотки сумісності між клієнтськими парами.

Розроблений метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, відрізняється від існуючих тим, що дозволяє порівнювати клієнтські анкети не лише за категоріальними даними, а й за текстовими описами. Це дає змогу значно глибше проаналізувати бажання клієнта та підібрати партнера з аналогічними поглядами, захопленнями та життєвими цілями.

Практичне значення одержаних результатів. Було створено інформаційну систему визначення сумісності клієнтів шлюбних агентств, що є прикладною програмною реалізацією методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, забезпечує користувача можливістю реєстрації та авторизації на веб-сайті шлюбного агентства з подальшою можливістю заповнення клієнтської анкети та пошуку сумісних партнерів з використанням розробленого методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

Інформаційна структура системи складається із трьох підсистем та бази даних: «Підсистема формування клієнтської бази шлюбного агентства», «Прикладна підсистема (веб-сайт шлюбного агентства)», «Підсистема визначення сумісності клієнтів шлюбного агентства», що дозволяють визначати сумісності клієнтів шлюбних агентств.

Відповідно до проведених досліджень було встановлено, що точність визначення оцінок подібності текстових описів у випадку сумісних пар становить 86%, в той час як для несумісних пар – 78%, загальна точність методу порівняння текстових описів становить 82%.

Апробація результатів кваліфікаційної роботи магістра та публікації. Основні наукові і практичні результати роботи доповідались в доповіді «Datalogic Model for Determining the Clients Compatibility Based on Questionnaires Data Analysis by Artificial Intelligence Means» на I Міжнародній науково-практичній конференції «Innovative Solutions in Science: Balancing Theory and Practice» (San Francisco, USA) 28-30 жовтня 2024 року та в доповіді «Підхід до визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних» на XVI

Всеукраїнській науково-практичній конференції «Актуальні проблеми комп'ютерних наук АПКН-2024» (м. Хмельницький) 15-16 листопада 2024 року. За темою кваліфікаційної роботи магістра автором виконано три наукових публікації:

1. Zhuk D., Mazurets O., Sobko O., Klimenko V. Datalogic Model for Determining the Clients Compatibility Based on Questionnaires Data Analysis by Artificial Intelligence Means. Innovative Solutions in Science: Balancing Theory and Practice. Proceedings of the 1st International scientific and practical conference. October 28-30, 2024. San Francisco, USA. 2024. Pp. 80-85.

2. Жук Д.І., Мазурець О.В., Кадинська В.Д., Тищенко О.О. Підхід до визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 208-214.

3. Жук Д.І., Мазурець О.В., Залуцька О.О., Денисенко Б.О. Реляційна персоніфікована даталогічна модель даних для визначення сумісності анкетних даних нейромережевими засобами. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2025. №1. (Довідка з редакції).

Структура та обсяг роботи. Кваліфікаційна робота магістра складається з реферату, завдання, змісту, переліку скорочень, вступу, 4 розділів, висновків, переліку посилань із 51 найменувань та 6 додатків. Обсяг основного тексту кваліфікаційної роботи магістра становить 87 сторінок. У роботі наведено 26 рисунків і 28 таблиць.

Ключові слова: шлюбне агентство, сумісність, інтелектуальний аналіз даних, клієнтська анкета.

Зміст

Перелік скорочень	3
Вступ.....	4
РОЗДІЛ 1 Аналіз предметної області визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних	8
1.1 Огляд особливостей діяльності шлюбних агентств та специфіки їхньої роботи	8
1.2 Огляд методів інтелектуального аналізу даних для визначення сумісності клієнтів шлюбних агентств за їхніми анкетами.....	11
1.3 Аналіз наукових публікацій з напрямку інтелектуального аналізу даних	13
1.4 Аналіз існуючого програмного забезпечення для визначення сумісності анкет клієнтів шлюбних агентств	15
1.5 Постановка задачі.....	19
РОЗДІЛ 2 Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних	21
2.1 Огляд процесу інтелектуального аналізу даних	21
2.2 Схема підходу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних	23
2.3 Опис методу ідентифікації сумісності клієнтів шлюбного агентства за фреймовою моделлю подання анкетних даних:	27
2.4 Опис методу нейромережевої ідентифікації сумісності клієнтів шлюбного агентства за їхніми текстовими описами.....	30
2.5 Особливості формування бази анкет кандидатів шлюбного агентства	32
2.6 Автоматизована генерація описів клієнтів на основі їхніх анкетних даних за допомогою промптів та ChatGPT	34
Висновки до розділу 2	35
РОЗДІЛ 3 Проектування інформаційної системи визначення сумісності клієнтів шлюбних агентств	37
3.1 Структура інформаційної системи визначення сумісності клієнтів шлюбних агентств	37

3.2 Функціональна структура інформаційної системи визначення сумісності клієнтів шлюбних агентств	39
3.2 Даталогічна модель бази даних інформаційної системи визначення сумісності клієнтів шлюбних агентств	42
3.4 Вибір засобів розробки для реалізації веб-сайту шлюбного агентства.....	53
Висновки до розділу 3	56
РОЗДІЛ 4 Дослідження ефективності методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних	57
4.1 Програмна архітектура інформаційної системи визначення сумісності клієнтів шлюбних агентств	57
4.2 Розробка прикладних компонентів інформаційної системи визначення сумісності клієнтів шлюбних агентств	60
4.3 Прикладне тестування інформаційної системи визначення сумісності клієнтів шлюбних агентств	66
4.4 Особливості використання веб-сайту шлюбного агентства	73
4.5 Дослідження ефективності роботи методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних	77
Висновки до розділу 4	80
Загальні висновки.....	82
Перелік посилань.....	84
Додатки	

Перелік скорочень

Скорочення, термін, позначення	Пояснення
KPM	Кваліфікаційна робота магістра
КН	Комп'ютерні науки
ПІБ	Прізвище, ім'я, по батькові
ІАД	Інтелектуальний аналіз даних
FNN	Feedforward Neural Networks
SNN	Siamese Neural Networks
NLP	Natural Language Processing
ШІ	Штучний інтелект
LSTM	Long short-term memory
CNN	Convolutional neural network
ПЗ	Програмне забезпечення
AI	Artificial intelligence
GPT	Generative Pre-trained Transformer
API	Application Programming Interface
БД	База даних
ID	Identity Document
FK	Foreign Key
SQL	Structured Query Language
IDE	Integrated Drive Electronics
ORM	Object Relational Mapper
HTTP	HyperText Transfer Protocol

Вступ

Кваліфікаційна робота магістра розв'язує задачу покращення визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Результатом роботи є розроблений метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, що дозволяє визначати відсоток сумісності між парами клієнтів на основі їхніх анкетних даних.

Актуальність теми. У сучасному світі процес пошуку партнера для побудови міцних та довготривалих стосунків значно ускладнився через напружений темп життя, зростаючі соціальні вимоги та високий рівень міграції. У таких умовах, шлюбні агентства відіграють велику роль, допомагаючи людям шукати бажаних партнерів відповідно до їхніх вимог. Проте, традиційні методи підбору сумісних партнерів втрачають свою актуальність, через величезну кількість клієнтів, анкети яких неможливо проаналізувати вручну. Тому, автоматизація процесу підбору сумісних партнерів з використанням методів інтелектуального аналізу даних дає змогу обробляти запити від великої кількості клієнтів та знаходити підходящих партнерів з мінімальними часовими затратами та виключаючи суб'єктивні оцінки працівників традиційного шлюбного агентства.

Таким чином, в епоху цифровізації та постійного збільшення обсягів інформації та цифрових даних, покращення визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних є актуальним завданням. Це дасть змогу шлюбним агентствам обслуговувати більшу кількість клієнтів та значно розширити базу кандидатів для пошуку сумісностей.

Мета і задачі роботи. Метою роботи є покращення визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Для досягнення поставленої мети необхідно вирішити такі задачі:

– провести аналіз предметної області, пов'язаної із дослідженням діяльності шлюбних агентств та методами інтелектуального аналізу анкетних даних клієнтів шлюбних агентств;

- розробити метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних;
- обрати датасет для формування клієнтської бази;
- створити набір промптів для автоматизованої генерації текстових описів клієнтів через ChatGPT;
- спроектувати структуру веб-застосунку для пошуку партнера з високою сумісністю;
- спроектувати структуру даталогічної моделі бази даних веб-застосунку для пошуку партнера з високою сумісністю;
- виконати вибір засобів розробки для реалізації веб-застосунку для пошуку партнера з високою сумісністю;
- виконати програмну реалізацію веб-застосунку для пошуку партнера з високою сумісністю, який використовуватиме запропонований метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних;
- провести тестування розробленого застосунку для пошуку партнера з високою сумісністю;
- виконати дослідження ефективності запропонованого методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних використовуючи реалізований веб-застосунок.

Об’єкт дослідження. Процес визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

Предмет дослідження. Моделі, методи та засоби визначення сумісності суб’єктів.

Методи дослідження. Засоби інтелектуального аналізу даних, фреймове порівняння даних.

Наукова новизна одержаних результатів. Результати виконання кваліфікаційної роботи магістра містять наукову новизну, зокрема було покращено визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Використовуючи розроблений метод, у вигляді підходу, що має на меті паралельне використання методу порівняння анкет за фреймовою моделлю та

нейромережевого визначення оцінок схожості текстових описів клієнтів, було отримано відсотки сумісності між клієнтськими парами.

Розроблений метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, відрізняється від існуючих тим, що дозволяє порівнювати клієнтські анкети не лише за категоріальними даними, а й за текстовими описами. Це дає змогу значно глибше проаналізувати бажання клієнта та підібрати партнера з аналогічними поглядами, захопленнями та життєвими цілями.

Практичне значення одержаних результатів. Було створено інформаційну систему визначення сумісності клієнтів шлюбних агентств, що є прикладною програмною реалізацією методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, забезпечує користувача можливістю реєстрації та авторизації на веб-сайті шлюбного агентства з подальшою можливістю заповнення клієнтської анкети та пошуку сумісних партнерів з використанням розробленого методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

Інформаційна структура системи складається із трьох підсистем та бази даних: «Підсистема формування клієнтської бази шлюбного агентства», «Прикладна підсистема (веб-сайт шлюбного агентства)», «Підсистема визначення сумісності клієнтів шлюбного агентства», що дозволяють визначати сумісності клієнтів шлюбних агентств. Відповідно до проведених досліджень було встановлено, що точність визначення оцінок подібності текстових описів у випадку сумісних пар становить 86%, в той час як для несумісних пар – 78%, загальна точність методу порівняння текстових описів становить 82%.

Апробація результатів кваліфікаційної роботи магістра та публікації. Основні наукові і практичні результати роботи доповідались в доповіді «Datalogic Model for Determining the Clients Compatibility Based on Questionnaires Data Analysis by Artificial Intelligence Means» на I Міжнародній науково-практичній конференції «Innovative Solutions in Science: Balancing Theory and Practice» (San Francisco, USA) 28-30 жовтня 2024 року та в доповіді «Підхід до визначення сумісності клієнтів

шлюбних агентств за інтелектуальним аналізом анкетних даних» на XVI Всеукраїнській науково-практичній конференції «Актуальні проблеми комп'ютерних наук АПКН-2024» (м. Хмельницький) 15-16 листопада 2024 року. За темою кваліфікаційної роботи магістра автором виконано три наукових публікації:

1. Zhuk D., Mazurets O., Sobko O., Klimenko V. Datalogic Model for Determining the Clients Compatibility Based on Questionnaires Data Analysis by Artificial Intelligence Means. Innovative Solutions in Science: Balancing Theory and Practice. Proceedings of the 1st International scientific and practical conference. October 28-30, 2024. San Francisco, USA. 2024. Pp. 80-85.

2. Жук Д.І., Мазурець О.В., Кадинська В.Д., Тищенко О.О. Підхід до визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 208-214.

3. Жук Д.І., Мазурець О.В., Залуцька О.О., Денисенко Б.О. Реляційна персоніфікована даталогічна модель даних для визначення сумісності анкетних даних нейромережевими засобами. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2025. №1. (Довідка з редакції).

Структура та обсяг роботи. Кваліфікаційна робота магістра складається з реферату, завдання, змісту, переліку скорочень, вступу, 4 розділів, висновків, переліку посилань із 51 найменувань та 6 додатків. Обсяг основного тексту кваліфікаційної роботи магістра становить 87 сторінок. У роботі наведено 26 рисунків і 28 таблиць.

РОЗДІЛ 1 Аналіз предметної області визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних

1.1 Огляд особливостей діяльності шлюбних агентств та специфіки їхньої роботи

Шлюбні агентства відіграють важливу роль у сучасному суспільстві, допомагаючи людям в пошуку партнерів для побудови відносин. Існування шлюбних агентств стає особливо актуальним в часи виникнення безлічі соціальних проблем, пов'язаних із важким темпом життя та постійно зростаючими вимогами до партнерів. Загалом, шлюбні агентства – це спеціалізовані організації, які допомагають в пошуках та підборі партнерів з метою створення сімейних зв'язків [1]. Їх головна мета – допомогти клієнтам знайти партнера з високим рівнем сумісності, враховуючи побажання обох сторін.

Необхідність існування шлюбних агентств обумовлена низкою соціальних проблем, які ускладнюють процес пошуку партнера з яким можна побудувати міцні сімейні зв'язки. Серед ключових факторів, які стають на заваді при пошуку партнера, є зайнятість на роботі, високі вимоги до партнера та місце проживання [2].

Сучасний темп життя, особливо у містах, спричиняє нестачу вільного часу, який можна було б присвятити побудові особистого життя, включаючи створення сім'ї. Надмірне навантаження на роботі, що забирає значну частину часу та особистих ресурсів, позбавляє людей можливості для знайомств, обмежуючи пошук потенційних майбутніх партнерів. У таких умовах звернення до послуг шлюбного агентства може стати ефективним рішенням, що дозволить знайти партнера, економлячи значну частину часу на пошуки.

Ще однією проблемою сучасності є підвищена вимогливість до партнерів, адже сучасне суспільство з кожним роком стає все більш вимогливим. Це пов'язано із значним розвитком та популяризацією психології, завдяки чому люди стають більш обізнані в тому як поводити себе в стосунках. Тому, людям значно важче знайти партнера, який буде відповідати всім поставленим якостям.

Місце проживання, також, має великий вплив на процес пошуку партнерів, що особливо загострилось під час війни, коли багато людей були змушені виїхати за кордон [3]. В таких умовах, звернення до шлюбних агентств дає можливість врахувати географічне розташування при підборі партнерів.

Основна особливість діяльності шлюбних агентств полягає в ретельному підборі кандидатів для клієнтів, враховуючи їхні вимоги та побажання. Цей процес може охоплювати як традиційні індивідуальні консультації із спеціалістами, так і сучасні системи інтелектуального аналізу даних, що здатні визначати рівень сумісності на основі інформації з клієнтських анкет. Крім того, шлюбні агентства можуть надавати деякі додаткові послуги, такі як: психологічна підтримка, сімейна терапія та навчання по покращенню соціальних навичок.

Шлюбні агентства можуть різнитись за спеціалізацією та методами підбору партнерів, тому їх поділяють на три основні типи: традиційні шлюбні агентства, нішеві шлюбні агентства та шлюбні агентства онлайн [4].

Традиційні шлюбні агентства роблять акцент на особистому контакті з клієнтом. Вони зазвичай мають офіс, в який клієнти можуть звертатись для отримання послуг. Фахівці шлюбного агентства проводять з клієнтами індивідуальні консультації та створюють профіль клієнта, що слугуватиме основою для підбору потенційних партнерів. Додатково, клієнти можуть заповнювати анкети про себе, вказуючи особисту інформацію, таку як вік, стать та рід діяльності. Після збору інформації про клієнта, спеціалісти підбирають кандидата або декількох кандидатів та влаштовують їм особисті зустрічі, на яких люди мають змогу познайомитись та дізнатись більше один про одного, що допоможе їм визначити чи підходять вони одне одному [5]. Основною перевагою традиційного шлюбного агентства є те, що всі питання, пов'язані з підбором кандидатів і організацією зустрічей бере на себе агентство, що дозволяє клієнту зекономити багато часу. Однак, такі послуги можуть виявитись досить дорогими та довготривалими через обмежену клієнтську базу, яка може бути на момент звернення.

Нішеві шлюбні агентства зосереджені на пошуку партнерів для окремих груп. Наприклад, такі агентства можуть обслуговувати лише заможних клієнтів, які

готові інвестувати великі суми для пошуку партнерів рівних за соціальним статусом. Також, нішеві шлюбні агентства можуть спеціалізуватися на підборі партнерів для клієнтів із певних релігійних груп, національностей чи культурних спільнот [6]. Ці агентства орієнтовані на індивідуальні особливості групи людей із якими вони працюють, тому й звернення до нішевих шлюбних агентств може бути корисним для людей із певних соціокультурних груп.

Інший тип шлюбних агентств – це шлюбні агентства онлайн, які працюють виключно в інтернет-просторі. Вони можуть бути представлені у вигляді веб-сайтів або мобільних додатків, де клієнти самостійно реєструються та заповнюють анкетні дані. Пошук потенційних партнерів відбувається за допомогою спеціальних алгоритмів, які аналізують анкетні дані користувачів і пропонують їм список кандидатів із високим рівнем сумісності [7]. Користувачі самі обирають з ким вони хотіли б познайомитися. Якщо, обоє з користувачів мають бажання дізнатись більше один про одного, вони можуть розпочати спілкування за допомогою вбудованих чатів.

Враховуючи соціальні проблеми сучасності, саме шлюбні агентства у онлайн форматі є найпопулярнішим та найефективнішим видом шлюбних агентств, через їхню простоту використання та доступність у будь-який час та незалежно від розташування. Крім цього, це досить економний варіант, адже користувачам подібних платформ доступний базовий безкоштовний варіант доступу до функціоналу та можливість придбання підписок різної вартості [8].

Важливим аспектом використання послуг шлюбних агентств онлайн є заповнення анкети користувача. Саме на підставі анкетних даних відбувається підбір потенційних партнерів з найвищим рівнем сумісності. Зазвичай анкета містить таку інформацію [9]:

- Персональні дані: ПІБ, вік, стать, місце проживання, освіта, професія. цей блок є базовим і включає інформацію про соціально-демографічний статус особи.

- Опис зовнішності: зріст, вага, колір очей та волосся, наявність татуювань. сюди також додаються особисті фото, які надають змогу створити візуальне уявлення про людину.

– Індивідуальні особливості: інтереси, хобі, життєві цінності та погляди. дані параметри є одними з найважливіших, адже вони слугують індивідуальним описом особистості.

– Вимоги до майбутнього партнера: даний блок містить перелік критеріїв, яким повинен відповідати партнер.

Отже, саме на основі анкетних даних відбувається пошук потенційних партнерів для клієнта шлюбного агентства. Оскільки, шлюбні агентства можуть мати великі клієнтські бази то виникає необхідність в автоматизації процесу обробки клієнтських анкет. Це допоможе швидко отримувати результати підбору партнерів та зробіть цей процес більш точним, завдяки використанню методів інтелектуального аналізу даних. Крім цього, через величезну популярність шлюбних агентств у онлайн форматі, виникає необхідність в розробці методів для автоматизації процесу пошуку партнера на основі анкетних даних.

1.2 Огляд методів інтелектуального аналізу даних для визначення сумісності клієнтів шлюбних агентств за їхніми анкетами

Для автоматизованого визначення сумісності клієнтів шлюбних агентств за їхніми анкетними даними необхідно використовувати сучасні методи інтелектуального аналізу даних. ІАД являє собою процес виявлення прихованих, невідомих або неочевидних закономірностей у великих обсягах даних на основі використання математичних моделей, статистичних методів чи алгоритмів штучного інтелекту [10]. Основною метою інтелектуального аналізу даних є перетворення необроблених даних на інформацію, яка може бути використана для прийняття рішень, прогнозування чи оптимізаційних процесів.

Для визначення сумісності клієнтів шлюбних агентств за допомогою статистичних методів інтелектуального аналізу даних можна застосовувати процес обчислення відповідності на основі числових та категоріальних даних з анкет, таких як вік, освіта, захоплення чи інтереси. Основна ідея полягає в тому, щоб оцінити схожість між людьми на основі порівняння ключових анкетних даних. Для цього,

необхідно порівнювати перелік вимог до потенційного партнера із анкетами інших клієнтів. Після порівняння зазначених даних можна підрахувати відсоток сумісності, тобто число, яке буде обчислено на основі співпадінь за конкретними категоріями. Таким чином, чим більше даних співпадатиме, тим вищим буде відсоток сумісності клієнтів шлюбного агентства [11].

Для інтелектуального аналізу анкетних даних клієнтів шлюбних агентств можна застосовувати методи та системи штучного інтелекту, включаючи сучасні моделі глибокого навчання, які здатні аналізувати різні типи даних (числові, категоріальні, текстові). Оскільки дане завдання полягає в порівнянні двох сутностей з метою визначення ступеня їхньої схожості, то для цього можна використовувати архітектуру нейронних мереж прямого поширення або сіамських нейронних мереж.

Нейронні мережі прямого поширення Feedforward Neural Networks (FNN) є базовими моделями глибокого навчання, які складаються з вхідного шару, кількох прихованих шарів та вихідного шару. Основне завдання FNN – знаходити залежності між вхідними даними, якими є анкети клієнтів шлюбного агентства, і цільовою змінною, тобто рівнем сумісності. Архітектура FNN добре підходить для обробки числових та категоріальних даних, з яких складається переважна більшість анкет клієнтів шлюбних агентств [12].

Для визначення схожості між об'єктами часто використовуються сіамські нейронні мережі Siamese Neural Networks (SNN). Особливістю SNN є наявність двох однакових підмереж, які відповідають за обробку вхідних даних. Це дозволяє проводити початковий аналіз об'єктів порівняння незалежно один від одного. Кожна підмережа генерує векторне представлення вхідних даних, після чого вектори порівнюються за допомогою функції порівняння для оцінки їхньої схожості [13].

Оскільки, анкети клієнтів шлюбних агентств можуть містити текстову інформацію, наприклад описи про себе або побажання щодо майбутнього партнера, то виникає потреба в аналізі такої інформації. Обробка текстових анкетних даних відноситься до сфери обробки природної мови. Обробка природної мови Natural Language Processing (NLP) – це процес автоматизованого аналізу людської мови

[14]. Методи NLP дозволяють виділяти корисну інформацію з текстів, наприклад, ключові слова, що можуть стати основою для подальшої роботи алгоритмів. Оскільки, детальні текстові розповіді про себе та описи бажаного партнера, часто містять важливу інформацію про особу, їхній автоматизований аналіз може суттєво допомогти у прийнятті рішень, щодо сумісності клієнтів шлюбних агентств.

Для вирішення NLP задач часто використовуються моделі глибокого навчання, наприклад Bidirectional Encoder Representations from Transformers (BERT). Ця модель розроблена компанією Google AI та базується на архітектурі трансформерів. Особливістю BERT є її здатність враховувати контекст слів у двох напрямках, як зліва на право так і справа на ліво. Це дозволяє отримати глибше представлення слів у тексті, особливо коли їхнє значення залежить від порядку використання в реченні [15].

Отже, для визначення сумісності клієнтів шлюбних агентств доцільно використовувати поєднання декількох підходів, що може дозволити не лише оптимізувати процес інтелектуального аналізу анкетних даних, а й отримати точніші результати, щодо сумісності клієнтів. Тому, статистичний аналіз числових та категоріальних даних може бути корисний при фільтрації анкет з метою формування групи потенційних партнерів, які відповідають зазначеним вимогам. Використання систем ШІ, зокрема моделі глибокого навчання BERT аналізу текстових описів, дасть змогу обчислити відсоток сумісності між клієнтами на основі текстової інформації з їхніх анкет.

1.3 Аналіз наукових публікацій з напряму інтелектуального аналізу даних

Важливим етапом в розробці методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних є аналіз наукових публікацій з напряму ІАД, в яких пропонуються методи визначення подібності між об'єктами. Нижче наведено огляд розглянутих досліджень.

У дослідженні [16] запропоновано метод нечіткого пошуку з використанням таблиці подібності символів, основна мета якого – підвищення швидкості та точності релевантних документів до заданої пошукової фрази. При пошуку відповідності, запропонований метод враховує наявність орфографічних помилок у пошуковій фразі та використання синонімів, скорочень або альтернативних форм слова. Для цього, пропонується використовувати таблицю подібності символів для врахування різних рівнів схожості між символами, що допомагає підвищити точність порівняння. Використання даного підходу дозволяє враховувати семантичну схожість символів навіть при наявності неточностей. Даний підхід може бути адаптований для визначення подібності текстових даних у клієнтських анкетах. Наприклад, таблиця подібності символів може бути використана для врахування помилок у текстових описах, написаних клієнтами (наприклад, розповідь про себе), оскільки тексти неформального типу дуже часто містять багато помилок.

У дослідженні [17] розглянуто використання сіамських нейронних мереж у системах рекомендацій, зокрема для аналізу схожості між об'єктами. Принцип роботи SNN базується на використанні однакових підмереж для вимірювання схожості між об'єктами. Відповідно до цього, схожі об'єкти матимуть меншу відстань між векторами ознак. В обраному дослідженні даний підхід використовується в контексті формування користувацьких рекомендацій на основі їхніх уподобань, проте, цю ідею можна використати й для створення переліку потенційних партнерів з високою сумісністю на основі порівняння схожості анкет клієнтів. Також, в дослідженні встановлено основні виклики, які виникають при використанні SNN, такі як потреба в використанні сучасних моделей глибокого навчання для вирішення більш складних завдань.

Наступне дослідження присвячене використанню нейронних мереж для визначення схожості текстів українською мовою [18]. В дослідженні було розглянуто та реалізовано дві моделі: LSTM та CNN. Для порівняння результатів був проведений експеримент, сенс якого полягав в аналізі текстів з використанням лематизації та без. Результати проведеного дослідження показали, що лематизація значно покращує процес аналізу та порівняння текстів. За метрикою «Ассурасу»,

точність аналізу лематизованих текстів склала 64%, в той час як нелематизованих – 61%. Отже, попередня обробка тексту є важливим етапом в процесі аналізу текстової інформації. Також, було встановлено, що модель LSTM показує точніші результати порівняння україномовних текстів.

Ще одне дослідження, пов'язане із сферою NLP, присвячене дослідженню ефективності семантичного аналізу текстів з використанням моделей BERT та DistilBER [19]. В роботі розглядається налаштування зазначених моделей для семантичного аналізу україномовних текстів. Для цього, моделі були додатково навчені на україномовних коментарях про ресторани, готелі, кафе, кінотеатри та інші громадські заклади. Для дослідження ефективності роботи навчених моделей були обчислені три метрики: Precision, Recall, F1-Score. В ході дослідження було встановлено, що модель DistilBER потребує значно менше обчислювальних ресурсів та є швидшою за модель BERT, проте загальна точність роботи для обох моделей склала 91%.

Отже, було проведено аналіз наукових публікацій з напряму інтелектуального аналізу анкетних даних користувачів з метою визначення їх сумісності. Було розглянуто підходи до порівняння текстових даних, для яких використовуються методи та системи штучного інтелекту, включаючи моделі глибокого навчання, що показують досить точні результати роботи із текстами.

1.4 Аналіз існуючого програмного забезпечення для визначення сумісності анкет клієнтів шлюбних агентств

У онлайн просторі існує безліч ресурсів для пошуку партнерів, від веб-сайтів до мобільних застосунків. Тому, для дослідження принципів роботи шлюбних агентств у форматі онлайн розглянуто веб-сайт шлюбного агентства «Mega Love» [20] (Рисунок 1.1).

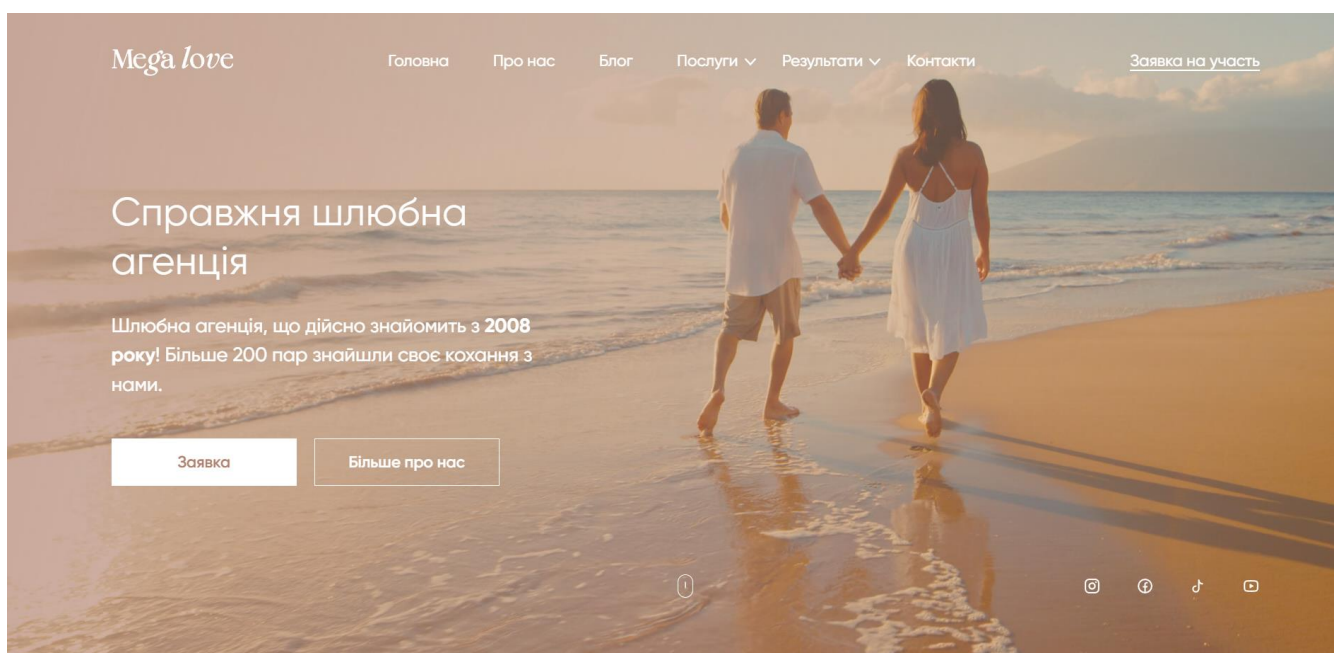


Рисунок 1.1 – Головна сторінка шлюбного агентства «Mega Love» [20]


Щоб скористатись послугами шлюбного агентства «Mega Love», клієнту потрібно подати заявку, вказавши свої персональні дані та контактну інформацію для зворотного зв'язку (Рисунок 1.2). Після цього, менеджери агентства зв'язуються з клієнтом для уточнення деталей. Далі клієнт отримує доступ до персонального кабінету в якому можна заповнити детальну анкету про себе та побажання щодо партнера.


Окрім цього, фахівці агентства проводять індивідуальні консультації з метою кращого розуміння клієнтів, та допомагають заповнювати анкети. Заповнені анкети передаються в обробку для партнерів із високою сумісністю. Якщо сформується пара в якій обоє клієнтів згодні на особисте знайомство із запропонованим партнером, тоді агентство бере на себе відповідальність за організацію зустрічі.

Таким чином, важливим етапом у процесі підбору партнера є заповнення анкети, яка включає як особисту інформацію, так і побажання щодо майбутнього партнера. Тому, під час розробки програмного продукту слід реалізувати зручний інтерфейс для заповнення анкетних даних.

Для знайомства напишіть нам

Залишіть заявку і наш клієнт-менеджер зв'яжеться з Вами

 Viber

 Telegram

Ваше ім'я та прізвище *	Ваше місто *
E-mail *	Ваш телефон (+380..) *
Дата народження (день.місяць.рік) *	Кількість дітей
Посилання на вашу соціальну мережу *	Telegram (Nickname / URL)
Звідки про нас дізналися *	
Додайте 5 ваших фотографій *	<div style="display: flex; gap: 5px;"> <div style="border: 1px solid #ccc; width: 25px; height: 25px; text-align: center; line-height: 25px;">+</div> <div style="border: 1px solid #ccc; width: 25px; height: 25px; text-align: center; line-height: 25px;">+</div> <div style="border: 1px solid #ccc; width: 25px; height: 25px; text-align: center; line-height: 25px;">+</div> <div style="border: 1px solid #ccc; width: 25px; height: 25px; text-align: center; line-height: 25px;">+</div> <div style="border: 1px solid #ccc; width: 25px; height: 25px; text-align: center; line-height: 25px;">+</div> </div>
<div style="border: 1px solid #ccc; padding: 10px; min-height: 60px;"> <p>Будь ласка, розкажіть дещо про себе</p> </div>	

Рисунок 1.2 – Форма для подачі заявки в шлюбне агентство «Mega Love» [20]

Недоліком звернення до традиційного шлюбного агентства, яким є шлюбне агентство «Mega Love», можна назвати те, що клієнт не матиме доступу до всього переліку сумісних партнерів, адже агентство пропонуватиме клієнту лише тих кандидатів, яких вибере самостійно. Таким чином, при підборі партнерів велику роль може зіграти суб'єктивність спеціалістів шлюбного агентства, що займатимуться підбором кандидатів. Це може призвести до неврахування безлічі важливої інформації. Саме тому автоматизований підбір сумісних партнерів є актуальним завданням.

Наступним прикладом існуючого ПЗ є мобільний застосунок для знайомств «Tinder» (Рисунок 1.3) [21]. «Tinder» – це один з найпопулярніших мобільних застосунків для знайомств онлайн, заснований у 2012 році. Через його велику популярність, даний застосунок має величезну клієнтську базу по всьому світу, що значно збільшує ймовірність знаходження бажаного партнера.

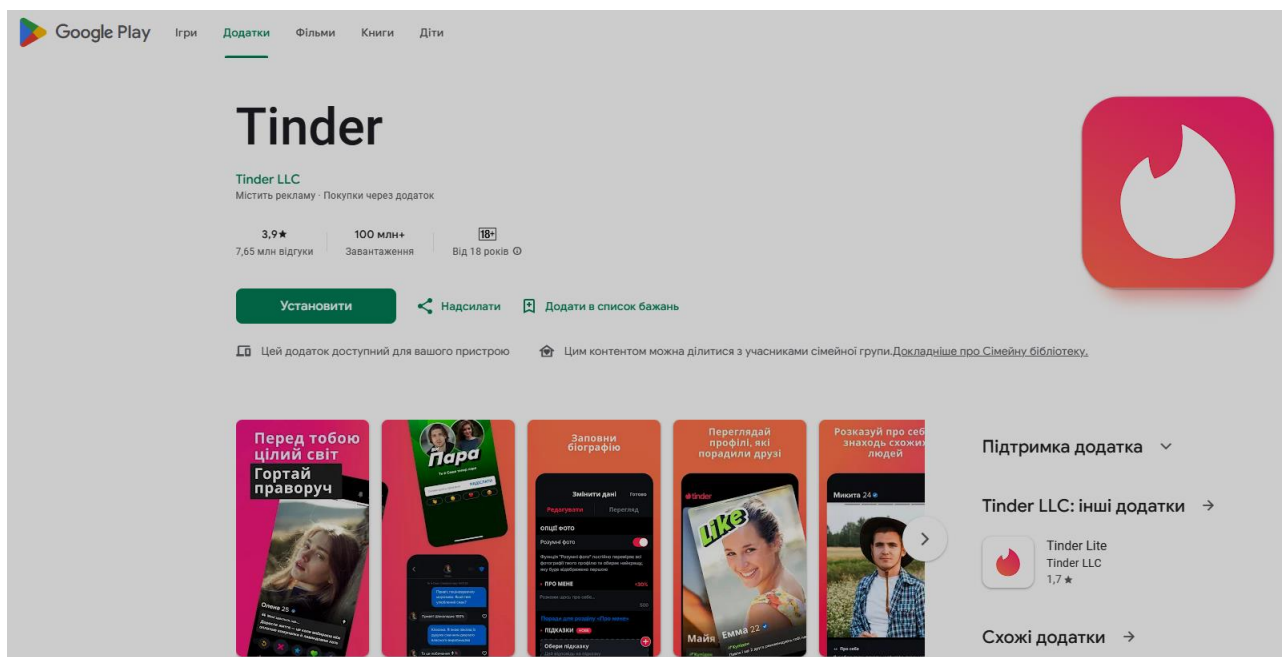


Рисунок 1.3 – Сторінка застосунку «Tinder» в Google Play [21]

Особливістю використання «Tinder» є його система свайпів (проведення пальцем по екрану), відповідно до якої свайп в ліворуч означає пропуск запропонованого партнера, а праворуч – вподобання (Рисунок 1.4). Якщо дві людини взаємно вподобали один одного їм відкривається можливість розпочати спілкування через внутрішній чат застосунку. Такий підхід значно пришвидшує та спрощує процес відбору партнерів.

Проте, така взаємодія є досить поверхневою, адже оцінка запропонованих партнерів відбувається переважно за їхніми фото. Це призводить до того, що інші ключові дані просто не враховуються. Також, існує велика ймовірність натрапити на фейкові профілі та шахраїв.

Також, недоліком даного застосунку є те, що тут не обчислюється ступінь сумісності користувачів, адже користувачам пропонуються усі існуючі профілі, які відповідають за статтю та вказаним місцем розташування. Для детального аналізу користувачі самі повинні переглядати анкети запропонованих партнерів та визначати чи підходять вони їм.



Рисунок 1.4 – Використання свайпів у «Tinder» [21]

Отже, було проведено аналіз існуючого програмного забезпечення для пошуку партнерів на прикладі веб-сайту шлюбного агентства та мобільного застосунку для знайомств. Проведений огляд дозволив встановити, що найважливішим етапом в пошуку партнера є процес заповнення анкети, адже саме за анкетними даними відбувається підбір партнерів. Також, на прикладі веб-сайту шлюбного агентства «Mega Love» встановлено, що підбір партнерів відбувається вручну спеціалістами агентства. Проте, такий підхід може викликати труднощі при розширенні клієнтської бази, адже чим більшу кількість даних потрібно проаналізувати тим більше часу та зусиль це забере у людини при ручному аналізі. Тому, автоматизація процесу визначення сумісності клієнтів шлюбного агентства є актуальним завданням у даній сфері. Крім цього, впровадження автоматизованого визначення сумісності користувачів мобільних застосунків та інших платформ схожих на «Tinder», допоможе користувачам зекономити час на пошуки партнера, не переглядаючи безліч нерелевантних пропозицій.

1.5 Постановка задачі

Метою кваліфікаційної роботи магістра є покращення визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, для чого необхідна розробка методу визначення сумісності клієнтів шлюбних агентств за

інтелектуальним аналізом анкетних даних та його прикладна реалізація у вигляді веб-сайту для пошуку партнерів.

Отож, щоб досягнути поставлену мету, необхідно виконати такі завдання:

- провести аналіз предметної області, пов'язаної із дослідженням діяльності шлюбних агентств та методами інтелектуального аналізу анкетних даних клієнтів шлюбних агентств;

- розробити метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних;

- обрати датасет для формування клієнтської бази;

- створити набір промптів для автоматизованої генерації текстових описів клієнтів через ChatGPT;

- спроектувати структуру веб-застосунку для пошуку партнерів;

- спроектувати структуру даталогічної моделі бази даних веб-застосунку для пошуку партнерів;

- виконати програмну реалізацію веб-застосунку для пошуку партнера з високою сумісністю, який використовуватиме запропонований метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних;

- провести тестування розробленого застосунку для пошуку партнера з високою сумісністю;

- виконати дослідження ефективності запропонованого методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних використовуючи реалізований веб-застосунок.

Практичним результатом виконання поставлених завдань є програмна реалізація методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних у вигляді веб-застосунку для пошуку партнерів, використовуючи який користувачі мають можливість заповнити анкети з можливістю автоматизованої генерації текстових описів на основі внесених анкетних даних, та подальшим пошуком партнерів з високою сумісністю.

РОЗДІЛ 2 Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних

2.1 Огляд процесу інтелектуального аналізу даних

Для розробки методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних необхідно використовувати сучасні методи інтелектуального аналізу даних. Процес інтелектуального аналізу даних складається з восьми послідовних етапів (Рисунок 2.1), починаючи зі збору та підготовки вхідних даних і закінчуючи впровадженням та використанням створеної моделі.



Рисунок 2.1 – Процес інтелектуального аналізу даних для визначення сумісності клієнтів шлюбних агентств

На першому етапі відбувається збір початкового набору даних, який потребує інтелектуального аналізу. Дані можуть бути зібрані з різних джерел: БД, текстових файлів, веб-сайтів або заздалегідь сформованих наборів (датасетів). В зібраних даних необхідно чітко розуміти значення кожного атрибута, їхній тип та формат, адже від цього залежить подальша обробка зібраної інформації.

Далі, зібрані дані необхідно підготувати до подальшого інтелектуального аналізу. Цей етап є одним із найважливіших, адже від якості вхідних даних залежатиме кінцевий результат. Підготовка даних може включати виявлення та видалення помилок, шуму чи відсутніх значень. Наприклад, в анкетах клієнтів шлюбних агентств можуть бути пусті поля, що ускладнюватиме підбір партнерів, тому виникає потреба в перевірці анкет на відсутність пустих полів та їхнє заповнення, за потреби.

Наступний етап полягає в перетворенні даних у формат, який більше підходитиме для аналізу. Це може включати агрегацію даних, тобто об'єднання кількох змінних у одну категорію для спрощення аналізу. Також, це може включати створення нових змінних на основі наявних або перетворення категоріальних змінних в числові (наприклад, якщо користувач вживає алкоголь, то ця інформація може бути записана числом «1», а якщо ні – «0»).

Коли дані готові для аналізу необхідно обрати метод, який найкраще підходитиме для вирішення конкретної задачі. Вибір методу залежить від типу даних, які потрібно аналізувати, мети дослідження та характеру проблеми. Інколи, використовується декілька методів та алгоритмів з подальшим порівнянням отриманих результатів, з метою вибору кращого варіанту. Зазвичай, для вирішення задачі інтелектуального аналізу даних використовуються моделі глибокого навчання.

Обрану модель потрібно адаптувати під вирішення конкретної задачі та навчити на відповідному наборі навчальних даних. Під час навчання модель аналізує зв'язки між вхідними даними та навчається знаходити закономірності.

Після проведення навчання необхідно провести оцінку точності роботи моделі. Модель оцінюється на даних, які не були використанні під час навчання, що дозволяє оцінити якість роботи моделі з незнайомою інформацією. Для цього, також, використовуються різні метрики, вибір яких залежить від обраної моделі.

Якщо навчання моделі досягло задовільного рівня то можна переходити до наступного етапу, який полягає в інтерпретації та візуалізації отриманих результатів. Для цього створюються різноманітні графіки та діаграми, які

допомагають краще зрозуміти отримані результати та прийняти подальше рішення щодо використання натренованого методу.

На фінальному етапі відбувається впровадження навченої та перевіреної моделі в проєкт, який потребує інтелектуального аналізу даних. Також, на цьому етапі відбувається відображення результатів роботи моделі у зрозумілій користувачу формі. Наприклад, результатом інтелектуального аналізу анкетних даних клієнтів шлюбних агентств буде перелік потенційних партнерів з найбільшою сумісністю.

Отже, для визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних потребує покрокового виконання вказаних етапів. Це дасть змогу отримати результат у вигляді переліку потенційних партнерів з високою сумісністю для цільового клієнта, на основі інтелектуального аналізу їхніх анкетних даних.

2.2 Схема підходу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних

Підхід для визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних має на меті паралельне застосування двох методів (Рисунок 2.2). У першому методі відбувається ідентифікація сумісності клієнтів шлюбного агентства за фреймовою моделлю подання анкетних даних, в той час як у другому методі – нейромережева ідентифікація сумісності клієнтів шлюбного агентства за їхніми текстовими описами.

Відповідно до першого методу, вхідними даними є: анкета цільового клієнта шлюбного агентства та база анкет кандидатів. Анкета цільового клієнта шлюбного агентства – це анкета клієнта для якого здійснюється підбір партнерів з високою сумісністю, використовуючи запропонований метод. База анкет кандидатів – це, фактично, клієнтська база шлюбного агентства, тобто перелік зареєстрованих клієнтів.

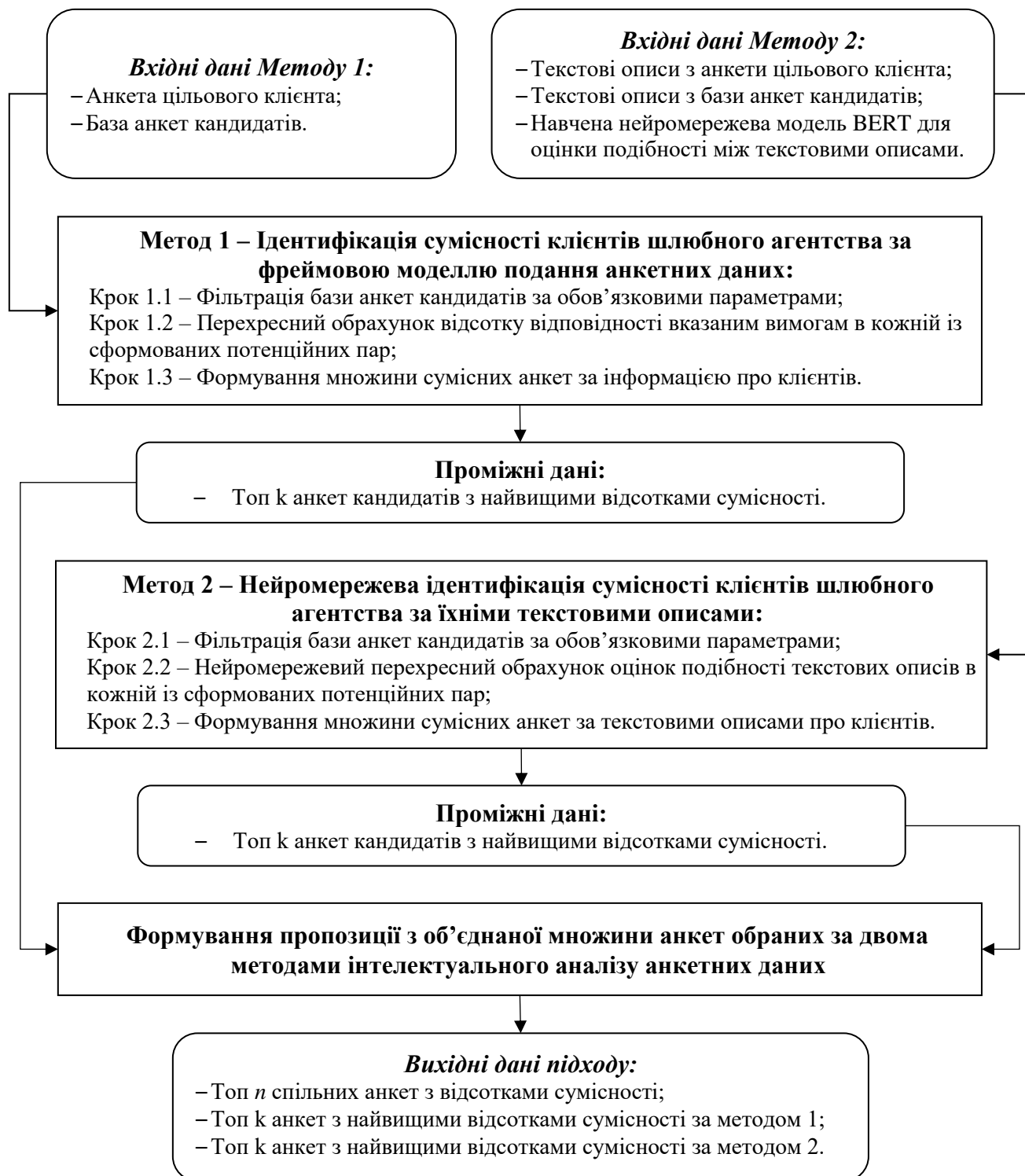


Рисунок 2.2 – Схема підходу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних

Оскільки, кожна клієнтська анкета містить перелік вимог до бажаного партнера, то на першому кроці обох методів відбувається фільтрація бази кандидатів за обов'язковими параметрами, які вказані в переліку вимог цільового клієнта. Це

необхідно для того, щоб скоротити кількість даних для подальшої обробки та відсіяти кандидатів, які однозначно не підійдуть цільовому клієнту. Наприклад, якщо клієнт шукає собі партнера протилежної статі, то немає необхідності визначати сумісність із кандидатами, які не задовольнятимуть дану вимогу.

Далі, відбувається перехресний обрахунок відсотку відповідності вказаним вимогам в кожній із сформованих пар. Тобто, кожен кандидат із клієнтської бази шлюбного агентства, який пройшов фільтрацію за обов'язковими параметрами, перетворюється на потенційного партнера для цільового клієнта. Відповідно до цього, формується певна кількість потенційних пар, яка відповідає кількості потенційних партнерів для цільового клієнта. Тоді, у кожній парі відбувається перехресне порівняння вимог до бажаного партнера із анкетною потенційного партнера. Таким чином, обчислюється відсоток відповідності не лише потенційного партнера відносно цільового клієнта, а й відсоток відповідності цільового клієнта відносно потенційного партнера. Це дає змогу враховувати відповідність вказаним вимогам відносно обох партнерів у потенційній парі.

Після цього, формується множина сумісних анкет за інформацією про клієнтів. Для цього, обчислюється середнє арифметичне між отриманими відсотками відповідності у кожній із сформованих потенційних пар. Отримане значення – це загальний відсоток сумісності у потенційній парі. Таким чином, в парах, в яких обоє партнерів мають високі відсотки відповідності зазначеним вимогами, будуть більші відсотки сумісності, ніж в парах, в яких ці значення сильно відрізняються. Також, для ідентифікації сумісності чи несумісності сформованої пари, прийнято встановити поріг в 50%. Тобто, якщо загальний відсоток сумісності перевищує 50%, то дана пара вважається сумісною.

Таким чином, використовуючи метод ідентифікації сумісності клієнтів шлюбного агентства за фреймовою моделлю подання анкетних даних, буде отримано множину сумісних анкет за інформацією про клієнтів. З цієї множини буде відібрано певну кількість анкет з найвищим відсотком сумісності, для проведення подальших досліджень.

Одночасно із першим методом буде відбуватись нейромережева ідентифікація сумісності клієнтів шлюбного агентства за їхніми текстовими описами. Вхідними даними для цього методу є: текстові описи з анкети цільового клієнта, текстові описи з бази анкет кандидатів, навчена нейромережева модель BERT для оцінки подібності між текстовими описами. Текстовими описами з анкети цільового клієнта є два текстові описи, перший з яких – це розповідь клієнта про себе, а другий – опис бажаного партнера. Аналогічні текстові описи має кожна анкета з бази анкет кандидатів.

Так само як і на попередньому методі, спочатку проводиться фільтрація за обов'язковими параметрами, з метою зменшення кількості даних для подальшого аналізу. Далі, відбувається перехресний обрахунок оцінок подібності текстових описів в кожній із сформованих потенційних пар використовуючи нейромережеву модель BERT. Відповідно до цього, порівнюється текстовий опис бажаного партнера від цільового клієнта, з розповіддю про себе від потенційного партнера. Таким чином, для кожної потенційної пари буде обчислено дві оцінки подібності текстових описів на основі яких буде обраховано загальний відсоток сумісності за формулою середнього арифметичного.

Отже, за результатами роботи другого методу буде сформовано множину сумісних анкет за текстовими описами про клієнтів, на основі якої будуть відібрані анкети з найвищим відсотком сумісності.

По завершенню роботи обох методів буде отримано два переліка анкет з найвищими відсотками сумісності (наприклад топ 10 анкет за кожним із методів). На основі цього буде сформовано пропозицію з об'єднаної множини анкет обраних за двома методами інтелектуального аналізу анкетних даних. Таким чином, в дану пропозицію потраплять тільки ті анкети, які є в переліках результатів за обома методами.

Таким чином, вихідними даними запропонованого підходу будуть:

- топ n спільних анкет з відсотками сумісності;

– топ k анкет з найвищими відсотками сумісності відповідно до методу ідентифікації сумісності клієнтів шлюбного агентства за фреймовою моделлю подання анкетних даних;

– топ k анкет з найвищими відсотками сумісності відповідно до методу нейромережевої ідентифікація сумісності клієнтів шлюбного агентства за їхніми текстовими описами.

Отже, було розроблено підхід для визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, який використовує паралельну роботу двох методів ідентифікації сумісності клієнтів. Даний підхід дозволяє оцінити роботу обох методів та отримати об'єднану множину анкет з високими відсотками сумісності для цільового клієнта.

2.3 Опис методу ідентифікації сумісності клієнтів шлюбного агентства за фреймовою моделлю подання анкетних даних:

Даний метод має на меті формування множини сумісних анкет за інформацією про клієнта, використовуючи фреймову модель подання анкетних даних. Відповідно до фреймової моделі, фрейм – це структура даних, яка використовується для організації інформації у зручному для обробки вигляді. Основною ознакою фрейму є наявність слотів або комірок, кожна з яких зберігає певний тип інформації [22]. У даному контексті, фреймом є абстрактне представлення анкети клієнта, що складається з комірок у яких зберігається інформація про клієнта шлюбного агентства (наприклад, для комірки «Вік» значенням буде число, яке відповідатиме віку клієнта).

У запропонованому методі, фреймова модель подання анкетних даних використовується для формування множини сумісних анкет за інформацією про клієнтів. Це відбувається шляхом порівняння комірок відповідних фреймів. Для цього, на основі кожної клієнтської анкети формується два фрейми: фрейм з інформацією про клієнта та фрейм з вимогами до бажаного партнера.

Отже, спочатку необхідно провести фільтрацію бази анкет кандидатів на відповідність обов'язковим параметрам. Таким чином, значно скоротиться кількість інформації, яка підлягатиме обробці, що полегшить процес інтелектуального аналізу анкетних даних, адже після фільтрації переважна більшість анкет відкидатиметься. Наприклад, якщо клієнт шукає дівчину віком від 20 до 25 років, то усі кандидати, які не задовольнятимуть цю умову не будуть братись до уваги.

Для проведення фільтрації бази анкет кандидатів необхідно встановити певні правила за якими буде сформовано вибірку потенційних партнерів. Для початку, до уваги будуть братись три основні параметри: стать, вік та розташування. Тобто, якщо у клієнта є чітка вимога до одного із цих параметрів, або до усіх трьох, то в вибірку потраплять лише ті кандидати, які повністю відповідають цим вимогам. Параметр «стать» повинен мати 100% збіг із зазначеною вимогою, адже, якщо клієнт шукає собі партнера лише протилежної статі, немає необхідності розглядати анкети, які не підходять за гендерною приналежністю. Вік також є важливим критерієм який має мати точний збіг, адже, зазвичай люди шукають собі партнера зі своєї вікової категорії і інші вікові групи їх не цікавлять [23]. Розташування не є обов'язковим параметром, який повинен мати 100% збіг, але тільки у випадку якщо воно не вказано у вимогах. Якщо у вимогах клієнта до партнера вказано конкретне місто, то це означає що клієнт шукає партнера лише зі свого міста і не має сенсу пропонувати людей, які не відповідають зазначеній геолокації.

Крім цих трьох важливих параметрів, перелік вимог до партнера може бути наповнений іншою додатковою інформацією, наприклад, вимогами до зовнішності чи інтересів. І хоч всі ці вимоги можуть бути важливими для клієнта, та знайти людину яка на 100% відповідала б зазначеним критеріям практично неможливо або шанси на це дуже малі. Крім цього, один кандидат може повністю відповідати клієнту з точки зору зовнішності але мати зовсім інші цінності та інтереси, або навпаки. Тому, виникає необхідність в обчисленні відсотку відповідності вимогам до бажаного партнера із анкетними даними потенційних партнерів, які пройшли фільтрацію за трьома основними параметрами. Цей відсоток буде обраховуватись

простим статистичним методом порівняння двох фреймів, відповідно до якого, чим більша кількість параметрів співпадатиме, тим вищим буде відсоток відповідності.

Оскільки, відсоток відповідності не може перевищувати 100%, то для кожного параметру необхідно обчислити скільки відсотків нараховуватиметься у випадку збігу. Для цього, спочатку обчислюється загальна кількість вказаних клієнтом параметрів у переліку вимог до бажаного партнера. Також, необхідно враховувати наявність списків у певних параметрах, наприклад список хобі. Таким чином, загальна формула для обчислення відсотку відповідності матиме наступний вигляд:

$$C = \sum_n \frac{\binom{1}{n}}{k} \cdot 100\%, \quad (2.1)$$

де n – загальна кількість вказаних параметрів, k – кількість варіантів вибору у параметрах у форматі списків.

Таким чином, якщо загальна кількість параметрів вказаних у вимогах до бажаного партнера становить 5, то за кожен збіг буде нараховуватись 20%. Проте, якщо одним із параметрів є перелік хобі в кількості 10 штук, то за кожен збіг по хобі буде нараховано по 2%.

Отже, в результаті проведених обчислень буде отримано відсоток відповідності потенційного партнера зазначеним вимогам, шляхом порівняння двох фреймів – фрейму з переліком вимог до бажаного партнера та фрейму з особистою інформацією потенційного партнера. Оскільки, важливо враховувати відповідність вимогам відносно обох партнерів з потенційної пари, то даний перелік операцій необхідно проводити відносно обох партнерів із пари. Таким чином буде отримано два відсотки відповідності вказаним вимогам для однієї потенційної пари клієнтів шлюбного агентства. За отриманими відсотками відповідності буде обчислено загальний відсоток сумісності партнерів для кожної потенційної пари.

2.4 Опис методу нейромережевої ідентифікації сумісності клієнтів шлюбного агентства за їхніми текстовими описами

Метод нейромережевої ідентифікації сумісності клієнтів шлюбного агентства за їхніми текстовими описами має на меті обчислення оцінки схожості текстових описів клієнтів на основі порівняння їхніх векторних представлень. Саме для формування векторних представлень текстових описів використовується нейромережева модель BERT.

Метод нейромережевої ідентифікації сумісності клієнтів шлюбного агентства за їхніми текстовими описами зображено на рисунку 2.3. Цей метод складається з трьох компонентів: два вхідні тексти для порівняння (опис бажаного партнера та розповідь потенційного партнера про себе), навчена нейромережева модель BERT для формування векторного представлення вхідних текстів та функція порівняння, якою обрано косинусну відстань (Рисунок 2.3).

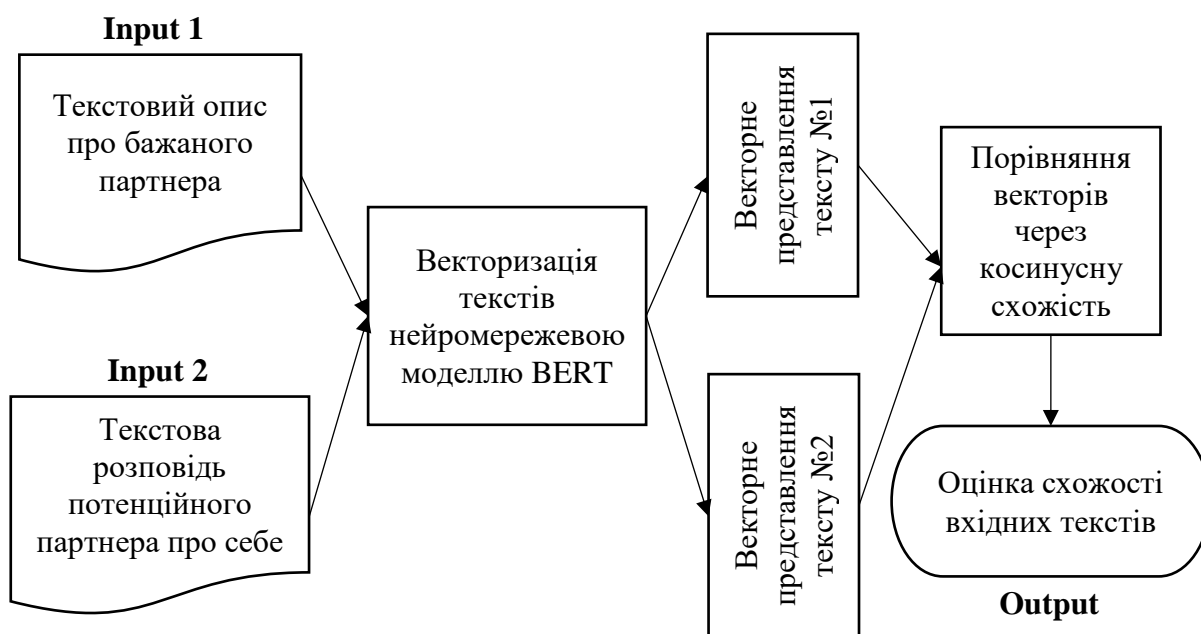


Рисунок 2.3 – Схема методу нейромережевої ідентифікації сумісності клієнтів шлюбного агентства за їхніми текстовими описами

Отже, для формування векторного представлення вхідних текстових описів використовується нейромережева модель BERT. Загалом, векторизація – це процес

перетворення будь-якої інформації у числовий формат [24]. У моделі BERT процес векторизації тексту відбувається через перетворення кожного слова або токена на багатовимірний вектор, який зберігає його значення відносно всього тексту. Процес векторизації тексту складається з декількох етапів: токенизація, додавання спеціальних токенів, перетворення в ідентифікатори, маскуванню послідовностей [25].

Для токенизації модель BERT використовує вбудований токенизатор на базі токенизатора WordPiece [26]. Токенизація – це процес розбиття текстового масиву на окремі елементи – токени. Також, токенизація застосовується для розбиття невідомих слів на частини [27]. Це дає змогу моделі BERT працювати зі словами, які відсутні в її словнику.

Крім цього, у текст додаються спеціальні токени для розділення. Спочатку додається токен для позначення початку тексту. Далі, додаються токени для розділення речень та позначення кінця тексту.

Після цього, перетворені в токени слова, перетворюються в числові ідентифікатори на основі словника моделі. Саме числові ідентифікатори визначають важливість слів із яких складається текст.

На завершення, відбувається маскуванню створеної послідовності. Тобто, якщо довжина тексту менша за максимально допустиму, то текст доповнюється масками. Це необхідно для того, щоб обробляти тексти різної довжини.

Обчислення оцінки подібності векторизованих текстових описів відбувається з використанням функції порівняння. Для цього, прийнято рішення використати косинусну схожість (Cosine Similarity), яка дозволяє обчислювати подібність між двома векторами використовуючи косинус кута між ними [28]. Косинусна схожість часто використовується для порівняння текстових даних, які представлені у вигляді векторів. Перевагою використання косинусної схожості є те, що обчислення не залежать від довжини вхідних текстів, яка може відрізнятись.

Формула для обчислення косинусної схожості для двох векторних представлень текстів A та B має наступний вигляд [28]:

$$CS = \cos(\theta) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \cdot \|\vec{B}\|}, \quad (2.2)$$

де $\vec{A} \cdot \vec{B}$ – скалярний добуток векторних представлень текстів A та B , $\|\vec{A}\|$ і $\|\vec{B}\|$ – довжини (норми) векторів A та B , θ – кут між векторами.

В результаті порівняння двох векторних представлень через косинусну схожість можна отримати число в діапазоні від «-1» до «1», де:

- «-1» – вхідні тексти повністю протилежні;
- «0» – у вхідних текстах відсутня подібність;
- «1» – вхідні тексти повністю схожі.

За виконанням описаної послідовності кроків буде отримано оцінку подібності двох текстових описів клієнтів. Для кожної потенційної пари описаний процес відбувається двічі, з метою врахування оцінок подібності текстових описів відносно обох партнерів. Далі, на основі отриманих оцінок подібності обчислює відсоток сумісності для кожної потенційної пари. Таким чином, формується множина сумісних анкет за текстовими описами про клієнтів.

Отже, використання нейромережевої моделі BERT для векторизації вхідних текстових описів забезпечує глибоке розуміння контексту слів. Обчислення подібності сформованих векторів через косинусну відстань дозволяє порівнювати вхідні тексти та отримувати відсоткову оцінку схожості текстових описів.

2.5 Особливості формування бази анкет кандидатів шлюбного агентства

Для забезпечення роботи методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних необхідно сформувати базу анкет кандидатів, тобто клієнтську базу шлюбного агентства.

Анкета клієнта шлюбного агентства містить велику кількість полів, які потребують заповнення. Сюди належить як особиста інформація (ім'я, стать, вік) так і зовнішні характеристики (ріст, вага) та захоплення (перелік хобі). Для заповнення анкет клієнтів шлюбного агентства та формування клієнтської бази будуть використанні дані із датасету, знайденого у вільному доступі на ресурсі Kaggle [29], та доповнені згенерованими даними через Faker.

Отже, для формування клієнтської бази буде використано датасет «OkCupid Profiles» [30]. OkCupid – це міжнародна платформа для знайомства, заснована американською компанією [31]. У датасеті зібрано інформацію про профілі користувачів цієї платформи. Проте, задля уникнення порушення конфіденційності та проблем із законодавством, у даному датасеті відсутня персональна інформація, така як ПІБ користувача та його контактні дані. Загалом, датасет містить 60 тисяч записів, що дає змогу сформувати 60 тисяч анкет клієнтів. Проте, така кількість анкет є занадто великою для використання в межах навчального проєкту і для демонстрації роботи методу та дослідження його ефективності достатньо сформувати близько 5 тисяч клієнтських анкет.

В обраному датасеті міститься така інформація про користувачів платформи: вік, сімейний статус, стать, орієнтація, статура, дієта, вживання алкоголю, вживання наркотиків, куріння, освіта, етнічна приналежність, ріст, дохід, професія, місце проживання, наявність дітей, наявність домашніх улюбленців, мови спілкування та знак зодіаку. Загалом, датасет містить 19 полів, переважна більшість яких є категоріальними, тобто користувачі можуть обирати підходящі варіанти із доступного переліку.

Оскільки, в датасеті відсутня інформація про імена клієнтів, їхні контактні дані, перелік інтересів та інші індивідуальні особливості, то виникає необхідність згенерувати цю інформацію за допомогою Faker. Faker – це спеціалізований програмний інструмент для генерації несправжніх даних, які можуть використовуватись для наповнення бази даних тестовими зразками [32]. Faker здатен генерувати дані, які будуть схожі на реальні, проте не порушуватимуть конфіденційність осіб, адже дані будуть вигадані та підібрані випадковим чином.

За допомогою Faker клієнтські анкети будуть доповнені відсутніми даними, такими як: ПІБ, дата народження, номер телефону, електронна пошта та перелік хобі. Використовуючи Faker можна генерувати не лише запропоновані дані із бібліотеки, а й доповнювати їх власними переліками. Наприклад, можна скласти власний перелік інтересів та доповнювати ними анкети клієнтів шлюбного агентства.

Таким чином, створено тисячу анкет клієнтів шлюбного агентства у рівній кількості для обох статей, використовуючи дані із датасету «OkCupid Profiles». Відсутні дані будуть згенеровані за допомогою інструменту Faker. Створені анкети будуть використані для апробації методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

2.6 Автоматизована генерація описів клієнтів на основі їхніх анкетних даних за допомогою промптів та ChatGPT

Оскільки, в обраному датасеті, на основі якого формуватимуться клієнтські анкети, немає текстових описів про клієнтів, виникає потреба в використанні технологій автоматизованої генерації текстової інформації. Для цього буде використано модель ChatGPT [33], яка інтегрується в програмний продукт через API OpenAI [34]. Використання API дозволяє відправляти промпти до моделі ChatGPT та отримувати відповіді.

Загалом, промпти – це текстові запити, які служать інструкціями для моделей штучного інтелекту, за допомогою яких можна отримати бажану інформацію. За допомогою промпту можна вказати безліч деталей на основі яких буде згенеровано відповідь [35]. Наприклад, через промпт можна задати тон (формальний, жартівливий, сумний) очікуваної відповіді, її величину та інші специфічні вимоги, які враховуватимуться під час генерації. Таким чином, промпти використовуються для адаптації використовуваної моделі під вирішення конкретної задачі.

Для автоматизованої генерації текстових описів клієнтів необхідно створити два промпти, які слугуватимуть шаблонами для отримання персоналізованих текстів про клієнта та його бажаного партнера. При формування промпту важливо враховувати як персональну інформацію клієнта так і його індивідуальні риси. Тому, ідеальний промпт повинен включати усю інформацію з анкети клієнта.

Отже, промпт для генерації текстового опису клієнта про себе на основі його персональних даних буде таким: *«Склади текстовий опис для профілю клієнта шлюбного агентства. Вік: {вік}, інтереси: {інтереси}, професія: {професія}*

Опиши його унікальні риси, хобі та життєві погляди». За допомогою цього промпту буде отримано текстовий опис клієнта, що враховуватиме його персональні характеристики.

За схожим підходом буде відбуватись генерація текстових описів бажаних партнерів на основі переліку вимог до партнера. Таким чином, другий промпт буде таким: *«Склади опис бажаного партнера для клієнта шлюбного агентства, відповідно до вимог та опису клієнта. Вік: {вік}–{вік}, інтереси: {інтереси}, професія: {професія} Опис клієнта: {текстовий опис клієнта}. Опиши особисті якості, які цінує клієнт»*. Варто відмітити, що при генерації опису бажаного партнера буде братись до уваги не лише перелік вимог, а й опис самого клієнта, адже важливо щоб у потенційної пари збігались інтереси та життєві цінності.

Використання ChatGPT для автоматизованої генерації текстових описів клієнтів шлюбного агентства дозволяє отримати якісні текстові описи з мінімальними часовими витратами. ChatGPT здатний генерувати тексти, які дуже схожі на природну мову [36]. Враховуючи особисті дані клієнтів, згенеровані тексти відображатимуть їхні унікальні емоційні та поведінкові характеристики, що особливо актуально в контексті пошуку партнерів та визначенні сумісності, адже це потребує врахування безлічі персональних деталей. Також, це допоможе уникнути банальності та зробить текстові описи більш унікальними.

Отже, автоматизована генерація текстових описів клієнтів є ключовим кроком на етапі формування клієнтських анкет. Використання ChatGPT через API OpenAI дозволяє інтегрувати обрану технологію в програмний продукт та генерувати текстові описи, використовуючи сформовані промпти. Автоматизована генерація текстових описів спрощує процес заповнення анкети, дозволяючи швидко створювати необхідні описи, які відображають індивідуальні особливості клієнта.

Висновки до розділу 2

Було розроблено метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних на основі підходу, який має на меті

паралельне застосування двох методів. Відповідно до першого методу, ідентифікація клієнтів шлюбного агентства відбувається на основі порівняння анкетних даних у вигляді фреймів. У другому методі ідентифікація сумісності клієнтів шлюбного агентства відбувається шляхом нейромережевого порівняння текстових описів клієнтів.

Метод ідентифікації клієнтів шлюбного агентства за фреймовою моделлю подання анкетних даних передбачає перехресний обрахунок відсотків відповідності партнера зазначеному переліку вимог. Для цього, з кожної клієнтської анкети було сформовано два фрейми: фрейм з особистою інформацією про клієнта та фрейм з переліком вимог до бажаного партнера. Результатами роботи даного методу є два відсотки відповідності для кожної із потенційних пар клієнтів, за якими обчислено загальний відсоток сумісності. Для оцінки схожості векторних представлень вхідних текстів використано косинусну відстань, яка дозволяє отримати числове значення схожості текстів. В результаті роботи даного методу було отримано дві оцінки схожості текстових описів для кожної потенційної пари клієнтів. На основі отриманих оцінок було обчислено загальний відсоток сумісності для кожної пари.

Також, було розглянуто особливості формування бази анкет кандидатів. Для цього було використано поєднання даних із датасету «OkCupid Profiles» зі згенерованими випадковими даними через інструмент Faker. Таким чином, було сформовано тисячу клієнтських анкет у рівній пропорції для обох статей. Оскільки, в обраному датасеті немає детальних текстових описів клієнтів, то було використано модель ChatGPT для автоматизованої генерації текстових описів про клієнтів на основі їхніх анкетних даних. Для цього було сформовано два промпти: один для генерації розповіді користувача про себе, другий – для опису бажаного партнера.

Отже, відповідно до запропонованого підходу, було сформовано дві множини анкет із високою сумісністю відносно цільового клієнта. Також, було сформовано пропозицію з об'єднаної множини анкет, обраних за двома методами інтелектуального аналізу анкетних даних.

РОЗДІЛ 3 Проектування інформаційної системи визначення сумісності клієнтів шлюбних агентств

3.1 Структура інформаційної системи визначення сумісності клієнтів шлюбних агентств

Структуру інформаційної системи визначення сумісності клієнтів шлюбних агентств поділено на три підсистеми та базу даних (Рисунок 3.1). Перша підсистема необхідна для реалізації функціоналу пов'язаного із наповненням клієнтської бази шлюбного агентства. Друга підсистема – це прикладна підсистема у вигляді веб-сайту шлюбного агентства, який використовуватиме запропонований метод для підбору партнерів. Третя підсистема – це, власне, реалізація методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. База даних необхідна, в першу чергу, для збереження клієнтських анкет та всієї допоміжної інформації, необхідної для функціонування веб-сайту.

В підсистемі формування клієнтської бази шлюбного агентства реалізовано функціонал, який відповідає за формування клієнтських анкет, використовуючи дані з датасету та згенеровані дані через Faker. Також, тут відбувається процес автоматизованої генерації текстових описів клієнтів про себе та бажаного партнера використовуючи промпти до ChatGPT.

Наступна підсистема є прикладною підсистемою у вигляді веб-сайту шлюбного агентства, тому тут зосереджено весь функціонал пов'язаний із роботою веб-сайту. У цій підсистемі реалізовано можливість реєстрації нових клієнтів шлюбного агентства та можливість авторизації для зареєстрованих клієнтів. Також, тут відбувається процес заповнення клієнтської анкети з можливістю вибору та редагування автоматично згенерованих текстових описів. Після збереження заповненої анкети, клієнту доступний перегляд введених даних на своїй персональній сторінці. Також, після збереження анкетних даних клієнту відкривається можливість до пошуку партнерів з високою сумісністю. Пошук партнерів здійснюється через розроблений метод визначення сумісності клієнтів

шлюбних агентств за інтелектуальним аналізом анкетних даних. Після проведення пошуку, клієнту доступний перегляд переліку знайдених партнерів з обчисленими відсотками сумісності.



Рисунок 3.1 – Схема структури інформаційної системи визначення сумісності клієнтів шлюбних агентств

У підсистемі визначення сумісності клієнтів шлюбного агентства використовується запропонований підхід. Тому, тут зосереджено функціонал, пов'язаний із програмною реалізацією методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Саме тут відбувається формування двох множин сумісних анкет та формування пропозиції з

об'єднаної множини анкет обраних за двома методами інтелектуального аналізу анкетних даних.

Не менш важливим елементом структури інформаційної системи для визначення сумісності клієнтів шлюбних агентств є база даних. У базі даних будуть зберігатись сформовані клієнтські анкети. Також, у БД будуть зберігатись категоріальні дані, необхідні для заповнення анкет (наприклад, перелік хобі). Це дасть можливість клієнтам швидко заповнювати свої анкети, вибираючи потрібну інформації з наданого переліку доступних варіантів. Також, це допоможе уникнути зайвих помилок при формуванні анкет.

Отже, сформовано структуру інформаційної системи для визначення сумісності клієнтів шлюбних агентств, яка складається з трьох взаємопов'язаних підсистем та бази даних. Кожна підсистема має зв'язок із БД, адже потребує як отримання даних із БД, так і збереження даних в БД. Цей підхід дозволяє створити гнучку структуру інформаційної системи, яку можна легко змінювати за потреби.

3.2 Функціональна структура інформаційної системи визначення сумісності клієнтів шлюбних агентств

Функціональна структура інформаційної системи визначення сумісності клієнтів шлюбних агентств складається з окремих груп функцій, які забезпечують користувача веб-сайту шлюбного агентства необхідним функціоналом. Таким чином, функціональна структура веб-сайту шлюбного агентства складається з трьох груп функцій: група функцій для автентифікації у системі, група функцій для роботи з клієнтською анкетною, група функцій для пошуку сумісних партнерів.

Перш за все варто розглянути групу функцій для автентифікації у системі, яка відповідає за доступ до такого функціоналу як реєстрація та авторизація (Рисунок 3.2). Без даної групи функцій, користувач веб-сайту не зможе стати клієнтом шлюбного агентства, адже не зможе створити свій особистий профіль.

Також, дана група функцій включає можливість відновлення та зміни паролю, що є дуже корисним у випадку коли користувач забув свій пароль для

входу або хоче його змінити. Крім цього, у дану групу входить можливість виходу з особистого профілю та видалення профілю. Видалення профілю призводить до видалення усіх даних про клієнта із БД, тому його анкета не буде аналізуватись при пошуку сумісних партнерів.

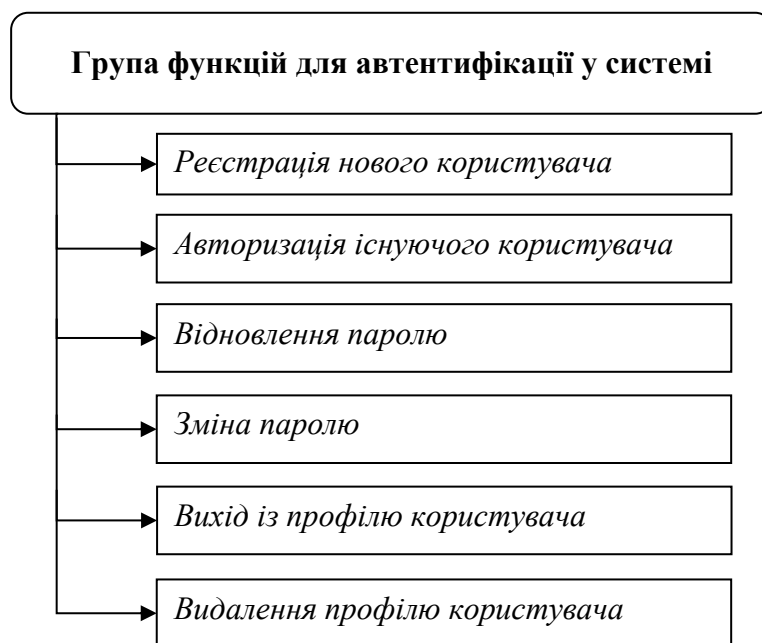


Рисунок 3.2 – Група функцій для автентифікації в системі

Наступною групою функцій є група функцій для роботи з клієнтською анкетною. Дана група функцій включає в себе функціонал, пов'язаний із процесом заповнення клієнтської анкети. Сюди входять такі функції: заповнення клієнтської анкети, перегляд клієнтської анкети, редагування клієнтської анкети, приховання клієнтської анкети, видалення клієнтської анкети (Рисунок 3.3). Також, дана група функцій включає в себе можливість автоматизованої генерації текстового опису на основі анкетних даних.

Для початку клієнту доступна можливість заповнення клієнтської анкети особистою інформацією та вимогами до бажаного партнера. Далі, клієнт має змогу переглянути заповнену анкету та відредагувати за потреби. Також, клієнт може видалити свою анкету, що призведе до видалення усіх особистих даних із бази даних. Таким чином, клієнт без заповненої анкети не буде братись до уваги під час пошуку сумісних партнерів. Проте, якщо клієнт бажає тимчасово обмежити

можливість аналізу його анкетних даних, він може приховати свою анкету. Таким чином, анкетні дані залишатимуться в БД, але не будуть враховуватись при аналізі.

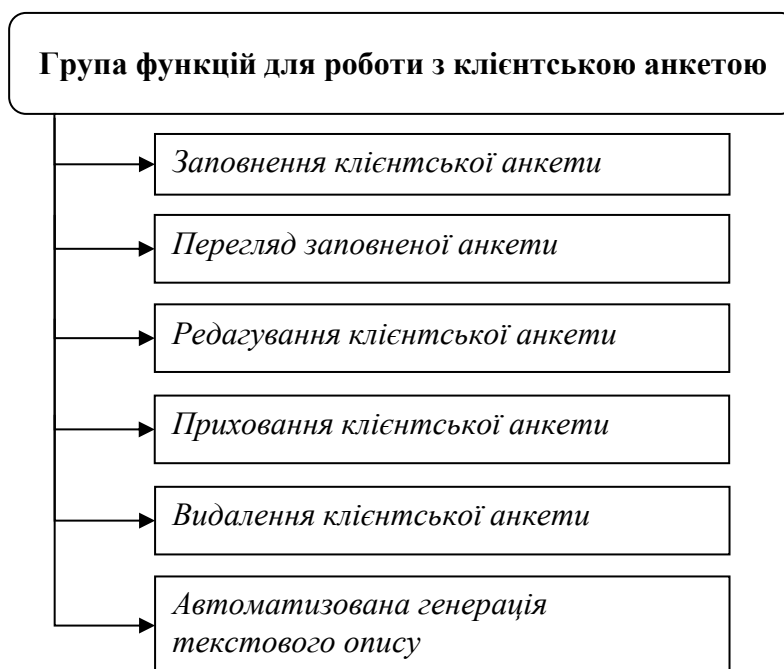


Рисунок 3.2 – Група функцій для роботи з клієнтською анкетною

Останньою групою функцій є група функцій для пошуку сумісних партнерів, яка включає функціонал, пов'язаний із пошуком бажаних партнерів. Дана група функцій використовує розроблений метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Тож, в групі функцій для пошуку сумісних партнерів міститься такий функціонал: пошук сумісних партнерів, перегляд множини сумісних партнерів за методом порівняння фреймів, перегляд множини сумісних партнерів за нейромережевим порівнянням текстових описів, перегляд об'єднаної пропозиції сумісних партнерів за обома методами, перегляд профілю запропонованого кандидата (Рисунок 3.3).

Маючи доступ до описаного функціоналу, користувач веб-сайту шлюбного агентства має можливість проводити пошук сумісних партнерів та переглядати результати у вигляді трьох переліків із можливими кандидатами. Перші два переліки формуються за допомогою двох методів, в той час як третій перелік – це об'єднана пропозиція кандидатів за обома методами. Користувач може переглядати

результати визначення сумісності для кожного кандидата окремо, та переходити на його особистий профіль для перегляду його анкетних даних.

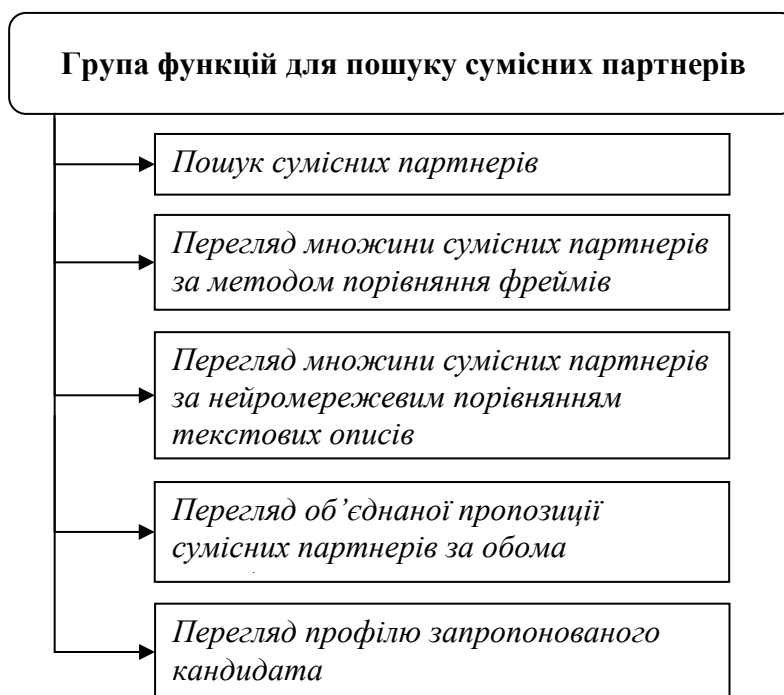


Рисунок 3.3 – Група функцій для пошуку сумісних партнерів

Таким чином, реалізація описаних функцій забезпечує користувача можливістю реєстрації та авторизації на веб-сайті шлюбного агентства з подальшою можливістю заповнення клієнтської анкети та пошуку сумісних партнерів з використанням розробленого методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

3.2 Даталогічна модель бази даних інформаційної системи визначення сумісності клієнтів шлюбних агентств

Було спроектовано структуру даталогічної моделі бази даних для інформаційної системи визначення сумісності клієнтів шлюбних агентств (Рисунок 3.2). БД складається із двадцяти трьох таблиць: Users, Profiles, Ideal_profiles, Speaks, Languages, Signs, Religions, Pets, Offspring, Ethnicities, Educations, Smokes, Drugs,

Drinks, Diets, Body_types, Orientations, Sex, Statuses, Jobs, Cities, Hobbies, Hobbies_lists.

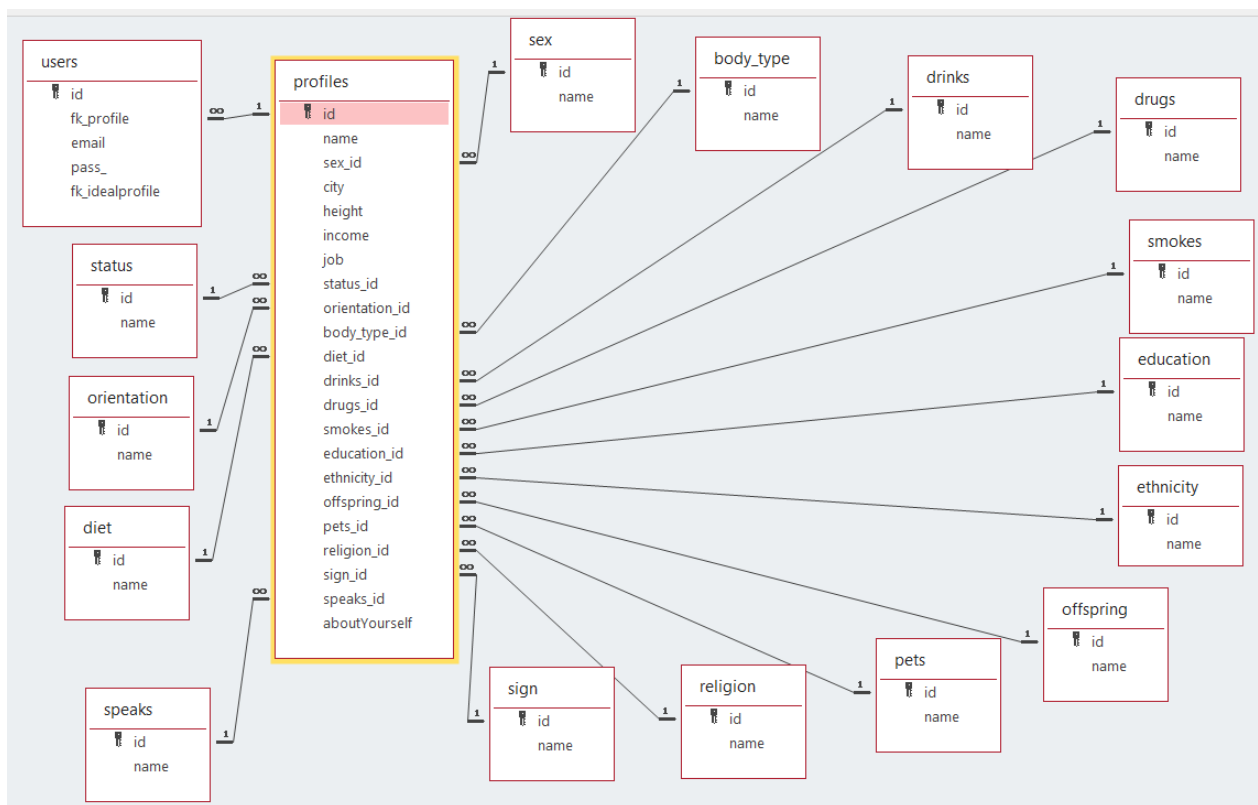


Рисунок 3.2 – Даталогічна модель БД інформаційної системи визначення сумісності клієнтів шлюбних агентств

Оскільки, основним функціоналом будь-якого подібного веб-сайту є можливість реєстрації та авторизації користувачів, то важливо створити таблицю в якій зберігалася б інформація про дані для автентифікації. Тому, було створено таблицю «Users» з такими полями: ID, Name, Surname, Email, Password, FK_profile, FK_ideal_profile (Таблиця 3.1).

Поле «ID» є унікальним ідентифікатором користувача. Це означає, що кожен користувач має свій унікальний номер, за допомогою якого його можна легко ідентифікувати у базі даних. Поля «Name» та «Surname» зберігають ім'я та прізвище користувача. В полі «Email» зберігаються дані про електронну пошту користувача до якої буде прив'язано особистий профіль. Це дасть змогу користувачеві відновлювати пароль у випадку, якщо користувач його забуде. Також, саме через

електронну пошту відбувається повторний вхід в особистий профіль. Поле «Password» зберігає пароль, який користувач придумав під час реєстрації. Пароль, також, використовується при повторному вході в систему. Поля «FK_profile» та «FK_ideal_profile» зберігають посилання на частини клієнтської анкети. Поле «FK_profile» – це особисті дані про клієнта шлюбного агентства, а «FK_ideal_profile» – перелік вимог до бажаного партнера. Таким чином, збереження цієї інформації дозволяє сформувати особистий профіль клієнта шлюбного агентства.

Таблиця 3.1 – Опис атрибутів таблиці «Users»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор клієнта шлюбного агентства
2.	Name	text	Ім'я клієнта шлюбного агентства
3.	Surname	text	Прізвище клієнта шлюбного агентства
4.	Email	text	Електронна пошта для реєстрації на веб-сайті шлюбного агентства
5.	Password	text	Пароль для автентифікації на веб-сайті шлюбного агентства
6.	FK_profile	int	Профіль користувача (заповнена анкета з персональною інформацією)
7.	FK_idealprofile	int	Вимоги до бажаного партнера

Кожному зареєстрованому користувачеві необхідно заповнити анкету клієнта шлюбного агентства, в якій потрібно вказати особисту інформацію. Для цього було створено таблицю «Profiles», яка включає перелік таких полів: ID, Age, FK_sex, FK_city, Height, Weight, Income, FK_job, FK_status, FK_orientation, FK_body_type, FK_diet, FK_drinks, FK_drugs, FK_smokes, FK_education, FK_ethnicity, FK_offspring, FK_pets, FK_religion, FK_sign, About_yourself (Таблиця 3.2).

Поле «ID» є унікальним ідентифікатором клієнтської анкети. Поля «Age», «Height», «Weight» та «Income» є числовими та зберігають числові дані про клієнта, такі як вік, зріст, вага та місячний дохід. Отже, при заповненні клієнтської анкети, користувач може ввести в дані поля лише числа.

Таблиця 3.2 – Опис атрибутів таблиці «Profiles»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор клієнтської анкети
2.	Age	int	Вік клієнта
3.	FK_sex	int	Стать клієнта
4.	FK_City	int	Місто в якому проживає клієнт
5.	Height	int	Ріст клієнта
6.	Weight	int	Вага клієнта
7.	Income	int	Дохід клієнта
8.	FK_job	int	Професія клієнта
9.	FK_status	int	Сімейний статус клієнта
10.	FK_orientation	int	Орієнтація клієнта
11.	FK_body_type	int	Тип статури клієнта
12.	FK_diet	int	Дотримання клієнтом дієт або харчових звичок (наприклад, вегетаріанство)
13.	FK_drinks	int	Вживання клієнтом алкогольних напоїв
14.	FK_drugs	int	Вживання клієнтом наркотичних речовин
15.	FK_smokes	int	Ставлення клієнта до куріння
16.	FK_education	int	Освіта клієнта
17.	FK_ethnicity	int	Етнічна приналежність клієнта
18.	FK_offspring	int	Наявність дітей у клієнта або ставлення до них (наприклад, за відсутності дітей клієнт може вказати, що планує їх у майбутньому або не хоче мати дітей взагалі)
19.	FK_pets	int	Наявність домашніх улюбленців у клієнта або ставлення до них (наприклад, клієнт може вказати, що не любить тварин)
20.	FK_religion	int	Релігія клієнта
21.	FK_sign	int	Знак зодіаку клієнта
22.	About_yourself	text	Текстовий опис клієнта (розповідь про себе)

Для формування переліку вимог до бажаного партнера створено таблицю «Ideal_profiles» з такими атрибутами: ID, FK_sex, FK_city, Min_age, Max_age, Min_height, Max_height, Min_weight, Max_weight, Min_income, Max_income, FK_job, FK_status, FK_orientation, FK_body_type, FK_diet, FK_drinks, FK_drugs, FK_smokes,

FK_education, FK_ethnicity, FK_offspring, FK_pets, FK_religion, FK_sign, About_partner (Таблиця 3.3).

Таблиця 3.3 – Опис атрибутів таблиці «Ideal_profiles»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор переліку вимог до бажаного партнера
2.	FK_sex	int	Стать бажаного партнера
3.	FK_city	int	Місто в якому проживає бажаний партнер
4.	Min_age	int	Мінімальний вік бажаного партнера
5.	Max_age	int	Максимальний вік бажаного партнера
6.	Min_height	int	Мінімальний ріст бажаного партнера
7.	Max_height	int	Максимальний ріст бажаного партнера
8.	Min_weight	int	Мінімальна вага бажаного партнера
9.	Max_weight	int	Максимальна вага бажаного партнера
10.	Min_income	int	Мінімальний дохід бажаного партнера
11.	Max_income	int	Максимальний дохід бажаного партнера
12.	FK_job	int	Професія бажаного партнера
13.	FK_status	int	Сімейний статус бажаного партнера
14.	FK_orientation	int	Орієнтація бажаного партнера
15.	FK_body_type	int	Тип статури бажаного партнера
16.	FK_diet	int	Дотримання бажаним партнером дієт або харчових звичок
17.	FK_drinks	int	Вживання бажаним партнером алкогольних напоїв
18.	FK_drugs	int	Вживання бажаним партнером наркотичних речовин
19.	FK_smokes	int	Ставлення бажаного партнера до куріння
20.	FK_education	int	Освіта бажаного партнера
21.	FK_ethnicity	int	Етнічна приналежність бажаного партнера
22.	FK_offspring	int	Наявність дітей у бажаного партнера або ставлення до них
23.	FK_pets	int	Наявність домашніх улюбленців у бажаного партнера або ставлення до них
24.	FK_religion	int	Релігія бажаного партнера
25.	FK_sign	int	Знак зодіаку бажаного партнера
26.	About_partner	text	Текстовий опис бажаного партнера

Далі, виникає необхідність зберігання безлічі даних, таких як стать, освіта, професія. Всі ці дані відносяться до категоріального типу, тобто потребують створення окремих таблиць. Таблиця «Cities» містить два поля «ID» та «Name». Для «ID» обрано цілочисловий тип даних, а для «Name» рядковий тип для запису назви міста (Таблиця 3.4).

Таблиця 3.4 – Опис атрибутів таблиці «Cities»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор міста (або іншого населеного пункту)
2.	Name	text	Назва міста

Таблиця «Sex» потрібна для збереження даних про стать клієнтів шлюбного агентства. Поле «ID» слугує унікальним ідентифікатором статі. Поле «Name» потрібне для збереження назви статі та відображення на сторінці (Таблиця 3.5).

Таблиця 3.5 – Опис атрибутів таблиці «Sex»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор статі
2.	Name	text	Назва статі

Таблиця «Body_type» потрібна для збереження записів про статуру тіла. Поле «ID» є унікальним ідентифікатором статури. Поле «Name» містить текстову назву статури тіла для відображення на сторінці (Таблиця 3.6).

Таблиця 3.6 – Опис атрибутів таблиці «Body_type»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор статури тіла
2.	Name	text	Назва статури тіла

Таблиця «Drinks» потрібна для збереження записів ставлення користувача до алкоголю. Поле «ID» є унікальним ідентифікатором запису про ставлення до

випивки. Поле «Name» містить текстове представлення ставлення до алкоголю (Таблиця 3.7).

Таблиця 3.7 – Опис атрибутів таблиці «Drinks»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор ставлення до алкоголю
2.	Name	text	Текстове представлення ставлення до алкоголю

Таблиця «Drugs» потрібна для збереження записів ставлення користувача до наркотиків. Усі поля та призначення полів схожі із таблицею «Drinks» і потрібні для відображення та збереження ставлення користувача до наркотиків (Таблиця 3.8).

Таблиця 3.8 – Опис атрибутів таблиці «Drugs»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор ставлення до наркотиків
2.	Name	text	Текстове представлення ставлення до наркотиків

Таблиця «Smokes» потрібна для представлення даних про ставлення до куріння. У цій таблиці є два атрибути «ID» та «Name». «ID» є первинним ключем, «Name» – текстове представлення ставлення до куріння (Таблиця 3.9).

Таблиця 3.9 – Опис атрибутів таблиці «Smokes»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор ставлення до куріння
2.	Name	text	Текстове представлення ставлення до куріння

Таблиця «Education» потрібна для збереження даних про навчальний заклад користувачів. Ця таблиця містить два поля: первинний ключ «ID» та текстова назва навчального закладу «Name» (Таблиця 3.10).

Таблиця 3.10 – Опис атрибутів таблиці «Education»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор навчального закладу
2.	Name	text	Назва навчального закладу

Таблиця «Ethnicity» містить в собі два атрибути для збереження даних про етнічні групи користувачів: ID, Name (Таблиця 3.11).

Таблиця 3.11 – Опис атрибутів таблиці «Ethnicity»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор етнічної групи
2.	Name	text	Назва етнічної групи

Таблиця «Offspring» потрібна для збереження даних про кількість дітей клієнта шлюбного агентства, або його ставлення до них. Дана таблиця містить такі атрибути: ID, Name (Таблиця 3.12).

Таблиця 3.12 – Опис атрибутів таблиці «Offspring»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор кількості дітей
2.	Name	text	Текстове представлення кількості дітей

Таблиця «Status» потрібна для збереження даних про сімейний статус клієнта шлюбного агентства. Дана таблиця містить такі атрибути: ID, Name (Таблиця 3.13).

Таблиця 3.13 – Опис атрибутів таблиці «Status»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор для сімейного статусу
2.	Name	text	Текстове представлення сімейного статусу пошуку

Таблиця «Orientation» потрібна для збереження даних про сексуальну орієнтацію клієнта шлюбного агентства. Дана таблиця містить такі атрибути: ID, Name (Таблиця 3.14).

Таблиця 3.14 – Опис атрибутів таблиці «Orientation»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор для сексуальної орієнтації
2.	Name	text	Назва сексуальної орієнтації

Таблиця «Speaks» потрібна для збереження даних про мову якою спілкується клієнт. Дана таблиця містить такі атрибути: ID, Name (Таблиця 3.15).

Таблиця 3.15 – Опис атрибутів таблиці «Speaks»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор для мови спілкування
2.	Name	text	Назва мови спілкування

Таблиця «Religion» потрібна для збереження даних про релігійне сповідання клієнта. Дана таблиця містить такі атрибути: ID, Name (Таблиця 3.16).

Таблиця 3.16 – Опис атрибутів таблиці «Religion»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор для релігійного сповідання
2.	Name	text	Назва релігійного сповідання

Таблиця «Diet» потрібна для збереження даних про дієту клієнта. Дана таблиця містить такі атрибути: ID, Name (Таблиця 3.17). Атрибут «ID» - це унікальний ідентифікатор для кожної із збережених дієт, який дозволяє створювати зв'язки із іншими таблицями у БД. Поле «Name» зберігає назви відомих дієт, перелік яких відображається користувачеві при заповненні клієнтської анкети.

Таблиця 3.17 – Опис атрибутів таблиці «Diet»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор дієти
2.	Name	text	Назва дієти

Таблиця «Sign» потрібна для збереження даних про знак зодіаку клієнта. Дана таблиця містить такі атрибути: ID, Name (Таблиця 3.18).

Таблиця 3.18 – Опис атрибутів таблиці «Sign»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор знаку зодіаку
2.	Name	text	Назва знаку зодіаку

Таблиця «Hobbies» потрібна для збереження даних про хобі та захоплення клієнта. Дана таблиця містить такі атрибути: ID, Name (Таблиця 3.19).

Таблиця 3.19 – Опис атрибутів таблиці «Hobbies»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор хобі
2.	Name	text	Назва хобі

Таблиця «Jobs» потрібна для збереження даних про місце роботи та посаду клієнта. Дана таблиця містить такі атрибути: ID, Name (Таблиця 3.20).

Таблиця 3.20 – Опис атрибутів таблиці «Jobs»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор місця роботи/посади
2.	Name	text	Назва місця роботи/посади

Таблиця «Pets» потрібна для збереження даних про наявність домашніх улюбленців у клієнта шлюбного агентства, або його ставлення до них. Дана таблиця містить такі атрибути: ID, Name (Таблиця 3.21).

Таблиця 3.21 – Опис атрибутів таблиці «Pets»

№	Назва	Тип даних	Опис
1.	ID	int	Унікальний ідентифікатор ставлення до домашніх улюбленців
2.	Name	text	Текстове представлення наявності або ставлення до домашніх улюбленців

Крім цього, буде створено дві розвідні таблиці для збереження переліку інтересів клієнта та переліку інтересів, якими повинен захоплюватись бажаний партнер. Створення розвідних таблиць дає змогу обирати більше ніж одне значення, що особливо корисне при виборі хобі. Отже, було створено розвідні таблиці «Hobbies_profile» та «Hobbies_ideal_profile». Структура таблиці «Hobbies_profile», яка включає перелік таких атрибутів: ID, FK_profile, FK_hobby (Таблиця 3.22).

Таблиця 3.22 – Опис атрибутів таблиці «Hobbies_profile»

№	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ
1.	FK_profile	int	Клієнтська анкета або перелік вимог до якого відноситься список хобі
2.	FK_hobby	int	Обране хобі

Структура таблиці «Hobbies_ideal_profile» аналогічна попередній та включає в себе такі атрибути: ID, FK_profile, FK_hobby (Таблиця 3.23).

Таблиця 3.23 – Опис атрибутів таблиці «Hobbies_ideal_profile»

№	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ
2.	FK_IdealProfile	int	Вторинний ключ для анкети бажаного партнера.
3.	FK_hobby	int	Вторинний ключ для хобі

Таким чином, було спроектовано даталогічну модель бази даних інформаційної системи для визначення сумісності клієнтів шлюбних агентств, яка складається із двадцяти трьох таблиць. Найголовнішими таблицями є таблиця користувачів веб-сайту шлюбного агентства, яка зберігає інформацію про дані для

автентифікації на веб-сайті, та таблиці із клієнтськими анкетами та переліком вимог до бажаного партнера. Інші таблиці є додатковими та зберігають допоміжну інформацію про категоріальні дані із клієнтських анкет.

3.4 Вибір засобів розробки для реалізації веб-сайту шлюбного агентства

Для реалізації веб-сайту шлюбного агентства було проведено вибір засобів розробки, що включає в себе вибір мови програмування, вибір середовища розробки, вибір системи керування базою даних та вибір додаткових компонентів та бібліотек, необхідних для програмної реалізації методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Обрано таке поєднання засобів розробки: мову програмування Python, середовище розробки PyCharm, фреймворк Flask, ORM бібліотеку SQLAlchemy, бібліотеку Faker, API OpenAI, бібліотеку PyTorch, бібліотеку Transformers, токенизатор BertTokenizer та нейромережеву модель BertModel.

Прикладну реалізацію методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних було створено з використанням високорівневої мови програмування Python. Основною перевагою Python є велика кількість бібліотек та інструментів, та можливість інтеграції з безліччю сторонніх платформ та додаткових API [37]. Також, за рейтингом мов програмування на 2024 рік, Python займає третє місце [38]. Популярність обраної мови сприяє розвитку активної спільноти, що є великою перевагою під час пошуку прикладів програмних рішень та допомоги.

PyCharm – це інтегроване середовище розробки Integrated Drive Electronics (IDE) для мови програмування Python. IDE PyCharm розроблене компанією JetBrains [39]. Обране середовище розробки надає розробникам великий набір інструментів для роботи з програмним кодом. В PyCharm можна як створювати програмні рішення так і тестувати та налагоджувати їх. З корисних функцій PyCharm можна виділити автоматичне завершення коду та виявлення помилок в реальному часі, що дозволяє одразу виправляти помилки. Крім цього, середовище дозволяє швидко та

легко встановити доступні бібліотеки та допоміжні інструменти для розробки. Описані переваги роблять IDE PyCharm одним із найпопулярніших середовищ розробки застосунків мовою програмування Python.

Для зручної та ефективної розробки веб-сайту шлюбного агентства обрано фреймворк Flask. Flask – це мікрофреймворк для розробки веб-застосунків мовою програмування Python. Flask називається мікрофреймворком через те, що на відмінну від великих фреймворків, таких як Django, Flask надає лише базові функції для створення веб-застосунків [40]. Це надає розробнику свободу вибору додаткових бібліотек, які можуть відрізнятись залежно від поставлених задач. В основному, Flask використовується для розробки серверної частини веб-сайту, управління маршрутизацією (наприклад, обробка запитів від клієнта) та відображення веб-сторінок.

Для роботи з базою даних обрано ORM бібліотеку SQLAlchemy. Оскільки, дана бібліотека створена за методологією Object-Relational Mapping, то її використання дозволяє взаємодіяти із БД, використовуючи об'єкти програмування замість написання SQL-запитів [41]. Також, перевагою даної бібліотеки є зручний користувацький інтерфейс через який можна взаємодіяти із БД. Проте, суттєвим недоліком є те, що дана бібліотека не надає змоги візуалізувати створену базу даних.

Для доповнення клієнтських анкет фіктивними даними використовується спеціальна бібліотека – Faker [42]. Дана бібліотека дозволяє генерувати правдоподібні дані такі як імена, адреси та контактні дані. Перевагою Faker є підтримка генерації даних різними мовами, зокрема й українською.

Для інтеграції моделі GPT, яка необхідна для автоматизованої генерації текстових описів клієнтів, використовується API OpenAI [43]. Взаємодія з платформою OpenAI відбувається за допомогою HTTP запитів, через які передаються промпти, відповідно до яких генеруються текстові описи клієнтів.

Оскільки, запропонований підхід передбачає нейромережеве порівняння текстових описів клієнтів, то для зручної реалізації зазначеного функціоналу використано бібліотеку PyTorch [44], яка забезпечує гнучкий інтерфейс для

створення нейронних мереж. Дана бібліотека підходить як для проведення навчання моделей так і для використання вже готових натренованих нейронних мереж.

Крім цього, оскільки нейромережева модель BERT відноситься до моделей трансформерів, то для її використання у проєкті необхідно використовувати бібліотеку Transformers [45] від платформи Hugging Face [46]. Саме через цю бібліотеку відбувається використання попередньо навчених моделей.

Окремим компонентом бібліотеки Transformers є BertTokenizer [47] – токенізатор для моделі BERT. Токенізатор розбиває вхідний текст на токени, підходящі для обробки моделлю BERT. Також, при розбитті тексту через BertTokenizer додаються спеціальні токени для позначення початку і кінця речень.

Для генерації векторних представлень текстових описів клієнтів використовується модель BERT, доступ до якої надається через компонент BertModel [48]. До особливостей BertModel варто віднести врахування контексту слів при формування векторного представлення тексту.

Отже, для розробки веб-сайту шлюбного агентства було обрано поєднання сучасних засобів розробки, які забезпечують зручність та високу ефективність розробки. За мову програмування було обрано Python – одну з найпопулярніших мов програмування на сьогоднішній день. Також, було обрано інтегроване середовище розробки PyCharm, що забезпечує ефективну розробку ПЗ мовою програмування Python. Основним інструментом для розробки серверної частини веб-сайту обрано мікрофреймворк Flask. Для зберігання даних та їх обробки обрано ORM бібліотеку SQLAlchemy, яка спрощує взаємодію з базою даних завдяки об'єктно-реляційному підходу. Генерацію фіктивних даних для анкет клієнтів забезпечено за допомогою бібліотеки Faker, яка підтримує генерацію даних українською мовою. Для автоматизованої генерації текстових описів клієнтів інтегровано API OpenAI, що дозволяє генерувати текст на основі заданих параметрів через взаємодію з моделлю GPT за допомогою промптів. Для забезпечення можливості використання нейромережевої моделі BERT використано такі бібліотеки як PyTorch та Transformers, а також додаткові компоненти – BertTokenizer та BertModel.

Висновки до розділу 3

У розділі було розроблено структуру інформаційної системи визначення сумісності клієнтів шлюбних агентств, яка складається із трьох підсистем та бази даних. У першій підсистемі реалізовано функціонал, який забезпечує наповнення клієнтської бази анкетами кандидатів. Друга підсистема є прикладним застосуванням запропонованого методу у вигляді веб-сайту шлюбного агентства, тому тут реалізовано функціонал пов'язаний із користуванням програмного застосунку. Третя підсистема відповідає за роботу методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

Для збереження клієнтських анкет та допоміжної інформації для функціонування веб-сайту було розроблено даталогічну модель бази даних, яка складається із двадцяти трьох взаємопов'язаних таблиць. Основними таблицями є таблиця з особистими даними про клієнта та таблиця з переліком вимог до бажаного партнера. Саме на основі даних із цих таблиць відбувається підбір сумісних партнерів використовуючи запропонований метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

Також, було обрано набір засобів розробки для реалізації веб-сайту шлюбного агентства. Серверну частину реалізовано за допомогою мікрофреймворку Flask в поєднанні з ORM бібліотекою SQLAlchemy для взаємодії з базою. Генерацію фіктивних даних для клієнтських анкет виконано через бібліотеку Faker. Автоматизовану генерацію текстових описів клієнтів реалізовано за допомогою моделі GPT, інтегрованої в проєкт через API OpenAI. Для роботи з нейромережевою моделлю BERT застосовано бібліотеки PyTorch та Transformers, а також компоненти BertTokenizer і BertModel.

Отже, розроблена інформаційна система для визначення сумісності клієнтів шлюбних агентств поєднує три підсистеми та базу даних, забезпечуючи повний цикл роботи шлюбного агентства: від заповнення клієнтських анкет до автоматизованого підбору сумісних партнерів.

РОЗДІЛ 4 Дослідження ефективності методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних

4.1 Програмна архітектура інформаційної системи визначення сумісності клієнтів шлюбних агентств

Для демонстрації роботи методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних було реалізовано інформаційну систему у вигляді веб-сайту шлюбного агентства. Програмна реалізація веб-застосунку складається з 26 класів, переважна більшість з яких створенні для представлення сутностей бази даних (Рисунок 4.1).

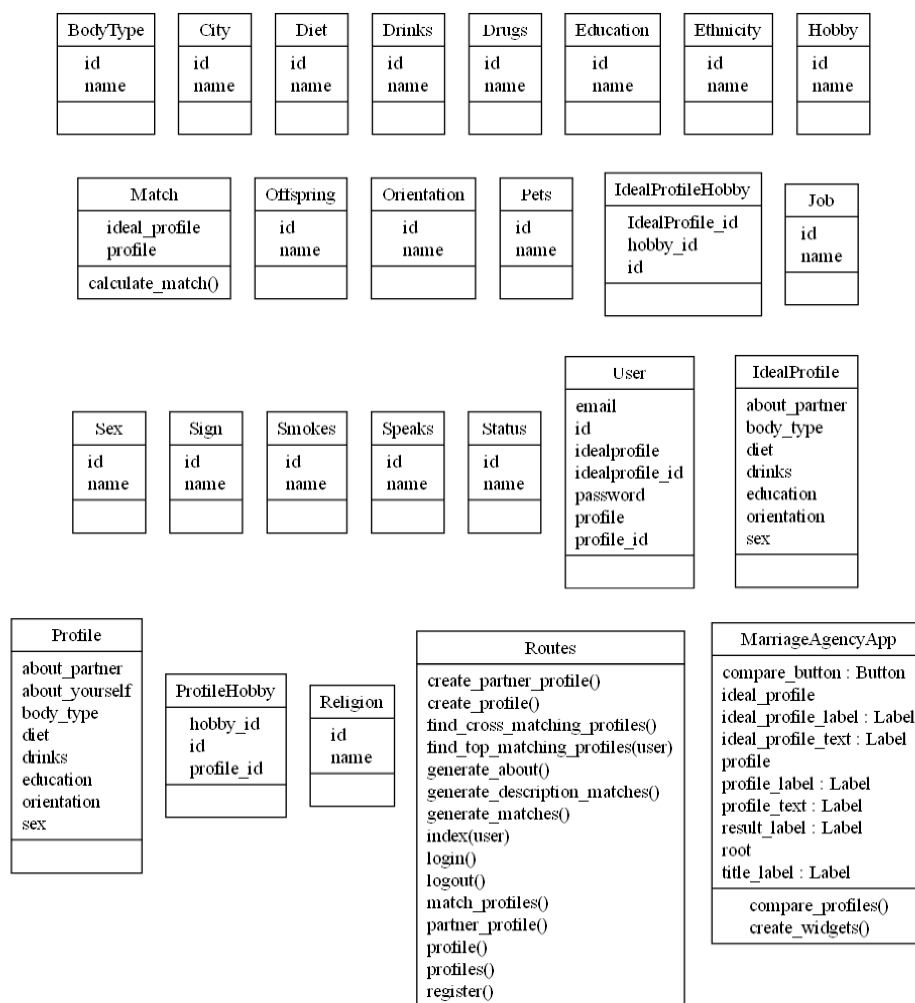


Рисунок 4.1 – Діаграма класів інформаційної системи визначення сумісності клієнтів шлюбних агентств

Для представлення сутностей із бази даних створено 23 класи: BodyType, City, Diet, Drinks, Drugs, Education, Ethnicity, Hobby, Offspring, Orientation, Pets, Job, Sex, Sign, Smokes, Speaks, Status, Religion, ProfileHobby, IdealProfileHobby, User, IdealProfile, Profile. Усі вони служать для отримання інформації для формування клієнтських анкет. Нижче наведено опис кожного класу:

- BodyType – клас для отримання інформації про тип будови тіла клієнта;
- City – клас для отримання інформації про місце проживання клієнта;
- Diet – клас для отримання інформації про те, чи дотримується клієнт дієти та якої;
- Drinks – клас для отримання інформації про те, чи вживає клієнт алкогольні напої і з якою періодичністю;
- Drugs – клас для отримання інформації про те, чи вживає клієнт наркотичні речовини і з якою періодичністю;
- Education – клас для отримання інформації про освіту клієнта;
- Ethnicity – клас для отримання інформації про етнічну приналежність клієнта;
- Hobby – клас для збереження інформації про доступні до вибору хобі;
- Offspring – клас для отримання інформації про наявність дітей та ставлення до них;
- Orientation – клас для отримання інформації про орієнтацію клієнта;
- Pets – клас для отримання інформації про наявність домашніх улюбленців та ставлення до них;
- Job – клас для отримання інформації про професію клієнта;
- Sex – клас для отримання інформації про стать клієнта;
- Sign – клас для отримання інформації про знак зодіаку клієнта;
- Smokes – клас для отримання інформації про те, чи клієнт курить з якою періодичністю;
- Speaks – клас для отримання інформації про мову спілкування клієнта;

- Status – клас для отримання інформації про сімейний статус клієнта;
- Religion – клас для отримання інформації про релігію клієнта;
- ProfileHobby – клас для отримання інформації про хобі клієнта;
- IdealProfileHobby – клас для отримання інформації про хобі бажаного партнера;
- User – клас для отримання інформації про зареєстрованих на веб-сайті користувачів;
- IdealProfile – клас для отримання інформації про перелік вимог до бажаного партнера;
- Profile – клас для отримання інформації про клієнта.

Для управління веб-сайтом шлюбного агентства створено клас «MarriageAgencyApp». У даному класі реалізовано методи, які відповідають за візуалізацію інтерфейсу користувача та забезпечують взаємодію користувача із веб-сайтом.

Для обчислення відсотків сумісності створено клас «Match». Даний клас отримує інформацію про клієнта та перелік вимог до бажаного партнера та інформацію про профілі кандидатів. Методи з даного класу використовується як для порівняння профілів за фреймовою моделлю подання, так і для порівняння текстових описів клієнтів.

Також, реалізовано клас «Routes» у якому зосереджено методи для реагування на дії користувача. Тобто, тут містяться методи, які виконують свою роботу після певних користувацьких дій, наприклад, якщо користувач натискає на певну кнопку то відбувається конкретна дія, передбачена певним методом.

Таким чином, було реалізовано прикладне застосування методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних у вигляді веб-сайту шлюбного агентства. Програмна структура інформаційної системи складається з 26 взаємопов'язаних класів, які взаємодіють між собою та забезпечують доступ до методу через користувацький інтерфейс.

4.2 Розробка прикладних компонентів інформаційної системи визначення сумісності клієнтів шлюбних агентств

Для прикладного застосування розробленого методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних було створено інформаційну систему у вигляді веб-сайту шлюбного агентства.

При відкритті веб-сайту шлюбного агентства відкривається головна сторінка, на якій користувачеві доступна можливість реєстрації або авторизації (у випадку, якщо користувач зареєстрований на веб-сайті шлюбного агентства). Реєстрація та авторизація є початковими кроками роботи клієнта із платформою, адже це дає змогу створити свій особистий профіль та відкриває можливості для подальших взаємодій із системою. Процес реєстрації користувача складається з таких послідовних кроків: відкриття веб-сайту, натиснення на кнопку «Вхід», заповнення полів необхідними даними, натиснення на кнопку «Зареєструватися».

Після реєстрації користувачеві відкривається його особистий профіль з можливістю заповнення клієнтської анкети. Також, після реєстрації, в базу даних додається запис про нового користувача веб-сайту. Збереження даних, які використовувались при реєстрації, забезпечити можливість повторного входу в особистий профіль, використовуючи персональний логін та пароль. Блок-схему алгоритму реєстрації користувача на веб-сайті шлюбного агентства наведено на рисунку 4.2.

В процесі реєстрації, відбувається перевірка на те, чи всі поля заповненні, адже реєстрація неможлива при відсутності необхідної інформації про клієнта. Якщо користувач спробує зареєструватись не ввівши необхідні дані, його буде повернуто на сторінку з формою для реєстрації та проінформовано про помилку, яка виникла. Також, якщо під час реєстрації виникатимуть інші непередбачувані помилки, користувача буде проінформовано про них.

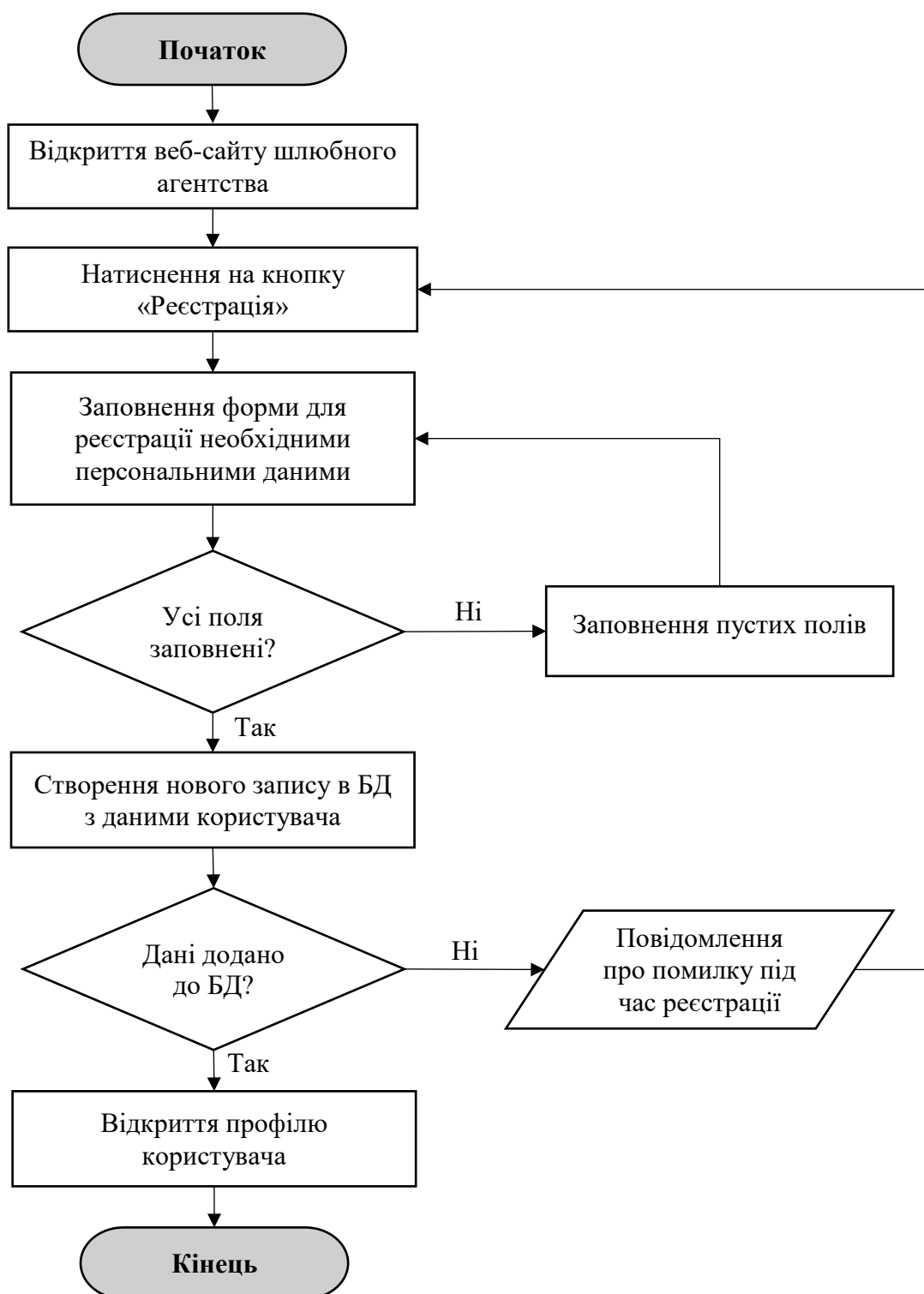


Рисунок 4.2 – Блок-схема алгоритму реєстрації нового користувача на веб-сайті шлюбного агентства

Після того, як користувач зареєструвався або авторизувався на веб-сайті шлюбного агентства, для нього стає доступною можливість заповнення клієнтської

анкети. Клієнтська анкета включає в себе дві частини: перша частина стосується безпосередньо клієнта, друга – перелік вимог до бажаного партнера.

Для цього, користувачеві необхідно заповнити дві форми, після чого стане доступною можливість автоматизованої генерації текстових описів, як для самого клієнта так і для бажаного партнера. Алгоритм заповнення клієнтської анкети включає в себе кроки, пов'язані із перевіркою коректності заповнення полів. Коли клієнт правильно заповнює анкету, його дані додаються до БД, що дозволяє проводити пошук потенційних партнерів. Блок-схему роботи алгоритму по заповненню клієнтських анкет наведено на рисунку 4.3.

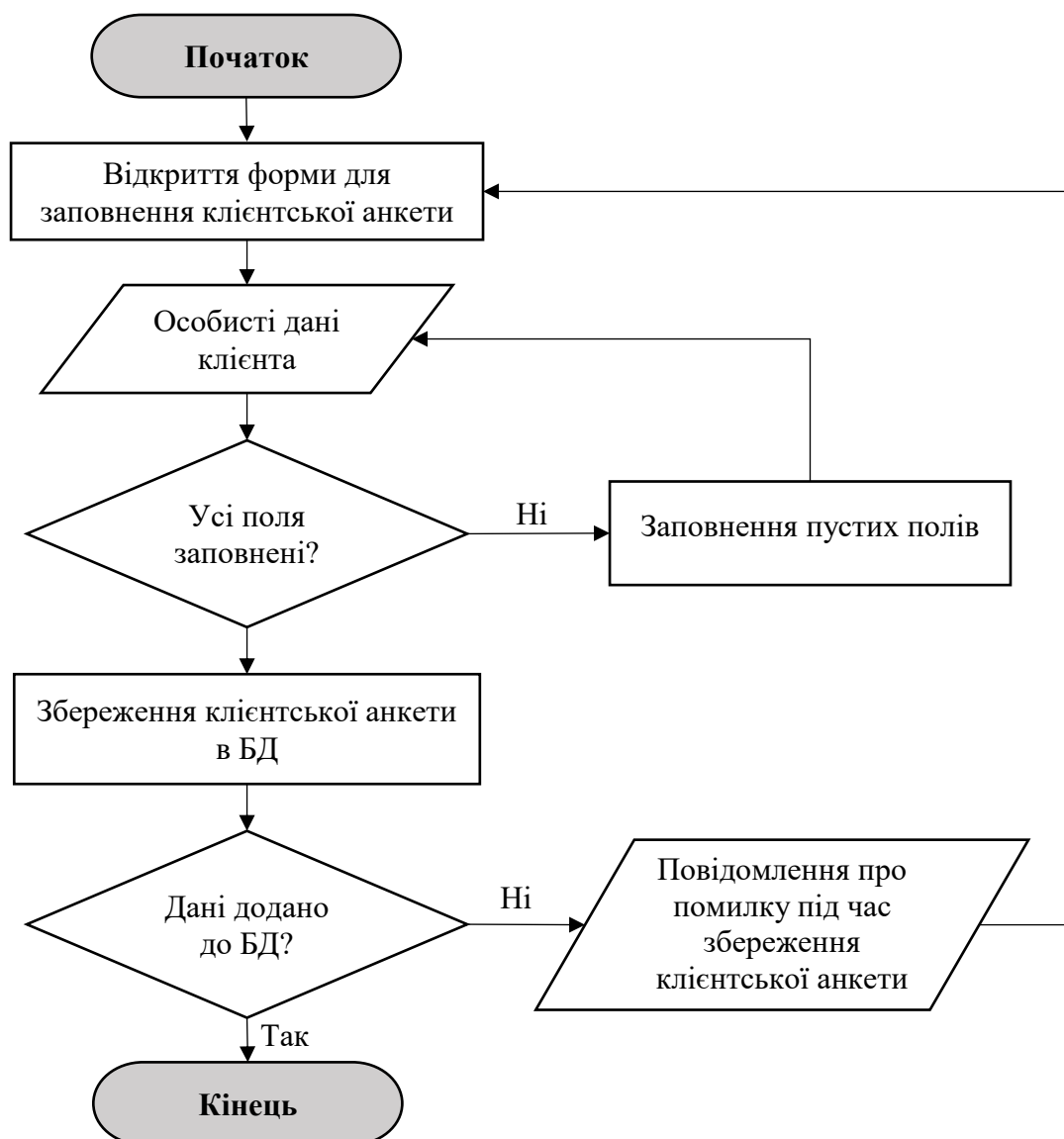


Рисунок 4.3 – Блок-схема алгоритму заповнення клієнтської анкети

На рисунку 4.3 продемонстровано заповнення тієї частини клієнтської анкети, яка стосується самого клієнта. Проте, алгоритм для заповнення переліку вимог до бажаного партнера аналогічний процесу заповнення персональної інформації.

Також, для кожної анкети відбувається автоматизована генерація текстових описів. Проте, інколи під час цього процесу можуть виникати помилки, що призводить до подальшого неправильного порівняння текстових описів. Однією із таких помилок є відсутність згенерованого текстового опису. Через це, такі анкети в полі для текстового опису можуть мати значення «NULL», що унеможлиблює їхнє порівняння (Рисунок 4.4).

id ^	height	weight	age	income	aboutYourself	sex_id	city_id	job_id	status_id	orienta...	body_t...
1	168	72	26	140146	Hi! I'm a 26-year-old ...	2	5	15	1	2	5
2	184	111	27	21846	Hello! I'm a 27-year-o...	1	7	18	4	4	6
3	168	80	43	50249	I'm a slender and acti...	2	5	15	1	2	5
4	174	106	39	45443	null	1	17	20	3	2	5
5	166	64	36	135673	I'm a 36-year-old wo...	2	20	19	3	5	6
6	161	119	56	50469	I'm a 56-year-old buil...	2	12	19	2	4	6
7	182	78	36	122389	I'm a 36-year-old cha...	1	6	1	2	3	5
8	160	107	23	117249	Hi! I'm a 23-year-old ...	2	13	4	5	3	3
9	165	61	29	61005	I'm a 29-year-old wo...	2	3	5	3	3	6
10	169	101	48	121410	Hello! I'm a charming ...	1	10	10	5	3	1

Рисунок 4.4 – Відсутність текстового опису в анкеті клієнта

Тому, прийнято рішення проводити додаткову перевірку анкет та видаляти ті анкети у яких відсутній текстовий опис. Блок-схему даного алгоритму наведено на рисунку 4.5. Описаний підхід допоможе уникнути помилок при подальшій роботі методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

Після того, як клієнт заповнить особисту анкету та складе перелік вимог до бажаного партнера йому стане доступна можливість пошуку потенційних партнерів. Для цього застосовується розроблений метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Відповідно до запропонованого підходу відбувається паралельне використання двох методів інтелектуального аналізу анкетних даних. Перший метод має на меті перехресний

обрахунок відсотків відповідності зазначеним вимогам, другий – перехресний обрахунок оцінки схожості текстових описів клієнтів.

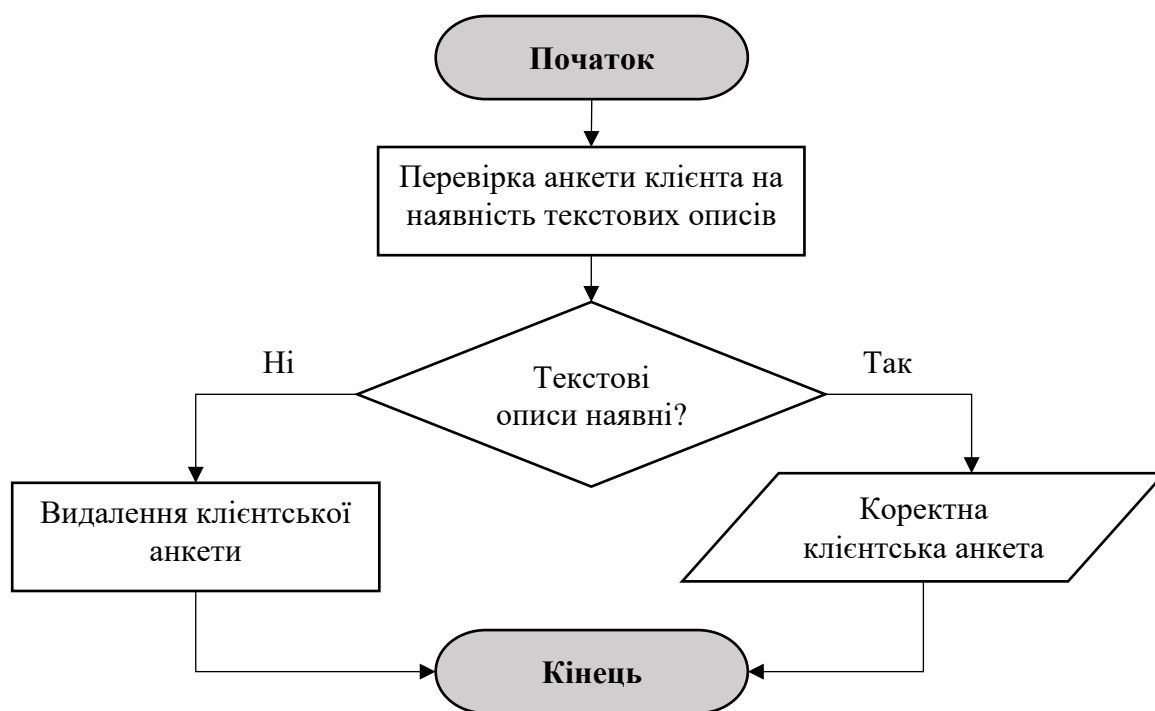


Рисунок 4.5 – Блок-схема алгоритму перевірки наявності текстових описів у анкетах клієнтів шлюбного агентства

Спочатку, у кожному з методів відбувається фільтрація клієнтської бази за обов'язковими параметрами. Після того як сформовано вибірку потенційних партнерів відбувається перехресне порівняння текстових описів. Тобто, в кожній потенційній парі порівнюються текстові описи бажаних партнерів із текстовими розповідями клієнтів про себе. Таким чином обчислюється дві оцінки подібності текстових описів, по одній для кожного партнера з потенційної пари. Після чого, за формулою середнього арифметичного обраховується загальний відсоток сумісності. Блок-схему роботи алгоритму перехресного порівняння текстових описів потенційної пари наведено на рисунку 4.6.

Аналогічним чином відбувається перехресний обрахунок відсотку відповідності зазначеним вимогам. В результаті роботи обох методів формується дві множини сумісних анкет. Далі, клієнту, який проводив пошук партнерів, відображається по десять потенційних партнерів з найвищими відсотками

сумісності, обраховані за двома методами. Крім цього, для клієнта формується пропозиція з об'єднаної множини сумісних партнерів, тобто партнерів, які потрапили у топ десять за результатами роботи двох методів.

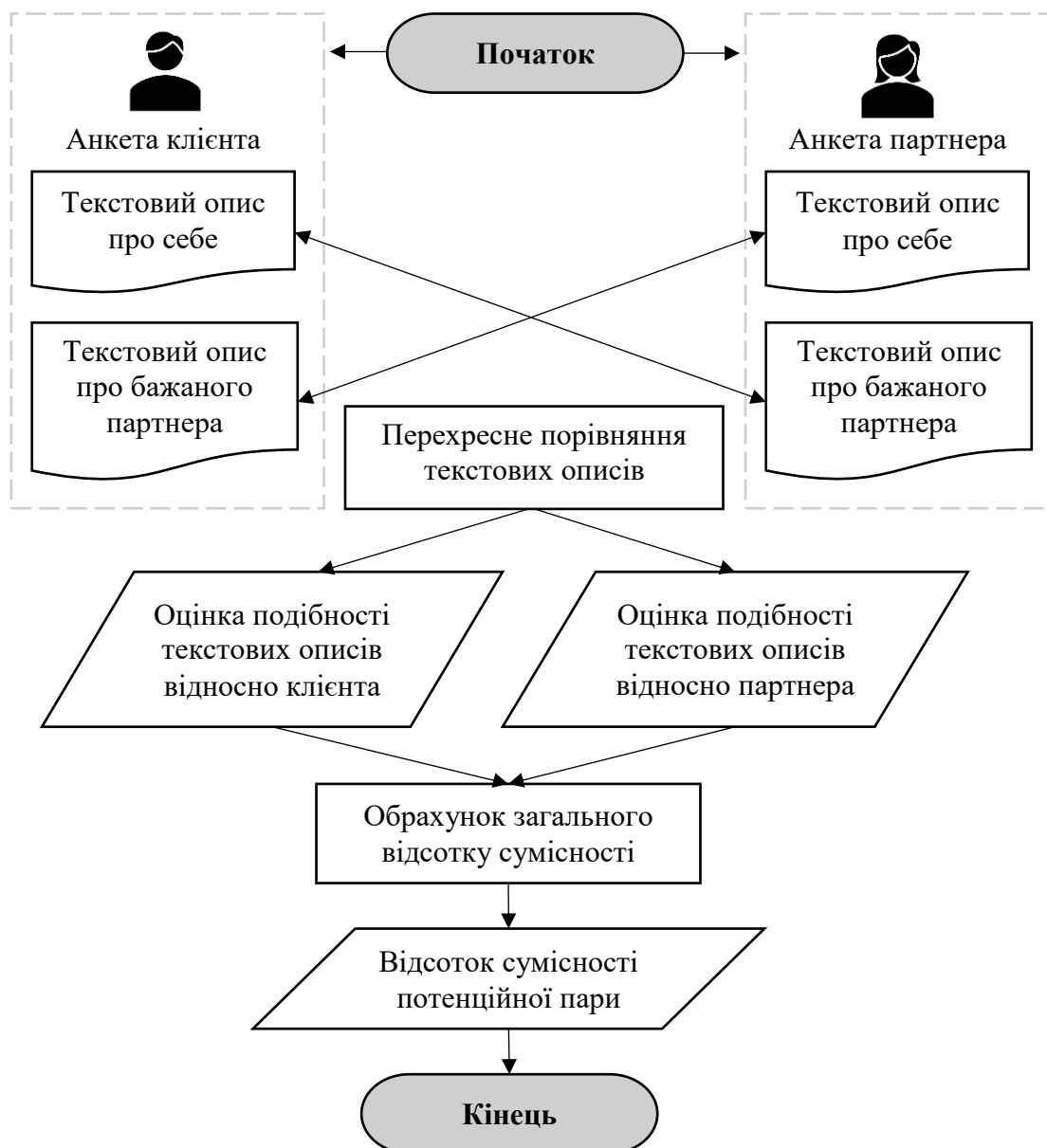


Рисунок 4.6 – Блок-схема алгоритму перехресного порівняння текстових описів потенційної пари

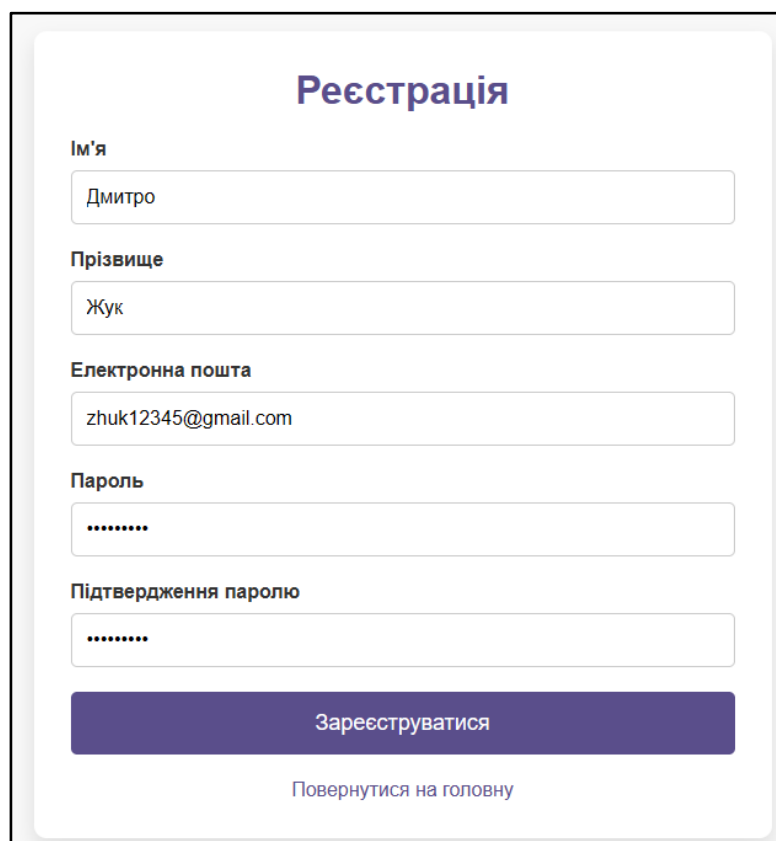
Отже, було розроблено веб-сайт шлюбного агентства, який використовує метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Розробка веб-сайту включала реалізацію алгоритму

автентифікації користувача в системі, процесу заповнення клієнтської анкети та обрахунок відсотків сумісності для потенційних пар.

4.3 Прикладне тестування інформаційної системи визначення сумісності клієнтів шлюбних агентств

Для проведення прикладного застосування інформаційної системи визначення сумісності клієнтів шлюбних агентств створено три тестові випадки: перевірка процесу реєстрація, перевірка процесу заповнення клієнтської анкети, перевірка процесу пошуку сумісних партнерів.

Оскільки, без реєстрації в системі клієнт не матиме змоги використовувати метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, то перший тест-кейс присвячений перевірці можливості реєстрації клієнта на веб-сайті шлюбного агентства. Для реєстрації клієнту необхідно заповнити відповідну форму персональними даними (Рисунок 4.7).



Реєстрація

Ім'я
Дмитро

Прізвище
Жук

Електронна пошта
zhuk12345@gmail.com

Пароль
.....

Підтвердження паролю
.....

[Зареєструватися](#)

[Повернутися на головну](#)

Рисунок 4.7 – Процес заповнення форми реєстрації персональними даними клієнта

Після того, як клієнт заповнить усі поля та натисне на кнопку «Зареєструватись» відбудеться процес перевірки заповнених полів і у випадку успіху користувача буде зареєстровано в системі (Таблиця 4.1).

Таблиця 4.1 – Тест-кейс АТ0001

Тест-кейс ID: АТ0001	Пріоритет: 1	Створено: 28.11.2024, Д. Жук
Назва: Перевірка процесу реєстрації користувача на веб-сайті шлюбного агентства		
Вхідні дані: Прізвище = «Жук», Ім'я = «Дмитро», Email = «zhuk12345@gmail.com», Пароль = «zhuk12345»		
Кроки	Очікуваний результат	
<p><i>Передумова:</i> користувач повинен бути не зареєстрованим у системі</p> <ol style="list-style-type: none"> Відкрити веб-сайт шлюбного агентства Натиснути кнопку «Реєстрація» Заповнити форму реєстрації персональними даними Натиснути кнопку «Зареєструватись» Порівняти фактичний результат з очікуваним 	<p>Сторінка з формою для реєстрації закрилась, та відкрилась персональна сторінка клієнта шлюбного агентства.</p> <p>В БД у таблиці «Profiles» та таблиці з'явився новий запис з персональною інформацією про щойно зареєстрованого користувача.</p>	
Результат виконання тест-кейсу: пройдено успішно		

Для можливості повторного входу на персональну сторінку клієнта шлюбного агентства, усі дані, які використовувались при реєстрації будуть збережені у базу даних. Таким чином, після успішної реєстрації нового клієнта

шлюбного агентства у таблиці «Users» додається новий запис із його персональними даними (Рисунок 4.8).

id	first_na...	last_na...	fk_profile	email	pass_	fk_ideal...
172	Malik	Mcgee	172	richarda...	8hI^l*Rs\$_	172
173	Patricia	Miller	173	cherylke...	_*\$7Bu...	173
174	Andre	Archer	174	beasleys...	BJ5C@Q...	174
175	Vanessa	Figueroa	175	jill49@ex...	#92Y8f+...	175
176	Sarah	Edwards	176	walkerm...	N+5HZ0...	176
177	Michael	Smith	177	collinsja...	ot6bfEg...	177
178	Richard	Fox	178	currysco...	4xABOJt...	178
179	Ashley	Carter	179	wilsonga...	p4pVh%...	179
180	Володя	Жупансь...	null	51423di...	qwerty123	null
181	Дмитро	Жук	null	zhuk123...	zhuk12345	null

Рисунок 4.8 – Новий запис в БД з переліком персональних даних нового клієнта

Після реєстрації на веб-сайті шлюбного агентства, користувач повинен заповнити анкету клієнта. Анкета складається з двох блоків: інформація про клієнта та вимоги до бажаного партнера. Для того, щоб здійснити пошук сумісних партнерів потрібно заповнити усі поля анкети особистою інформацією (Рисунок 4.9). До особистої інформації відноситься: мова спілкування, стать, місто проживання, зріст, дохід, професія, сімейний статус, орієнтація, будова тіла, дотримання дієти, вживання алкоголю чи наркотиків, куріння, освіта, етнічна приналежність, наявність дітей, наявність домашніх улюбленців, релігія та знак зодіаку.

Для заповнення клієнтської анкети, користувачеві веб-сайту необхідно заповнити усі поля форми, інакше, анкету не буде збережено. Це необхідно для того, щоб користувачі не мали змоги залишати поля з анкети пустими. Інакше, при аналізі анкетних даних, з метою пошуку сумісних партнерів, анкети з пустими полями будуть перешкоджати роботі методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

Створення нового профілю

Мови	Англійська ▾
Стать	чоловік ▾
Місто	Lake Dennisbury
Зріст (см)	182
Доход	35090
Робота	Sales manager
Статус	У стосунках ▾
Орієнтація	Пансексуальний ▾
Тип тіла	Атлетичний ▾
Дієта	Веган ▾
Алкоголь	Ні ▾
Наркотики	Так ▾
Куріння	Іноді ▾
Освіта	Середня школа ▾
Етнічна приналежність	Європеєць ▾
Діти	Одна дитина ▾
Тварини	Собака ▾
Релігія	Християнство ▾
Знак зодіаку	Стрілець ▾

Створити профіль

Рисунок 4.9 – Процес заповнення анкети клієнта

Якщо усі поля анкети заповнені, користувачеві необхідно зберегти свою анкету. У випадку успіху, анкетні дані користувача будуть збережені у базі даних. Отже, для перевірки описаного процесу створено другий тест-кейс (Таблиця 4.2). Даний тест-кейс створено з метою дослідження коректності роботи процесу збереження клієнтської анкети у базі даних. Адже, наявність заповненої анкети це ключова вимога для можливості проведення подальшого пошуку сумісних партнерів.

Таблиця 4.2 – Тест-кейс АТ0002

Тест-кейс ID: АТ0002	Пріоритет: 1	Створено: 28.11.2024, Д. Жук
Назва: Перевірка процесу заповнення анкети клієнта шлюбного агентства		
Вхідні дані: Особисті дані клієнта		
Кроки	Очікуваний результат	
<p><i>Передумова:</i> користувач повинен бути авторизованим у системі</p> <ol style="list-style-type: none"> 1. Відкрити веб-сайт шлюбного агентства 2. Авторизуватись у системі 3. Відкрити персональну сторінку клієнта шлюбного агентства 4. Натиснути кнопку «Заповнити анкету» 5. Заповнити усі поля особистою інформацією 6. Натиснути на кнопку «Зберегти анкету» 7. Порівняти фактичний результат з очікуваним 	<p>Сторінка з формою для заповнення анкетних даних закрилась. Відкрилась персональна сторінка клієнта шлюбного агентства з переліком особистих даних, які були додані при заповненні анкети.</p> <p>В БД у таблицях «Profiles» та «Ideal_profiles» з'явилися нові записи. У таблицю «Profile» додано особисті дані про клієнта, а у таблицю «Ideal_profiles» – вимоги до бажаного партнера.</p>	
Результат виконання тест-кейсу: пройдено успішно		

Після успішного збереження анкети клієнта, користувач веб-сайту шлюбного агентства має змогу здійснити пошук потенційних партнерів. Для цього, необхідно натиснути на кнопку «Пошук партнерів» на персональній сторінці клієнта. Після цього, система здійснить інтелектуальний аналіз анкетних даних та сформує три множини сумісних партнерів. Для перевірки процесу пошуку сумісних партнерів проведено третю перевірку (Таблиця 4.3).

Для проходження даного тест-кейсу встановлено поріг в 60% сумісності. Це дозволить клієнту шлюбного агентства розглянути найбільш релевантних кандидатів. За оптимальну кількість кандидатів, що перевищують встановлений поріг сумісності, прийнято три.

Таблиця 4.3 – Тест-кейс АТ0003

Тест-кейс ID: АТ0003	Пріоритет: 2	Створено: 28.11.2024, Д. Жук
Назва: Перевірка процесу пошуку сумісних партнерів		
Вхідні дані: Анкета клієнта, база анкет кандидатів		
Кроки	Очікуваний результат	
<p><i>Передумова:</i> користувач повинен бути авторизованим у системі</p> <ol style="list-style-type: none"> 1. Відкрити веб-сайт шлюбного агентства 2. Авторизуватись у системі 3. Відкрити персональну сторінку клієнта шлюбного агентства 4. Натиснути кнопку «Пошук партнерів» 5. Порівняти фактичний результат з очікуваним 	<p>Очікуваний результат: В обох множинах є не менше трьох партнерів із відсотком сумісності більшим за 60%.</p> <p>В об'єднаній множині є не менше трьох сумісних партнерів.</p> <p>Фактичний результат: В обох множинах є не менше трьох партнерів із відсотком сумісності більшим за 60%.</p> <p>В об'єднаній множині є не менше трьох сумісних партнерів.</p>	
Результат виконання тест-кейсу: пройдено успішно		

Отже, результати роботи методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних відображаються для користувача у вигляді трьох переліків (Рисунок 4.10).

У першому переліку користувач має змогу переглянути тих партнерів, які визначені сумісними за результатами роботи обох методів. Нижче, у другому переліку, користувач може переглянути сумісних партнерів, які були визначені за результатами роботи методу на основі фреймової моделі подання анкетних даних. У третьому переліку – партнери з високою сумісністю за результатами нейромережевого порівняння текстових описів. Таким чином, користувачеві буде надана детальна інформація про сумісних партнерів, які знайдені за допомогою різних методів визначення сумісності.

Шлюбне Агентство Головна Послуги Вхід

Найкращі кандидати за двома методами

 Amy Mooney Місто: Yorkport Стать: чоловік Вік: 44 Переглянути профіль	 Alyssa Wise Місто: Yorkport Стать: чоловік Вік: 44 Переглянути профіль	 Robert Dixon Місто: Yorkport Стать: чоловік Вік: 25 Переглянути профіль
--	---	--

Співпадіння за методом порівняння фреймів

 Amy Mooney Місто: Yorkport Стать: чоловік Вік: 44 Відсоток співпадіння: 72.5 Переглянути профіль	 David Lane Місто: Yorkport Стать: чоловік Вік: 44 Відсоток співпадіння: 67.5 Переглянути профіль	 Erka Sanchez Місто: Yorkport Стать: чоловік Вік: 44 Відсоток співпадіння: 63.75 Переглянути профіль
 Alyssa Wise Місто: Yorkport Переглянути профіль	 Robert Dixon Місто: Yorkport Переглянути профіль	 Richard Fox Місто: Yorkport Переглянути профіль

Порівняння за текстовими описами

 Jessica Burke Місто: Lake Dennisbury Стать: чоловік Вік: 38 Відсоток співпадіння: 73.35292 Переглянути профіль	 Barry Strong Місто: West Kevinmouth Стать: чоловік Вік: 38 Відсоток співпадіння: 69.97182 Переглянути профіль	 Christopher Fields Місто: Yorkport Стать: чоловік Вік: 36 Відсоток співпадіння: 67.32328 Переглянути профіль
 Robert Dixon Місто: Yorkport Стать: чоловік Вік: 25 Відсоток співпадіння: 66.2224 Переглянути профіль	 Jessica Zhang Місто: Gornelmouth Стать: чоловік Вік: 39 Відсоток співпадіння: 62.433774 Переглянути профіль	 Edwin Gardner Місто: Lake Kevinview Стать: чоловік Вік: 44 Відсоток співпадіння: 62.337864 Переглянути профіль

Рисунок 4.10 – Відображення результатів пошуку сумісних партнерів

Таким чином, проведене прикладне тестування розробленого веб-сайту шлюбного агентства доводить, що створений функціонал працює вірно. Користувач має змогу зареєструватись в системі та заповнити анкету клієнта шлюбного агентства особистою інформацією та вимогами до бажаного партнера. Також, користувач має змогу здійснити пошук сумісних партнерів та переглянути отриманий результат у вигляді трьох переліків сумісних партнерів із отриманими відсотками сумісності.

4.4 Особливості використання веб-сайту шлюбного агентства

При відкритті веб-сайту шлюбного агентства відкривається головна сторінка на якій розміщена загальна інформація про шлюбне агентство, перелік послуг та переваги звернення (Рисунок 4.11).

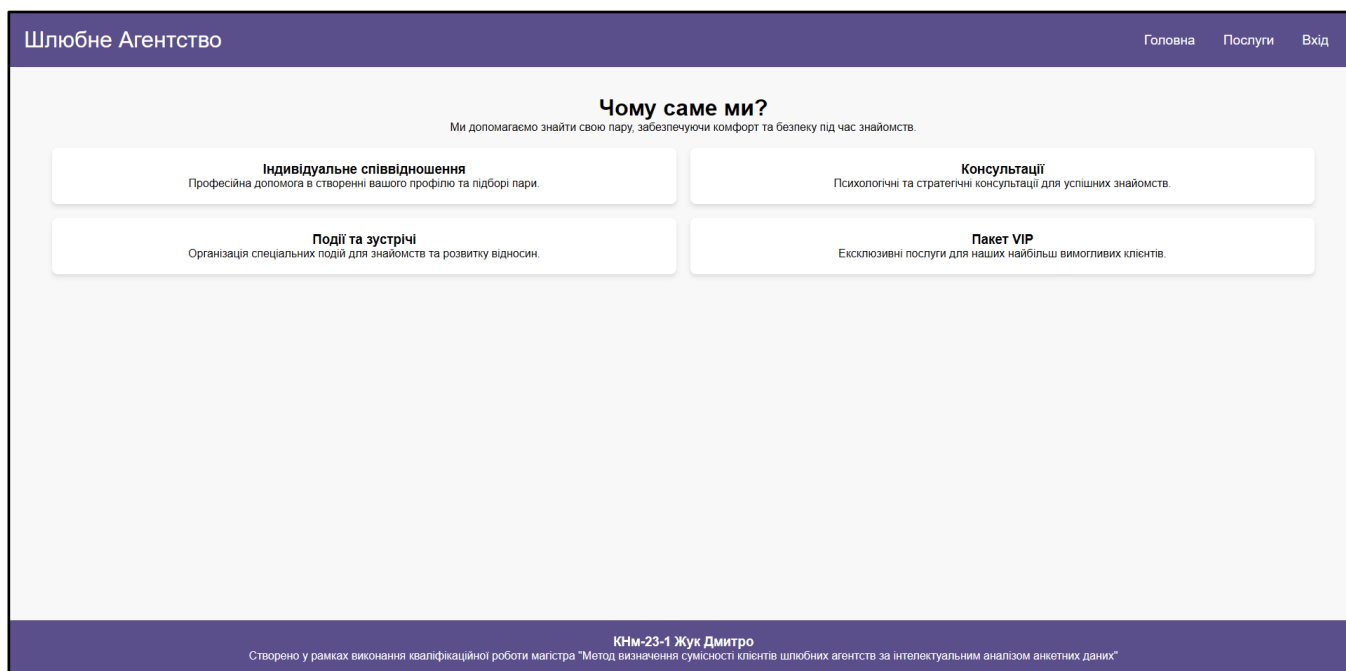


Рисунок 4.11 – Головна сторінка веб-сайту шлюбного агентства

Для використання веб-сайту шлюбного агентства, користувач повинен бути авторизованим у системі. Для цього потрібно зареєструватись та створити особистий профіль. Процес реєстрації включає заповнення відповідної форми персональними даними (Рисунок 4.12). Для вдалої реєстрації, користувачеві необхідно заповнити усі поля форми, інакше, при спробі зареєструватись буде відображено повідомлення про помилку та прохання заповнити пусті поля. Для реєстрації на веб-сайті шлюбного агентства, користувачеві необхідно ввести таку особисту інформацію: ім'я, прізвище, електронна пошта та пароль.

Реєстрація

Ім'я

Прізвище

Електронна пошта

Пароль

Підтвердження паролю

Зареєструватися

[Повернутися на головну](#)

Рисунок 4.12 – Вигляд форми для реєстрації на веб-сайті шлюбного агентства

В подальшому, зареєстровані користувачі мають змогу повторно увійти в особистий профіль, використовуючи логін та пароль, які були обрані при реєстрації. Логін та пароль необхідно вводити у форму для авторизації (Рисунок 4.13).

Вхід до акаунту

Email

Пароль

Увійти

[Немає акаунту?](#)
[Зареєструйтесь](#)

Рисунок 4.13 – Вигляд форми для авторизації

Далі, користувачеві необхідно заповнити анкету клієнта. Дана анкета складається з двох частин. Перша частина стосується самого клієнта, де йому потрібно вказати особисту інформацію про себе, свої вподобання та захоплення. Друга частина включає перелік запитань стосовно бажаного партнера. Тобто, тут клієнту необхідно описати, які якості повинні бути в шуканій людині. Отже, для заповнення анкети клієнта, необхідно ввести особисті дані в відповідну форму (Рисунок 4.14).

The image shows a web form for creating a new profile. The title is 'Створення нового профілю'. The form contains 18 fields, each with a label on the left and a corresponding input field on the right. The input fields are either dropdown menus or text boxes. At the bottom of the form is a purple button labeled 'Створити профіль'.

Field Label	Value
Мови	Англійська
Стать	чоловік
Місто	Введіть місто
Зріст (см)	Введіть зріст
Дохід	Введіть дохід
Робота	Введіть вашу роботу
Статус	Самотній
Орієнтація	Гетеросексуальний
Тип тіла	Стрункий
Дієта	Вегетаріанець
Алкоголь	Так
Наркотики	Так
Куріння	Так
Освіта	Середня школа
Етнічна приналежність	Європеєць
Діти	Без дітей
Тварини	Без тварин
Релігія	Християнство
Знак зодіаку	Овен

Створити профіль

Рисунок 4.14 – Форма для заповнення клієнтської анкети

Після заповнення анкети клієнта, користувачеві шлюбного агентства відкривається можливість здійснити пошук сумісних партнерів, згідно із зазначеними вимогами. Для пошуку партнерів використовується запропонований метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Результатами підбору партнері є три переліки можливих кандидатів (Рисунок 4.15).

Шлюбне Агентство Головна Послуги Вхід

Найкращі кандидати за двома методами

<p>Amy Mooney Місто: Yorkport Стать: чоловік Вік: 44</p> <p>Переглянути профіль</p>	<p>Alyssa Wise Місто: Yorkport Стать: чоловік Вік: 44</p> <p>Переглянути профіль</p>	<p>Robert Dixon Місто: Yorkport Стать: чоловік Вік: 25</p> <p>Переглянути профіль</p>
--	---	--

Співпадіння за методом порівняння фреймів

<p>Amy Mooney Місто: Yorkport Стать: чоловік Вік: 44 Відсоток співпадіння: 72.5</p> <p>Переглянути профіль</p>	<p>David Lane Місто: Yorkport Стать: чоловік Вік: 44 Відсоток співпадіння: 67.5</p> <p>Переглянути профіль</p>	<p>Erika Sanchez Місто: Yorkport Стать: чоловік Вік: 44 Відсоток співпадіння: 63.75</p> <p>Переглянути профіль</p>
<p>Alyssa Wise Місто: Yorkport Стать: чоловік</p> <p>Переглянути профіль</p>	<p>Robert Dixon Місто: Yorkport Стать: чоловік</p> <p>Переглянути профіль</p>	<p>Richard Fox Місто: Yorkport Стать: чоловік</p> <p>Переглянути профіль</p>

Порівняння за текстовими описами

<p>Jessica Burke Місто: Lake Dennisbury Стать: чоловік Вік: 30 Відсоток співпадіння: 73.02232</p> <p>Переглянути профіль</p>	<p>Barry Strong Місто: West Kewenmouth Стать: чоловік Вік: 39 Відсоток співпадіння: 69.97162</p> <p>Переглянути профіль</p>	<p>Christopher Fields Місто: Yorkport Стать: чоловік Вік: 36 Відсоток співпадіння: 67.02326</p> <p>Переглянути профіль</p>
<p>Robert Dixon Місто: Yorkport Стать: чоловік Вік: 25 Відсоток співпадіння: 66.2224</p> <p>Переглянути профіль</p>	<p>Jessica Zhang Місто: Gometmouth Стать: чоловік Вік: 39 Відсоток співпадіння: 62.43874</p> <p>Переглянути профіль</p>	<p>Edwin Gardner Місто: Lake Kewenmouth Стать: чоловік Вік: 44 Відсоток співпадіння: 62.337854</p> <p>Переглянути профіль</p>

Рисунок 4.15– Результати пошуку сумісних партнерів

У першому переліку користувач може переглянути тих кандидатів, які потрапили у результуючі множини за методом порівнянням фреймів та порівняння за текстовими описами. Відповідно, у другому переліку наведено тих кандидатів, чий відсоток сумісності виявились найбільшими, використовуючи метод порівняння фреймів. Аналогічно, у третьому переліку – кандидати з найвищим відсотком сумісності за порівнянням текстових описів.

Таким чином, використовуючи розроблений веб-сайт шлюбного агентства, користувач має змогу зареєструватись та заповнити клієнтську анкету особистою інформацією. На основі особистої інформації та вимог до бажаного партнера, клієнт може здійснити пошук потенційних партнерів, використовуючи метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

4.5 Дослідження ефективності роботи методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних

Дослідження ефективності роботи методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних полягає в дослідженні впливу кількості сумісних партнерів зі сформованих множин на кількість партнерів у об'єднаній множині. Тобто, змінюючи параметр k , який відповідає за кількість партнерів з найвищими відсотками сумісності, які будуть відібрані в множини за результатами роботи обох методів, буде змінюватись кількість партнерів n , які потраплять у фінальну об'єднану множину.

Для дослідження ефективності роботи запропонованого підходу було проведено чотири варіанти пошуку сумісних партнерів для одного клієнта. У кожному з чотирьох варіантів використано різне значення параметру k (Таблиця 4.4). Також, варто відмітити, що параметр k , завжди однаковий для обох методів.

Таблиця 4.4 – Вплив параметру k на кількість анкет в об'єднаній множині

Кількість анкет, які будуть відібрані в множини результатів (k)	Кількість анкет, які потрапили в об'єднану множину (n)
5	0
10	2
15	6
20	9

Таким чином, за результатом проведеного експерименту можна зробити висновок, що при збільшенні розміру множин сумісних партнерів, збільшується

кількість анкет, які потрапляють в об'єднану множину. Крім цього, при низькому значенні k , в об'єднану множину може не потрапити жодної анкети.

Даний експеримент було повторено для 10 різних анкет. Для кожного значення параметру k було знайдено середнє значення параметру n . Результати даного дослідження наведено на діаграмі, зображеній на рисунку 4.16.

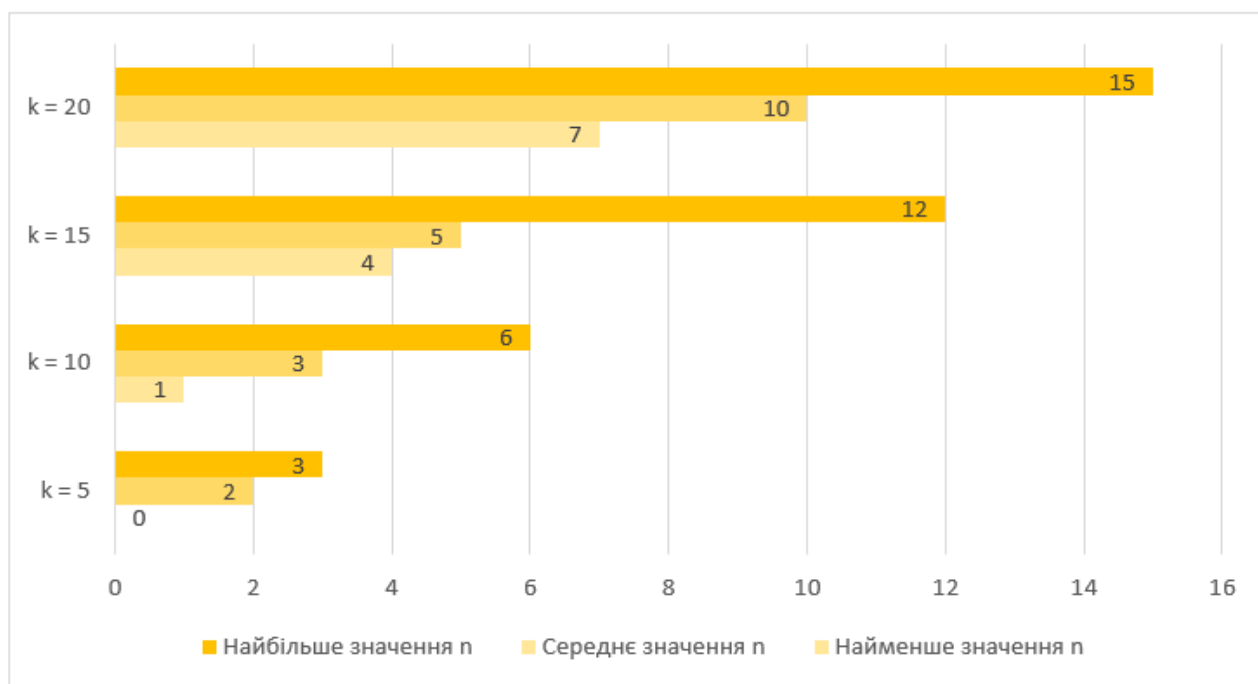


Рисунок 4.16 – Діаграма залежності кількості сумісних партнерів в об'єднаній множині від кількості сумісних партнерів в результуючих множинах обох методів

Результати описаного дослідження доводять, що при збільшенні кількості сумісних анкет, які потрапляють в результуючі множини методів, збільшується кількість анкет, які потрапляють в об'єднану множину. Це дозволить клієнту шлюбного агентства переглянути більшу кількість кандидатів та обрати серед них бажаного партнера з найбільш підходящими якостями.

Також, для перевірки ефективності створеного методу та дослідження коректності його роботи було проведено порівняння між роботою методу та експертним висновком від моделі GPT-4o [49], доступ до якої здійснювався через ChatGPT (Рисунок 4.17).

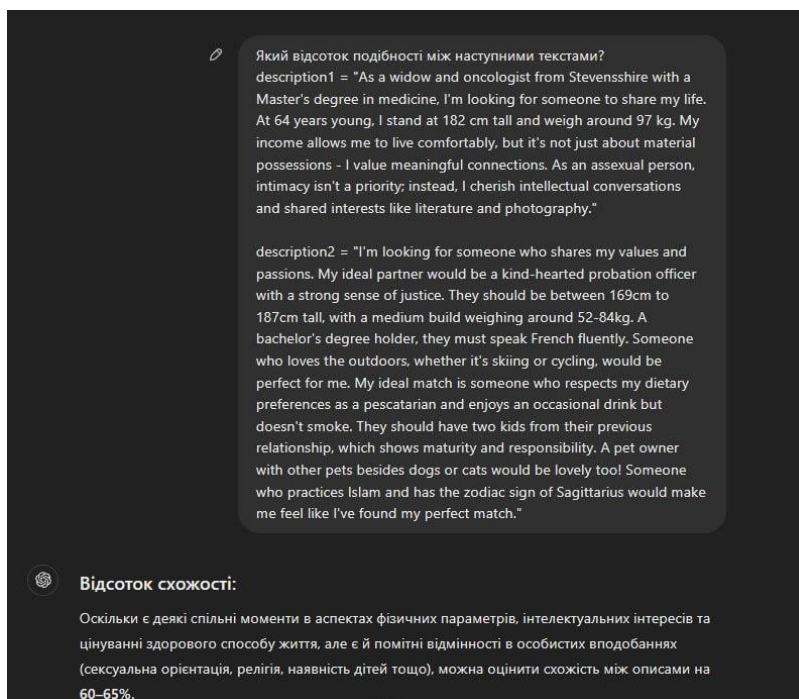


Рисунок 4.17 – Приклад отримання експертного висновку від ChatGPT

В ході дослідження порівнювались текстові описи клієнтів із бази даних шлюбного агентства, з метою обрахунку оцінки подібності текстів. Також, було встановлено допустиме відхилення від експертного висновку в межах 5%. Крім цього, було встановлено поріг сумісності та несумісності клієнтів, відповідно до якого, якщо оцінка подібності текстових описів менша 50% – пара вважається несумісною.

Для першої перевірки було використано 50 пар текстових описів клієнтів, які мали високий відсоток співпадіння за методом порівняння фреймів. В результаті проведеного дослідження, 43 перевірки із 50 виявились успішними. Таким чином, точність визначення оцінок подібності текстових описів в парах із високою сумісністю становить 86% відносно порівняння із моделлю GPT-4o.

В основному, створений метод надає більші відсотки подібності, ніж модель GPT-4o. В 38 випадках оцінка подібності текстових описів визначена створеним методом перевищувала оцінку подібності в отриманому експертному висновку. І лише в 12 випадках оцінка подібності визначена методом була меншою. У 3 вказаних випадках, оцінки подібності текстових описів були менші 50%, що свідчить про помилково визначену несумісність партнерів.

Також, було проведено перевірку визначення оцінок подібності текстових описів для несумісних партнерів. Отже, 39 перевірок із 50 виявились успішними, що говорить про те, що точність визначення оцінки схожості текстових описів в несумісних парах становить 78% відносно порівняння із моделлю GPT-4o.

Під час другого експерименту також виявилось, що переважна більшість оцінок подібності текстових описів визначених створеним методом є більші за оцінки отримані експертним висновком (42 рази із 50). Крім цього, у 6 випадках визначена оцінка подібності текстових описів перевищувала 50%, що означає, що дані пари були помилково визнані сумісними.

За отриманою інформацією було створено матрицю плутанини, яка узагальнює отримані результати проведених досліджень (Таблиця 4.5).

Таблиця 4.5 – Матриця плутанини

	Сумісні	Несумісні
Сумісні	47	6
Несумісні	3	44

Отже, відповідно до проведених досліджень було встановлено, що в порівнянні з моделлю GPT-4o, точність визначення оцінок подібності текстових описів у випадку сумісних пар становить 86%, в той час як для несумісних пар – 78%. Отже, загальна точність методу порівняння текстових описів становить 82%. Також, встановлено, що в переважній більшості проведених перевірок, створений метод надавав більшу оцінку подібності текстових описів ніж в експертному висновку, отриманому з використанням моделі GPT-4o.

Висновки до розділу 4

Було реалізовано метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, який дає змогу за вхідними даними у вигляді анкети клієнта та бази анкет кандидатів обраховувати відсотки сумісності на

основі порівняння анкетних даних за фреймовою моделлю подання та порівняння текстових описів. Прикладне застосування розробленого методу відбувається через реалізований веб-сайт шлюбного агентства. Використовуючи реалізований веб-сайт користувачі мають змогу зареєструватись та заповнити анкету клієнта, яка включає особисту інформацію про клієнта та вимоги до бажаного партнера. На основі даних із анкети клієнт може здійснити пошук сумісних партнерів.

Програмна архітектура розробленого веб-сайту складається із двадцяти шести класів, переважна більшість яких є класами для представлення сутностей із бази даних. Через ці класи відбувається отримання та додавання даних у БД. Також, реалізовано класи які відповідають за побудову графічного інтерфейсу та взаємодію користувача із веб-сайтом. Для реалізації методу визначення сумісності між клієнтами, також створено окремі класи.

Таким чином, створена програмна архітектура реалізовує процес визначення сумісності клієнтів шлюбного агентства за їхніми анкетними даними. Проведене прикладне тестування реалізованого веб-сайту шлюбного агентства доводить, що користувачі мають змогу зареєструватись, заповнити клієнтську анкету та отримати вихідні результати у вигляді переліків сумісних партнерів.

Також, було проведено дослідження ефективності роботи розробленого методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних на основі якого було встановлено, що чим більшою є кількість кандидатів у результуючих множинах обох методів, тим більше кандидатів включатиме об'єднана множина. Це дозволяє клієнту розглянути більшу кількість кандидатів та обрати бажаного партнера. Крім цього, було проведено дослідження з використанням експерта у вигляді моделі GPT-4o. В ході дослідження порівнювались результати визначення оцінки схожості двох текстових описів отримані запропонованим методом та експертним висновком. Таким чином, було встановлено, що точність роботи методу порівняння текстових описів становить 82%. Отже, даний метод може бути застосований для визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

Загальні висновки

Кваліфікаційна робота магістра розв'язує задачу покращення визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Результатом роботи є розроблений метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, представлений у вигляді підходу, що має на меті паралельне використання методу порівняння анкет за фреймовою моделлю та нейромережевого визначення оцінок схожості текстових описів клієнтів. Таким чином, використання запропонованого методу дозволяє визначати відсоток сумісності між парами клієнтських анкет, порівнюючи їхні анкетні дані.

Отже, в ході виконання кваліфікаційної роботи магістра було поставлено та вирішено такі завдання:

1. Проведено аналіз предметної області, пов'язаної із дослідженням діяльності шлюбних агентств та методами інтелектуального аналізу анкетних даних клієнтів шлюбних агентств.

2. Розроблено метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, у вигляді підходу, що має на меті паралельне використання двох методів.

3. Створено датасет для тестування роботи методу та дослідження його ефективності. За основу взяті дані які були запозичені із обраного датасету.

4. Реалізовано прикладне застосування методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних у вигляді веб-сайту шлюбного агентства.

5. Створено механізм промптів для генерації описів користувачів за анкетними даними.

6. Спроектовано структуру веб-застосунку шлюбного агенства.

7. Спроектовано структуру даталогічної моделі бази даних веб-застосунку для пошуку партнера з високою сумісністю.

8. Реалізовано веб-застосунок для пошуку партнерів за критеріями та текстовими порівняннями.

9. Досліджено ефективність роботи методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

Проведені дослідження ефективності роботи методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, доводять, що розроблений метод здатний визначати відсотки сумісності між парами клієнтів шлюбного агентства. Відповідно до проведених досліджень було встановлено, що точність визначення оцінок подібності текстових описів у випадку сумісних пар становить 86%, в той час як для несумісних пар – 78%, загальна точність методу порівняння текстових описів становить 82%. Таким чином, створений метод може бути практично застосований для визначення відсотку подібності між парами клієнтів у різноманітних платформах для знайомств та пошуку партнерів.

За темою кваліфікаційної роботи магістра автором виконано три наукові публікації. Основні наукові й практичні результати роботи доповідались [50, 51] у доповіді «Datalogic Model for Determining the Clients Compatibility Based on Questionnaires Data Analysis by Artificial Intelligence Means» на I Міжнародній науково-практичній конференції «Innovative Solutions in Science: Balancing Theory and Practice» (San Francisco, USA) 28-30 жовтня 2024 року та у доповіді «Підхід до визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних» на XVI Всеукраїнській науково-практичній конференції «Актуальні проблеми комп'ютерних наук АПКН-2024» (м.Хмельницький) 15-16 листопада 2024 року.

Перелік посилань

1. Бондарчук Ю. А., Бороліс, І. І.; Вишнеvsька, М. О. Англо-український словник юридичних термінів. Київ, 2021. 257 с.
2. Потапчук Є. М., Балукова А. І. Чинники міжособистісної сумісності майбутніх шлюбних партнерів. Причорноморський науково-дослідний інститут економіки та інновацій. Одеса, 2021. С167–171.
3. Потапчук Н., Чорноморець Д. Життєдіяльність сімей військовослужбовців в умовах війни як психологічна проблема. *Psychology Travelogs*, 2024. С 241–250.
4. StudFiles. Шлюбні агентства і зарубіжні сайти знайомств. URL: <https://studfile.net/preview/9297344/page:2/>
5. Brannen J., Collard J. *Marriages in trouble: The process of seeking help*. Taylor & Francis, 2023. 288 p.
6. Marriage Matching. Matchmaking. URL: <https://marriagematching.love/matchmaking/>
7. Ayyildiz. Choosing an Online Marriage Agency. URL: <https://ayyildizkurumsal.com/choosing-an-online-marriage-agency/>
8. Brander. Монетизація мобільних застосунків: стратегії та моделі. URL: <https://brander.ua/blog/monetyzatsiya-mobilnykh-zastosunkiv-stratehiyi-ta-modeli>
9. Cinque. Our Matchmakers Top 49 Ultimate Questions To Ask Someone in 2024 To Get To Know Them Quickly On A Date. URL: <https://cinque.com/datingincalifornia/>
10. Томенко В. І., Сачанюк-Кавецька Н. В., Дмитрієнко О. О. Інтелектуальний аналіз даних. Вісник Хмельницького національного університету. № 3. 2020 С 77–87.
11. Кушнір О. К., Чаплінський В. Р. Статистичні методи аналізу великих даних. *Modern Economics*. 2023. № 39. С 75–81. URL: [https://doi.org/10.31521/modecon.V39\(2023\)-11](https://doi.org/10.31521/modecon.V39(2023)-11).
12. Sharkawy A. N. Principle of neural network and its main types. *Journal of Advances in Applied & Computational Mathematics*. Volume 7. 2020. pp 8–19.

13. Chicco D. Siamese neural networks: An overview. *Artificial neural networks*. 2021. pp 73–94.
14. Вікіпедія. Обробка природної мови. URL: https://uk.wikipedia.org/wiki/Обробка_природної_мови
15. Choi H., Kim J., Joe S., Gwon Y. (2021, January). Evaluation of bert and albert sentence embedding performance on downstream nlp tasks. *International conference on pattern recognition*. 2021. pp. 5482–5487
16. Клещ, К. О. Підвищення ефективності алгоритмів нечіткого пошуку з використанням таблиці подібності символів. Київ, 2024. 137 с. URL: <https://ela.kpi.ua/handle/123456789/67661>
17. Serrano N., Bellogín A. Siamese neural networks in recommendation. *Neural Computing and Applications*. Volume 35. 2023. pp 13941–13953.
18. Брус А. Використання нейронних мереж для визначення схожості текстів українською мовою. Міністерство освіти і науки України, національний університет «Києво-Могилянська академія». Київ. 2020. 35 С.
19. Притула М., Оленич І. Сентимент-аналіз тексту з використанням моделей BERT та DISTILBERT. Міжнародна науково-практична конференція «проблеми комп'ютерних наук, програмного моделювання та безпеки цифрових систем». Львів, 2024. С 62–62.
20. Mega love. URL: <https://megalove.ua/>
21. Google Play. Tinder. URL: https://play.google.com/store/apps/details?id=com.tinder&utm_medium=cta&utm_campaign=website_home
22. Солодка Н. О., Шаповалов К. І., Умнов П. В. Вибір моделі представлення знань в експертних системах. Комп'ютерне моделювання та оптимізація складних систем. Матеріали VII Міжнародної науково-технічної конференції. Дніпро, 3–5 листопада, 2021 р. Дніпро, 2021. С. 115–116.
23. Корячко М. М. Соціально-психологічні чинники формування образу шлюбного партнера в юнацькому віці. Кваліфікаційна робота на здобуття освітнього

ступеня «бакалавр»: спец. 053 «Психологія». ЧНУ ім. Петра Могили. Миколаїв, 2023. 83 с.

24. Medium. Text Vectorization: Transforming Words into Meaningful Representations. URL: <https://medium.com/@divyansh3021/text-vectorization-transforming-words-into-meaningful-representations-e03b31bb3cb3>

25. Chris McCormick. BERT Word Embeddings Tutorial. URL: <https://mccormickml.com/2019/05/14/BERT-word-embeddings-tutorial>

26. Hugging Face. WordPiece tokenization. URL: <https://huggingface.co/learn/nlp-course/chapter6/6>

27. Oksim. Що таке токенизація в НЛП? URL: <https://www.oksim.ua/2023/07/25/shho-take-tokenizacziya-v-nlp/>

28. Козенко О.В., Мазурець О. В., Молчанова М. О., Собко О. В. Використання метрик косинусної схожості та індексу Жаккара для інтелектуального аналізу семантичної подібності текстових документів. Актуальні проблеми комп'ютерних наук 2023. Хмельницький, 2023. С 146–147

29. Kaggle. URL: <https://www.kaggle.com>

30. Kaggle. OkCupid Profiles. URL: <https://www.kaggle.com/datasets/andrewmvd/okcupid-profiles/data>

31. OkCupid. URL: <https://www.okcupid.com>

32. IT-notes. Python Faker – генерація фейкових даних. URL: <https://www.it-notes.wiki/python/python-faker/>

33. OpenAI. ChatGPT. URL: <https://openai.com/index/chatgpt>

34. OpenAI. OpenAI API. URL: <https://openai.com/index/openai-api/>

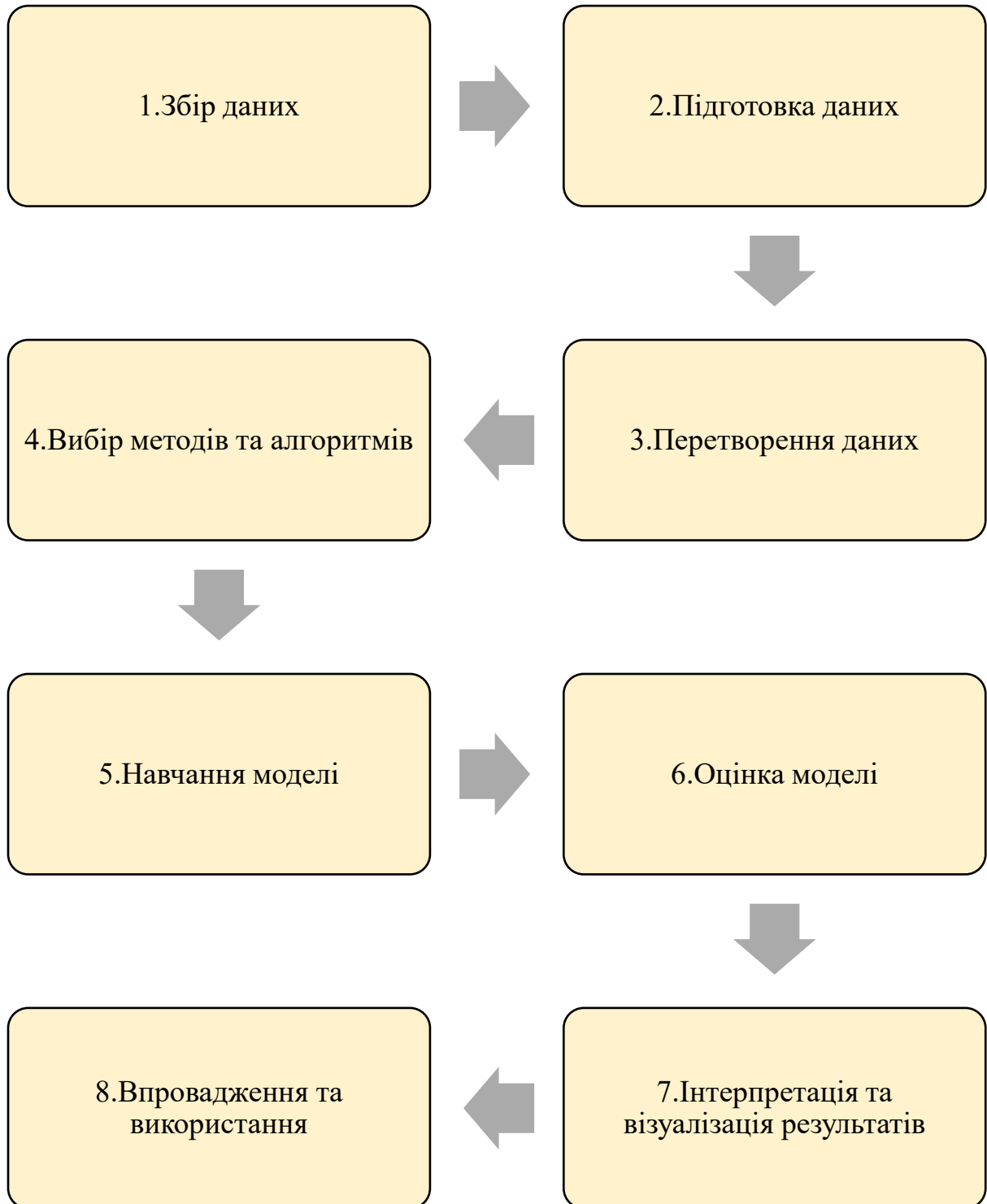
35. Prompt. Промти для ChatGPT українською мовою: як правильно скласти запит? URL: <https://prompt.com.ua/promti-chatgpt-ukrainskoju>

36. Mindner L., Schlippe T., Schaaff K. Classification of human-and ai-generated texts: Investigating features for chatgpt. International Conference on Artificial Intelligence in Education Technology. Springer, Singapore. 2023. pp 152–170

37. DAN IT. Що таке Python і де він використовується. URL: <https://dan-it.com.ua/uk/blog/python-cho-je-to-za-jazyk-programmirovanija-i-gde-ego-ispolzujut/>

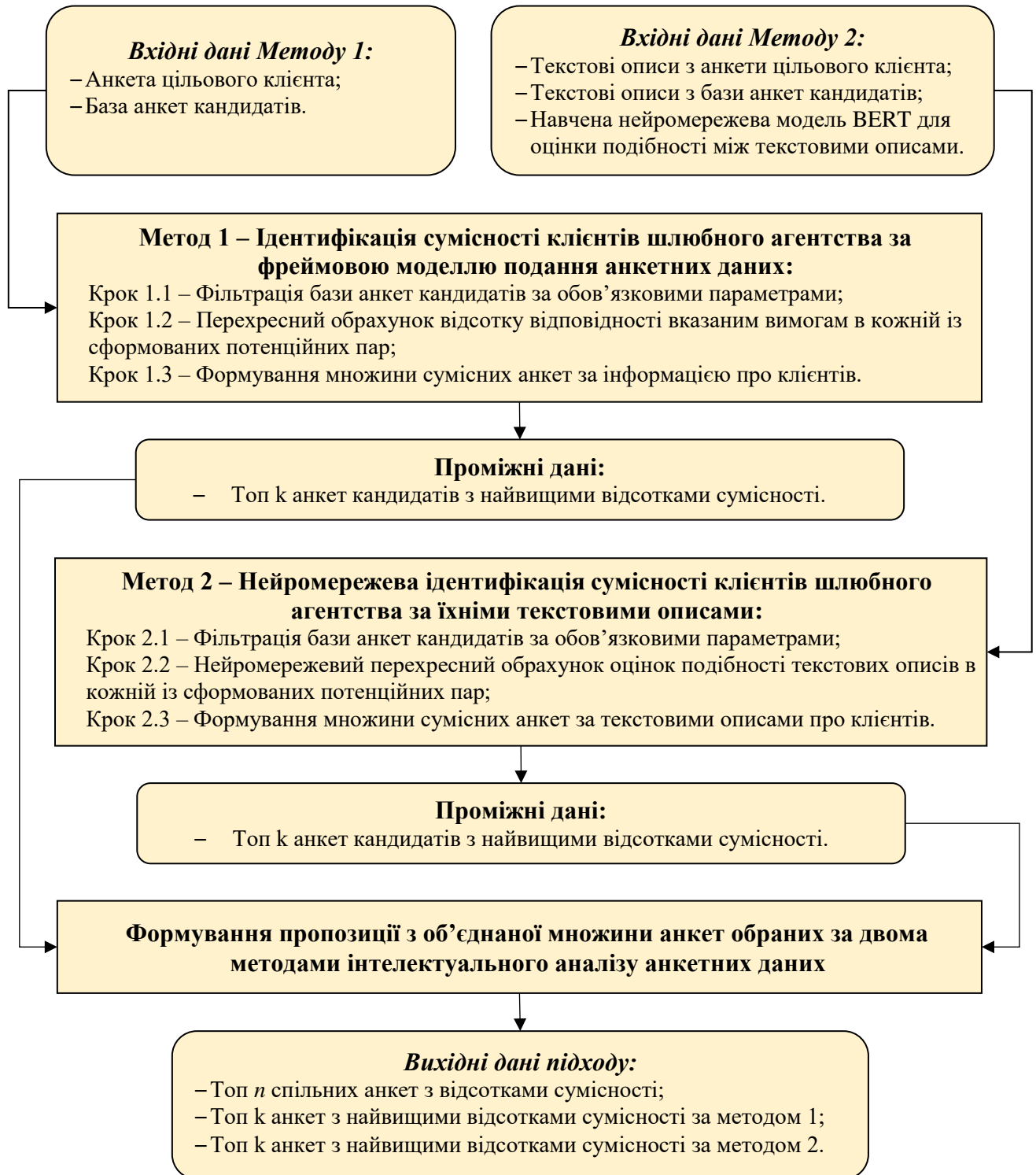
38. DOU. Рейтинг мов програмування 2024. URL: <https://dou.ua/lenta/articles/language-rating-2024/>
39. JetBrains. PyCharm. URL: <https://www.jetbrains.com/pycharm/>
40. Flask. Welcome to Flask. URL: <https://flask.palletsprojects.com/en/stable/>
41. SQLAlchemy. The Python SQL Toolkit and Object Relational Mapper. URL: <https://www.sqlalchemy.org/>
42. PyPI. Faker. URL: <https://pypi.org/project/Faker/>
43. SEO name. ChatGPT, перспективи розвитку, інтеграція. URL: <https://xn--7sbbaqhlkm9ah9aiq.net/ua/news-new/shatgpt-perspektivi-razvitiya-integratsiya.html>
44. PyTorch. URL: <https://pytorch.org>
45. Hugging Face. Transformers. URL: <https://huggingface.co/docs/transformers/index>
46. Hugging Face. URL: <https://huggingface.co>
47. Hugging Face. BertTokenizer. URL: https://huggingface.co/docs/transformers/v4.46.3/en/model_doc/bert#transformers.BertTokenizer
48. Hugging Face. BertModel. URL: https://huggingface.co/docs/transformers/model_doc/bert
49. OpenAI. Hello GPT-4o. URL: <https://openai.com/index/hello-gpt-4o/>
50. Zhuk D., Mazurets O., Sobko O., Klimenko V. Datalogic Model for Determining the Clients Compatibility Based on Questionnaires Data Analysis by Artificial Intelligence Means. Innovative Solutions in Science: Balancing Theory and Practice. Proceedings of the 1st International scientific and practical conference. October 28-30, 2024. San Francisco, USA. 2024. Pp. 80-85.
51. Жук Д.І., Мазурець О.В., Кадинська В.Д., Тищенко О.О. Підхід до визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 208-214.

ДОДАТКИ

Додаток А**Процес інтелектуального аналізу даних для визначення сумісності клієнтів шлюбних агентств**

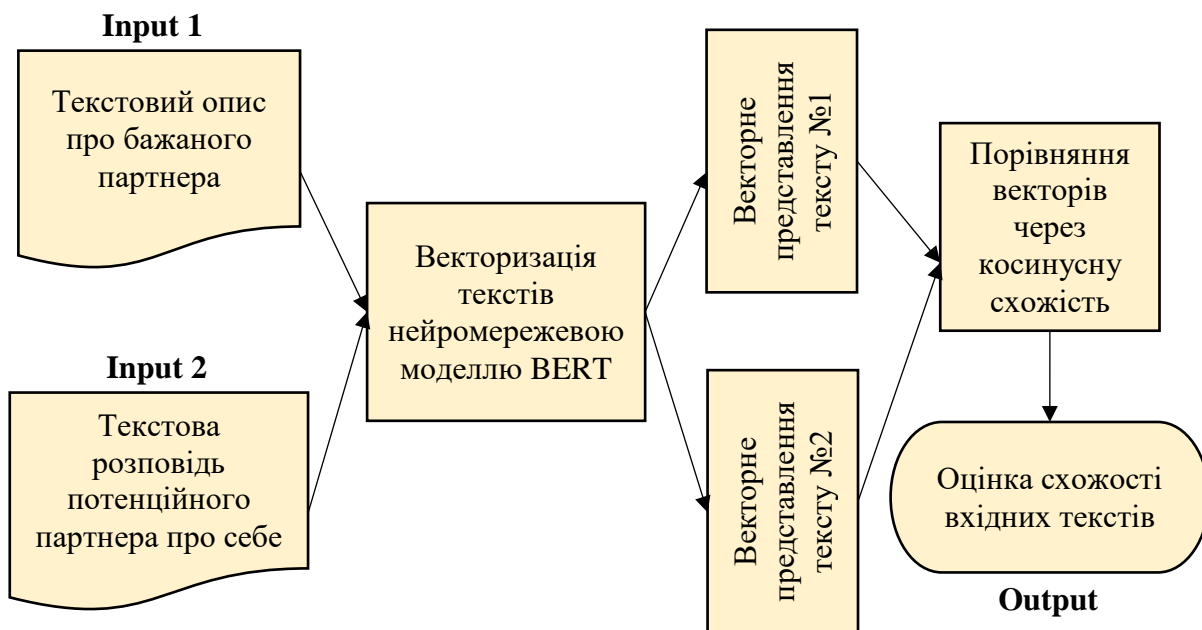
Додаток Б

Схема підходу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних



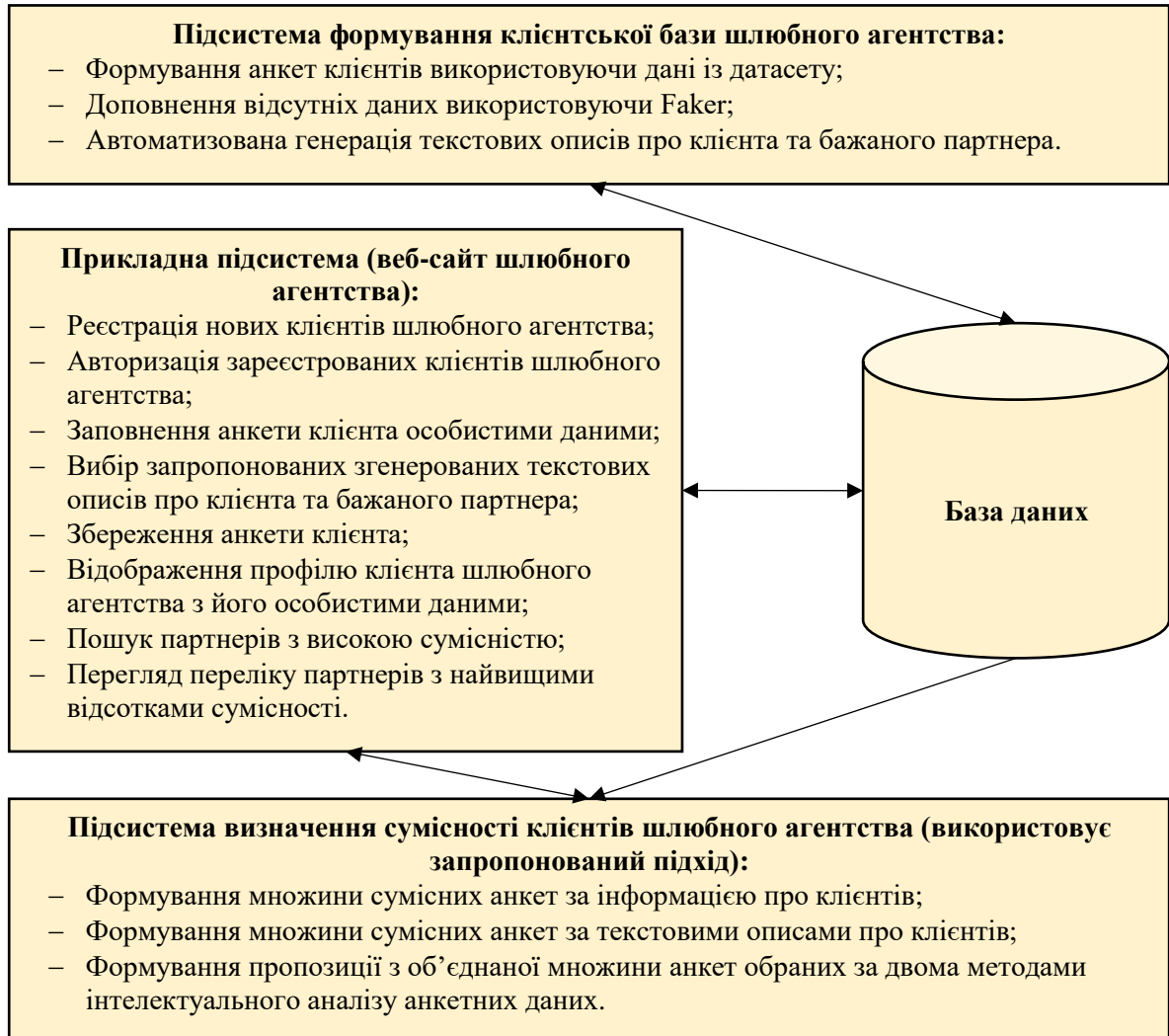
Додаток В

Схема методу обчислення оцінки схожості для однієї пари текстових описів



Додаток Г

Схема структури інформаційної системи визначення сумісності клієнтів шлюбних агентств



Додаток Д
Світлини наукових публікацій, виконаних при роботі
над кваліфікаційною роботою магістра

Наукові публікації:

1. Zhuk D., Mazurets O., Sobko O., Klimenko V. Datalogic Model for Determining the Clients Compatibility Based on Questionnaires Data Analysis by Artificial Intelligence Means. Innovative Solutions in Science: Balancing Theory and Practice. Proceedings of the 1st International scientific and practical conference. October 28-30, 2024. San Francisco, USA. 2024. Pp. 80-85.

2. Жук Д.І., Мазурець О.В., Кадинська В.Д., Тищенко О.О. Підхід до визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 208-214.

3. Жук Д.І., Мазурець О.В., Залуцька О.О., Денисенко Б.О. Реляційна персоніфікована даталогічна модель даних для визначення сумісності анкетних даних нейромережевими засобами. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2025. №1. (Довідка з редакції).

eoss-conf.com

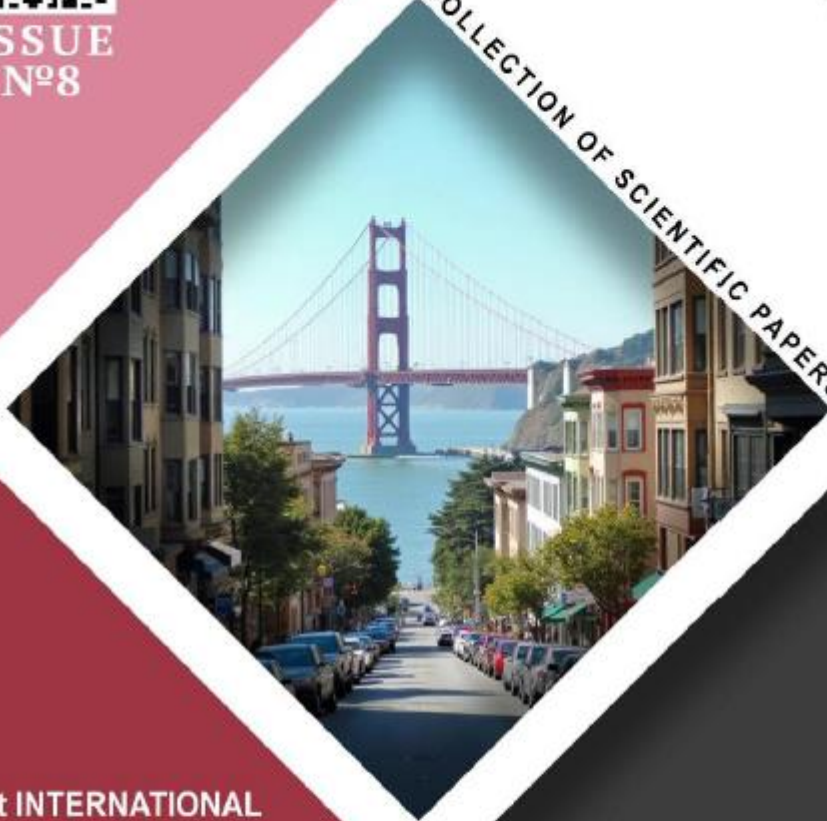


ISSUE
Nº8



EUROPEAN OPEN
SCIENCE SPACE

COLLECTION OF SCIENTIFIC PAPERS



1st INTERNATIONAL
SCIENTIFIC
AND PRACTICAL
CONFERENCE

INNOVATIVE SOLUTIONS
IN SCIENCE:
BALANCING THEORY
AND PRACTICE

OCTOBER 28-30, 2024, SAN FRANCISCO, USA



<i>Zhuk D., Mazurets O., Sobko O., Klimenko V.</i> DATALOGIC MODEL FOR DETERMINING THE CLIENTS COMPATIBILITY BASED ON QUESTIONNAIRES DATA ANALYSIS BY ARTIFICIAL INTELLIGENCE MEANS.....	80
<i>Zdolbitska N., Kurach N.</i> USE OF DYNAMIC FORMS FOR ANALYSIS OF DATA USING EXCEL AND VBA.....	85
<u>Section: Jurisprudence</u>	
<i>Піняк М.І., Цвігун Ю.А.</i> ВПЛИВ ЄВРОІНТЕГРАЦІЇ НА ДЖЕРЕЛА МІЖНАРОДНОГО ПРАВА.....	87
<i>Гольдберг Н.О., Дмитренко М.В.</i> ПРОБЛЕМИ І ТАКТИКА ПРОВЕДЕННЯ ОГЛЯДУ МІСЦЯ ПОДІЇ В УМОВАХ ВОЄННОГО СТАНУ.....	90
<i>Афтанасіє В., Романів Х.</i> ЗАХИСТ НАЙКРАЩИХ ІНТЕРЕСІВ ДИТИНИ В УМОВАХ ЗБРОЙНОГО КОНФЛІКТУ.....	93
<i>Піняк М., Витязанець Т.</i> СПІВВІДНОШЕННЯ МІЖНАРОДНОГО ПУБЛІЧНОГО ТА ПРИВАТНОГО ПРАВА В УМОВАХ ІНТЕРНАЦІОНАЛІЗАЦІЇ: СУПЕРЕЧНОСТІ НОРМ, ЇХ УЗГОДЖЕННЯ.....	96
<i>Вакал О.А., Журавель О.А.</i> СПЕЦІАЛЬНА ФІЗИЧНА ПІДГОТОВКА В УМОВАХ ОБМЕЖЕНОГО ПРОСТОРУ ТА РЕСУРСІВ.....	99
<i>Лаба Я.</i> ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ ПРОЦЕСУАЛЬНОГО СТАТУСУ АДВОКАТА В АДМІНІСТРАТИВНОМУ ПРОЦЕСІ.....	101
<u>Section: Management, Public Administration and Administration</u>	
<i>Васильчак С., Жидяк О., Соловій С., Бівчарук О.</i> СУТНІСТЬ ТА МІСЦЕ ЛОГІСТИЧНОЇ ДІЯЛЬНОСТІ В СИСТЕМІ УПРАВЛІННЯ ПІДПРИЄМСТВОМ.....	106



Використання сучасних технологій, таких як Python, PostgreSQL, SQLAlchemy та KivyMD, дозволило створити зручну та гнучку платформу для користувачів різного рівня знань. Завдяки інтеграції бази даних і функцій захисту облікових записів користувачів, система забезпечує безпечне та ефективне збереження даних.

Впровадження такої автоматизованої системи сприяє поліпшенню якості навчального процесу, дозволяючи користувачам більш ефективно засвоювати матеріал, працювати з інтерактивними вправами та отримувати миттєвий зворотний зв'язок на свої дії.

Список використаних джерел

1. Future language learning with emerging technologies - Journal of Computers in Education. SpringerLink. URL: <https://link.springer.com/article/10.1007/s40692-023-00285-9> (date of access: 20.10.2024)
2. Exploring the impact of technology on foreign language learning: a multivariate meta-meta-analysis study - Educational technology research and development. SpringerLink. URL: <https://link.springer.com/article/10.1007/s11423-024-10412-7> (date of access: 20.10.2024).
3. Self-directed use of mobile devices for language learning beyond the classroom | ReCALL | Cambridge Core. Cambridge Core. URL: <https://www.cambridge.org/core/journals/recall/article/selfdirected-use-of-mobile-devices-for-language-learning-beyond-the-classroom/C60A8CE7FA2F2D3CC4548398BDAB828B> (date of access: 20.10.2024).

DATALOGIC MODEL FOR DETERMINING THE CLIENTS COMPATIBILITY BASED ON QUESTIONNAIRES DATA ANALYSIS BY ARTIFICIAL INTELLIGENCE MEANS

Zhuk Dmytro
 Postgraduate student
 Mazurets Oleksandr
 Ph.D in Engineering Science, Associate Professor
 Sobko Olena
 Teacher
 Klimenko Valeriia
 Teacher
 Computer Science Department
 Khmelnytskyi National University, Ukraine

Previously, people met on the street, in bars or at other public events, that is, in places where other people are present. Today, one more method can be added to this method - online dating. Online dating has certain advantages. First of all, if you don't



like a person, you can simply ignore them. Also, the large influx and number of active users of such services provides a large selection of candidates for communication, which leads to the need for their filtering. In the 19th century, this procedure took place at the level of parents. Usually it was just told who you should marry or which girl to marry, but now this procedure has changed and in many cases freedom is given in communication and choosing a partner.

In order to start searching for people on dating sites, you need to create an account and fill out a questionnaire. In this questionnaire, users provide information about themselves, which helps to match interests and needs. Usually, personal data such as name, date of birth, height, weight, physique, as well as information about hobbies, work and bad habits are indicated in the questionnaire. In addition, in the questionnaire, you can indicate your dating goals, for example, looking for a serious relationship or just communication.

This information helps to select potential acquaintances that match the user's personal requirements and preferences. It helps to improve the chances of meeting the right person with whom you can develop a relationship based on common interests and goals.

The purpose of the work is the development of datalogic model for determining the clients compatibility based on questionnaires data analysis by artificial intelligence means

The frame comparison algorithm can be used to evaluate the compatibility of user questionnaires using a frame representation of knowledge. This algorithm compares the requirements and expectations of users, which were entered in the questionnaire, with the data of other users on the dating site.

In the frame comparison algorithm, criteria can be used to evaluate matches. In this case, these criteria include two surveys to form appropriate frames. The first frame contains information about the user, and the second frame contains information about the desired partner. The main criteria for comparing frames: common interests; compatibility of education; compatibility of marital status; compatibility of physical characteristics.

To start analyzing the questionnaire by the system, the user needs to set the characteristics and criteria by which the system will check the compatibility of the questionnaires. Verification of these questionnaires will be carried out by performing a certain test, in which points will be given for each matching criterion.

To determine the degree of compatibility between users, it is necessary to establish compatibility rules. For example, if user A is looking for a partner between the ages of 25 and 30, then users outside of that age range will not receive compatibility points.

After comparing questionnaires according to these criteria, the system displays the percentage of questionnaire compatibility. That is, the system compares two frames: the frame of the specific user and the frame of the user's expectations. The result of compatibility is a number that is calculated based on the matches of the given criteria. The more criteria match, higher percentage of compatibility will be.



The mathematical model of the compatibility method will have the following form $\sum_{n=1}^k \frac{1}{k}$, where k is the total number of evaluation criteria.

If the criteria have sub-criteria, then the maximum percentage for the criterion will be the sum of the matches of the sub-criteria. This is necessary, for example, for a hobby. If the total number of criteria is 10, then 10% is added for matching one criterion. If, for example, this criterion is a hobby of the user, then there can be many of them. For example, if the user has four hobbies, then 2.5% will be added for matching one.

The input data for the system is data about the desired user. This data set includes: gender, age, height, income level, education, occupation, city, hobbies, and marital status.

Stage 1. At this stage, the system forms a frame from the input data for further comparison.

Stage 2. The system compares user frames and the searched frame. According to the coincidence between the given frames, each frame has its own percentage of coincidence. After comparing all users, the system automatically filters out all questionnaires with a matching percentage of less than 50%.

Stage 3. The system indicates which of the candidate's parameters did not match the expected result. Sorts the list of applications by the percentage of matches, and sends the user the top 5 candidates.

Output data. The system forms an expert opinion for each questionnaire. The user is offered a list of the top 10 candidates.

The corresponding information system is intended for use by many users. Therefore, it is necessary to store data about these users somewhere [1]. For this, it is necessary to design and create a database [2, 3]. To create a database, you need to create tables and relations between these tables [4, 5]. It was necessary to create a data logic model of the database, which is depicted in Figure 1.

This policy has the following entities: frames of professions; user frames (UserFrames); hobby frames; cities; countries; degrees of education; education; sex

To describe the essence of the user's profession, you need to create a table with the following fields: Id, Name, Description.

In order to attach a frame to a user's account, you need to create a "UserFrames" table with the following fields: Id, Name, Surname, Education, Income, Age, Sex, PersonalityType, Height, Weight, Stature, ProfessionId, CityId.

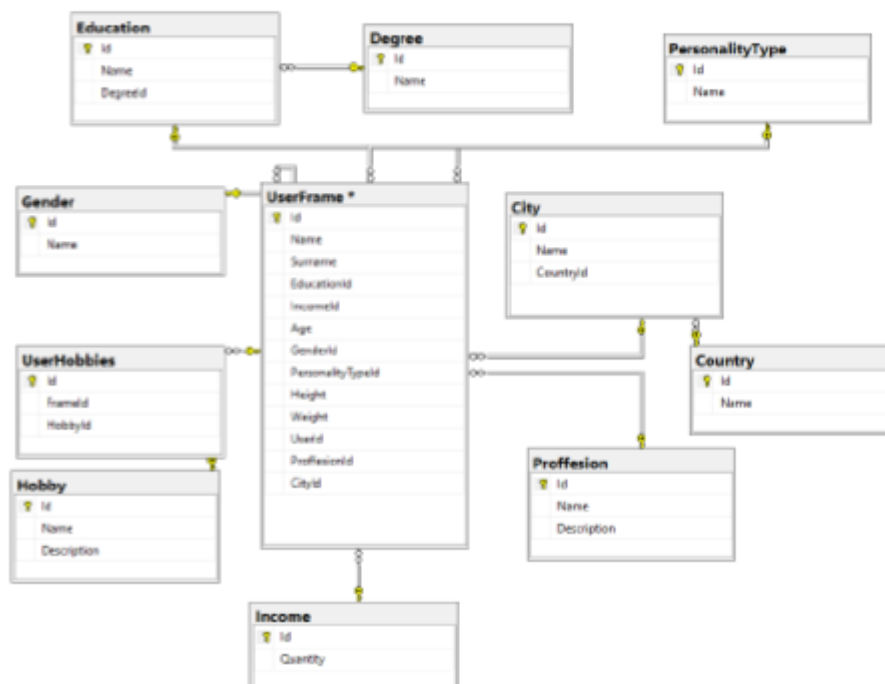


Figure 1. Datalogic model for determining the clients compatibility

To describe the essence of the user's city of residence, you need to create a table with the following fields: Id, Name, CountryId. To describe the entity of the country, you need to create a table with the following fields: Id, Name. To describe the hobby entity, you need to create a "Hobbies" table with the following fields: Id, Name, Description. A user can have many hobbies, and for this you need to make an additional table listing the user's hobbies. This table should have the following fields: Id, UserId, HobbyId. To describe the entity of gender, you need to create a table "Gender" with the following fields: Id, Name. To describe the entity of gender, you need to create a table "Gender" with the following fields: Id, Name. The "Degree" table is created to record degrees of education and has the following structure. To describe the personality type entity, you need to create a "PersonalityType" table with the following fields: Id, Name. To describe the income entity, you need to create the "Income" table with the following fields: Id, Quantity.

Also, the following data logical data model was generated to save the data of users of the dating website (Figure 2). This database data model is needed to manage user accounts and link frames to accounts.



Figure 2. Datalogic model of the website database

Therefore, the Datalogic Model for Determining the Clients Compatibility Based on Questionnaires Data Analysis by Artificial Intelligence Means was designed. To determine compatibility between users, the system uses a test in which points will be added for each matching criterion. After comparing the questionnaires according to these criteria, the system displays the percentage of compatibility of the questionnaires, which is calculated based on the matches of the given criteria. The more criteria match, the higher the match percentage. The need to create a database for the information system, which is intended for many users, was also described in detail. Table structures and fields for storing data on frames, professions, hobbies, cities and countries, gender, personality types, income, education, and education levels are described. It was also pointed out the need to create an additional table to store information about the hobbies of users, since they can have many different hobbies. Using a database will allow efficient and convenient storage and processing of data about users and their interests.



References

1. Mazurets O. V. Datalogic Projecting for Applied Adaptive Semantic Testing. Heritage of European Science. International collective monograph. Monographic series «European Science». Karlsruhe, Germany. Book 27. Part 3. Chapter 8. 2024. P.147-156.
2. Mazurets O., Sobko O., Klimenko V., Kozenko Y. Relation Datalogic Model for Determining the Diagnosis Based on Intellectual NLP-analysis of Symptom Description. Proceedings of XV International Scientific and Practical Conference «Innovative Development: Synthesis of Scientific Approaches in Various Fields of Research». March 20-22, 2024. Tallinn, Estonia. 2024. Pp. 61-66.
3. Mazurets O., Molchanova M., Klimenko V., Klopotivskiy D. Datalogic Model for Image Recognition by Convolutional Neural Network Using Cloud Services. Proceedings of XXII International Scientific and Practical Conference «Modern Scientific Research: Theoretical and Practical Aspects». May 8-10, 2024. Oslo, Norway. 2024. Pp. 64-68.
4. Mazurets O., Sobko O., Vit R., Pasternak V. Practical Approach for Detection by Deep Learning of Target Objects of Subject Area Based on Semantic Connectivity Indicators in Audio Database. Proceedings of XXIV International Scientific and Practical Conference «Modern Scientific Challenges are the Driving Force of the Development of Scientific Research». May 22-24, 2024. Bruges, Belgium. International Scientific Unity. 2024. Pp. 91-96.
5. Mazurets O., Uspenska K., Vit R., Tyschenko O. Intelligent System for Determining the Object Attributes Values by Neural Networks Means by Graphic Images in Databases. Current Trends in the Development of Scientific Research in Today's Conditions. Proceedings of XXV International scientific and practical conference. May 29-31, 2024. International Scientific Unity. Florence, Italy. 2024. Pp. 86-91.

USE OF DYNAMIC FORMS FOR ANALYSIS OF DATA USING EXCEL AND VBA

Zdolbitska Nina
 Ph.D., Associate Professor
 Department of Computer Sciences
 Kurach Nazar
 Lutsk National Technical University

Big data analytics as a new innovative tool in e-commerce plays an important role in helping enterprises. The study results can improve the implementation of big data analytics and digital technologies in organizations.

When working with a software product, there may be a need to dynamically change the number of controls on a user form, such as using dynamic buttons, option buttons, entering complex formulas, etc. This depends on the fact that it is not known in advance how many buttons may be needed when working with the application, or

Міністерство освіти і науки України
Хмельницький національний університет



ЗБІРНИК НАУКОВИХ ПРАЦЬ
за матеріалами XVI Всеукраїнської науково-практичної конференції
«Актуальні проблеми комп'ютерних наук АПКН-2024»

15-16 листопада 2024

Хмельницький 2024

Денисенко В.О., Мельников О.Ю. Вдосконалення наявного додатку для оброблення інформації про лісові насадження	175
Дидо Р.А., Мазурець О.В., Кліменко В.І. Інформаційна система для нейромережевої інтерактивної ідентифікації особистості за зображенням обличчя	180
Дідур В.О. Нейромережева класифікація залишків будівництва	187
Діхтяр М.О., Радюк П.М., Скрипник Т.К. Метод інтерпретування результатів виявлення патологій серця за зображенням МРТ	189
Драган Т.С., Галка А.О., Ніколайчук М.С., Джулій В.М. Алгоритм передачі конфіденційної інформації без спотворення растрового зображення	192
Жайворон Д.О., Пасічник О.А., Скрипник Т.К., Манзюк Е.А. Метод ідентифікації лікарських рослин за аналізом зображень нейромережевими засобами	196
Жарновський О.В., Казмірчук Я.М., Собко О.В., Мазурець О.В. Практична реалізація методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання	198
Жарчинський С.М. Система надійного зберігання даних на основі Openstack Object Storage	205
Жук Д.І., Мазурець О.В., Кадинська В.Д., Тищенко О.О. Підхід до визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних	208
Загребельний В.В., Кльоц Ю.П. Роль OSINT у протидії кіберзлочинності та веденні інформаційної війни	215
Зайцев І.О., Федоров Є.Є. Кіберфізична система для інтерактивного відображення доступності міської інфраструктури	218
Залуцька О.О. Метод автоматизованого оцінювання відповідності тональності відгуків на товари в інтернет-магазинах до їх користувацької оцінки з використанням нейромереж	221

УДК 004.8

Жук Д.І., Мазурець О.В., Кадинська В.Д., Тищенко О.О.

*Хмельницький національний університет***ПІДХІД ДО ВИЗНАЧЕННЯ СУМІСНОСТІ КЛІЄНТІВ ШЛЮБНИХ АГЕНТСТВ ЗА ІНТЕЛЕКТУАЛЬНИМ АНАЛІЗОМ АНКЕТНИХ ДАНИХ**

Запропоновано метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, який обчислює загальний відсоток сумісності пар, враховуючи вимоги до бажаних партнерів від обох сторін. Для реалізації запропонованого методу використовуються персональні дані клієнтів шлюбного агентства, взяті із знайденого датасету, який є у вільному доступі. На основі цих даних генеруються персональні текстові описи про клієнта і його вимоги до бажаного партнера, використовуючи існуючі нейромережі-трансформери. Для порівняння анкетних даних та обчислення відсотку сумісності використовується архітектура сіамської нейронної мережі.

The method for determining the compatibility of clients of marriage agencies based on the intelligent analysis of questionnaire data is proposed, which calculates the overall percentage of compatibility of couples, taking into account the requirements for desired partners from both sides. To implement the proposed method, the personal data of the clients of the marriage agency, taken from the found dataset, which is freely available, is used. On the basis of this data, personal text descriptions about the client and his requirements for the desired partner are generated, using existing neural networks-transformers. Siamese neural network architecture is used to compare questionnaire data and calculate the compatibility percentage.

Сучасний темп життя, особливо у містах, спричиняє нестачу вільного часу, який можна було б присвятити побудові особистого життя, включаючи створення сім'ї. Надмірне навантаження на роботі, що забирає значну частину часу та особистих ресурсів, позбавляє людей можливості для знайомств, обмежуючи пошук потенційних майбутніх партнерів. У таких умовах звернення до послуг шлюбного агентства може стати ефективним рішенням, що дозволить знайти партнера, економлячи значну частину часу на пошуки.

Ще однією проблемою сучасності є підвищена вимогливість до партнерів, адже сучасне суспільство з кожним роком стає все більш вимогливим. Це пов'язано із значним розвитком та популяризацією психології, завдяки чому люди стають більш обізнані в тому як поводити себе в стосунках. Тому, людям значно важче знайти партнера, який буде відповідати всім поставленим якостям.

Місце проживання, також, має великий вплив на процес пошуку партнерів, що особливо загострилось під час війни, коли багато людей були змушені виїхати за кордон [1]. В таких умовах, звернення до шлюбних агентств дає можливість врахувати географічне розташування при підборі партнерів.

Основна особливість діяльності шлюбних агентств полягає в ретельному підборі кандидатів для клієнтів, враховуючи їхні вимоги та побажання. Цей процес може охоплювати як традиційні індивідуальні консультації із спеціалістами, так і сучасні системи інтелектуального аналізу даних, що здатні визначати рівень сумісності на основі інформації з клієнтських анкет. Крім того, шлюбні агентства можуть надавати деякі додаткові послуги, такі як: психологічна підтримка, сімейна терапія та навчання по покращенню соціальних навичок.

Шлюбні агентства можуть різнитись за спеціалізацією та методами підбору партнерів, тому їх поділяють на три основні типи: традиційні шлюбні агентства, нішеві шлюбні агентства та шлюбні агентства онлайн [2].

Традиційні шлюбні агентства роблять акцент на особистому контакті з клієнтом. Вони зазвичай мають офіс, в який клієнти можуть звертатись для отримання послуг. Фахівці шлюбного агентства проводять з клієнтами індивідуальні консультації та створюють профіль клієнта, що слугуватиме основою для підбору потенційних партнерів. Додатково, клієнти можуть заповнювати анкети про себе, вказуючи особисту інформацію, таку як вік, стать та рід діяльності. Після збору інформації про клієнта, спеціалісти підбирають кандидата або декількох кандидатів та влаштовують їм особисті зустрічі, на яких люди мають змогу познайомитись та дізнатись більше один про одного, що допоможе їм визначити чи підходять вони одне одному [5]. Основною перевагою традиційного шлюбного агентства є те, що всі питання, пов'язані з підбором кандидатів і організацією зустрічей бере на себе агентство, що дозволяє клієнту зекономити багато часу. Однак, такі послуги можуть виявитись досить дорогими та довготривалими через обмежену клієнтську базу, яка може бути на момент звернення.

Нішеві шлюбні агентства зосереджені на пошуку партнерів для окремих груп. Наприклад, такі агентства можуть обслуговувати лише заможних клієнтів, які готові інвестувати великі суми для пошуку партнерів рівних за соціальним статусом. Також, нішеві шлюбні агентства можуть спеціалізуватися на підборі партнерів для клієнтів із певних релігійних груп, національностей чи культурних спільнот. Ці агентства орієнтовані на індивідуальні особливості групи людей із якими вони працюють, тому й звернення до нішевих шлюбних агентств може бути корисним для людей із певних соціокультурних груп.

Інший тип шлюбних агентств – це шлюбні агентства онлайн, які працюють виключно в інтернет-просторі. Вони можуть бути представлені у вигляді веб-сайтів або мобільних додатків, де клієнти самостійно реєструються та заповнюють анкетні дані. Пошук потенційних партнерів відбувається за допомогою спеціальних алгоритмів, які аналізують анкетні дані користувачів і пропонують їм список кандидатів із високим рівнем сумісності. Користувачі самі обирають з ким вони хотіли б познайомитися. Якщо, обоє з користувачів мають бажання дізнатись більше один про одного, вони можуть розпочати спілкування за допомогою вбудованих чатів.

Враховуючи соціальні проблеми сучасності, саме шлюбні агентства у онлайн форматі є найпопулярнішим та найефективнішим видом шлюбних агентств, через їхню простоту використання та доступність у будь-який час та незалежно від розташування. Крім цього, це досить економний варіант, адже користувачам подібних платформ доступний базовий безкоштовний варіант доступу до функціоналу та можливість придбання підписок різної вартості.

Отже, шлюбні агентства онлайн є найпопулярнішим та найефективнішим варіантом надання послуг з пошуку партнерів. Це зумовлено їхньою доступністю та простотою використання, адже звернутись до такого агентства можна з будь-якої точки світу. Ключовим етапом роботи шлюбних агентств є процес заповнення анкети клієнта, в яку людина додає як свою персональну інформацію, так і вимоги до бажаного партнера. Збереження анкетних даних користувачів у базі даних надає великий спектр можливостей для проведення інтелектуального аналізу збережених анкет, що є важливим при автоматизації процесів підбору партнерів. Тому, заповнення анкет та їхня подальша обробка є основою для реалізації методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

Метою роботи є розробка методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, та його прикладна реалізація у вигляді веб-сайту для пошуку партнера з високою сумісністю.

Для автоматизованого визначення сумісності клієнтів шлюбних агентств за їхніми анкетними даними необхідно використовувати сучасні методи інтелектуального аналізу даних. Інтелектуальний аналіз даних (ІАД) являє собою процес виявлення прихованих, невідомих або неочевидних закономірностей у великих обсягах даних на основі використання математичних моделей, статистичних методів чи алгоритмів штучного інтелекту [3, 4]. Основною метою інтелектуального аналізу даних є перетворення необроблених даних на інформацію, яка може бути використана для прийняття рішень, прогнозування чи оптимізаційних процесів [5].

Обраний метод потрібно адаптувати під вирішення конкретної задачі та навчити на відповідному наборі навчальних даних [6, 7]. Під час навчання метод аналізує зв'язки між вхідними даними та навчається знаходити закономірності [8].

Після проведення навчання необхідно провести оцінку точності роботи методу. Метод оцінюється на даних, які не були використанні під час навчання, що дозволяє оцінити якість роботи методу з незнайомою інформацією [9]. Для цього, також, використовуються різні метрики, вибір яких залежить від обраного методу [11].

Якщо навчання методу досягло задовільного рівня то можна переходити до наступного етапу, який полягає в інтерпретації та візуалізації отриманих результатів. Для цього створюються різноманітні графіки та діаграми, які допомагають краще зрозуміти отримані результати та прийняти подальше рішення щодо використання натренованого методу [12].

Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних складається із трьох окремих етапів (Рисунок 1). На першому етапі формуються анкети клієнтів на основі персональних даних із датасету. Другий етап передбачає підготовку анкетних даних для проведення інтелектуального аналізу. Останній третій етап присвячений обрахунку відсотків сумісності клієнтів шлюбних агентств на основі їхніх анкетних даних.

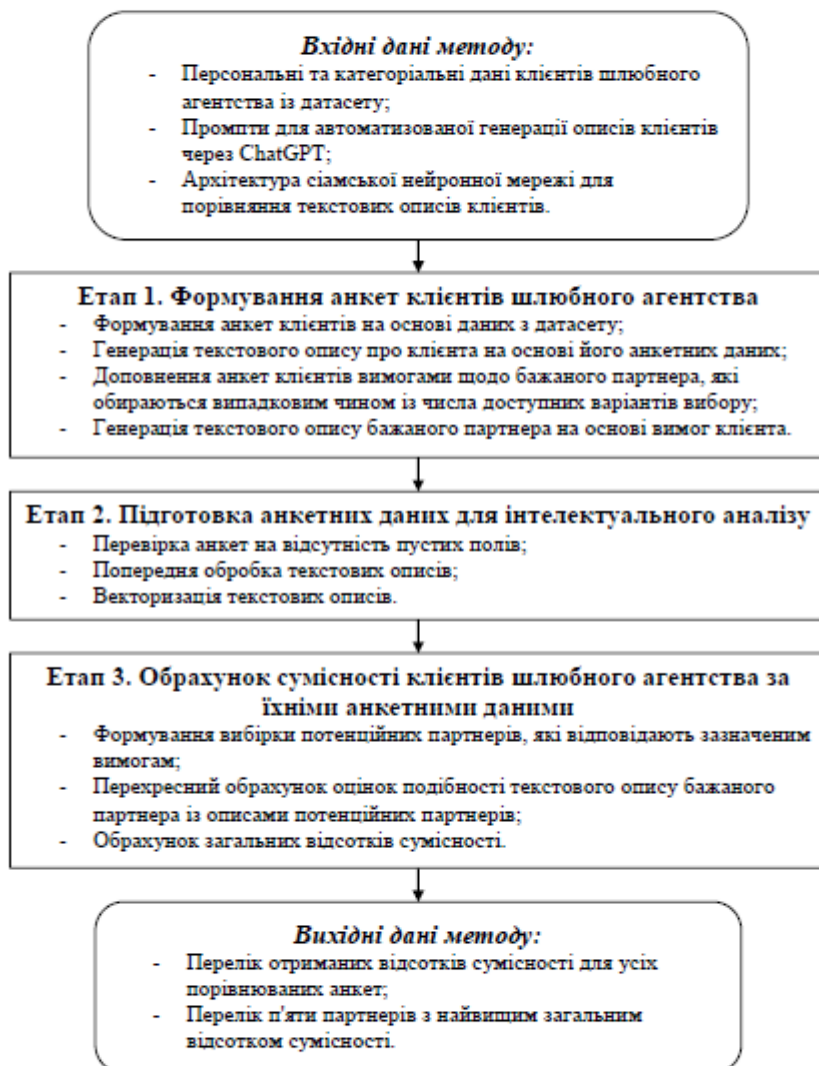


Рисунок 1 – Схема методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних

Для реалізації запропонованого методу використовуються персональні дані клієнтів шлюбного агентства, взяті із знайденого датасету, який є у вільному доступі. На основі цих даних генеруються персональні текстові описи про клієнта і його вимоги до бажаного партнера, використовуючи промпти та ChatGPT [20]. Для порівняння анкетних даних та обрахунку відсотку сумісності використовується архітектура сіамської нейронної мережі.

Отже, на першому етапі методу відбувається створення анкет клієнтів шлюбного агентства, використовуючи дані із датасету. Оскільки, обраний датасет містить лише персональні та категоріальні дані, такі як рівень освіти, список хобі чи знак зодіаку, було вирішено доповнити анкети згенерованими текстовими описами користувачів про себе. Також, кожна анкета буде доповнена інформацією про вимоги до бажаного партнера. Ця інформація буде обрана випадковим чином із числа доступних даних, що зберігатимуться у базі даних. На основі сформованих вимог, також, буде згенеровано текстовий опис бажаного партнера.

Оскільки, згенерований текст може не відповідати очікуванням клієнта, перед проведенням пошуку партнера, користувачеві веб застосунку буде доступна можливість обрати підходящі текстові описи з числа згенерованих чи відредагувати їх, або згенерувати нові текстові описи. Це допоможе уникнути аналізування погано згенерованих текстів, які можуть погіршувати результати пошуку бажаних партнерів.

Для проведення інтелектуального аналізу анкетних даних необхідно провести попередню обробку даних, щоб уникнути отримання некоректних результатів через наявність помилкових вхідних даних. Тому, всі підготовчі кроки відбуватимуться на другому етапі. Спершу, необхідно провести перевірку на наявність пустих полів в анкетах клієнтів. При відсутності даних їх буде додано випадковим чином із числа доступних варіантів вибору. Наприклад, якщо в користувача відсутні дані про його вік, йому буде присвоєно випадкове число в діапазоні від 18 до 122 років. За нижню межу допустимого віку прийнято вік досягнення повноліття [21], а за верхню – вік найстаршої людини в світі [22].

Оскільки, крім персональних та категоріальних даних у анкетах присутні текстові описи за якими будуть обраховуватись відсотки подібності, виникає необхідність провести попередню обробку текстів перед початком аналізу. Для цього буде застосовано нормалізацію текстового вмісту. Під нормалізацією тексту мається на увазі приведення текстового вмісту до єдиної форми, яка використовуватиметься для подальшого аналізу. Зазвичай, це включає переведення тексту в нижній регістр [23]. Це допомагає уникнути ситуації, коли одне і те саме слово сприймається програмою по різному, через різні варіанти написання. Також, даний крок включатиме видалення зайвих символів та знаків із тексту, адже в даному контексті вони не несуть жодного семантичного навантаження.

Далі, текстові описи необхідно векторизувати, адже векторизація тексту – це основний процес, який необхідний для обробки природної мови. В

загальному, векторизація – це представлення текстового вмісту у числовий формат, зрозумілий для алгоритмів [24]. Саме на основі векторизованого тексту, сучасні системи для обробки природної мови здатні вирішувати безліч завдань, починаючи від пошуку ключових слів і закінчуючи розумінням ідей із тексту.

На останньому третьому етапі методу відбувається обрахунок відсотків сумісності клієнтів шлюбного агентства за інтелектуальним аналізом анкетних даних. Для початку, формується вибірка потенційних партнерів, які відповідають зазначеним вимогам. Вибірка формується шляхом порівняння вимог до партнера із анкетними даними клієнтів зареєстрованих у базі даних. Для цього необхідно створити певні правила, за якими буде відбуватись фільтрація. Наприклад, якщо клієнт шукає партнера віком від 20 до 25 років то в вибірку потраплять лише ті люди, чий вік задовольнятиме вказану вимогу.

Далі, відбувається обрахунок оцінок подібності текстових описів клієнтів, використовуючи архітектуру сіамської нейронної мережі. Для цього відбувається перехресне порівняння текстових описів для кожної потенційної пари. Тобто, опис бажаного партнера для цільового клієнта порівнюється із текстовими розповідями потенційних партнерів про себе для отримання відсотку схожості текстових описів. Потім, відбувається обернене порівняння, де розповідь про себе від цільового клієнта порівнюється із описами бажаних партнерів від потенційних партнерів із вибірки. Це необхідно для того, щоб враховувати сумісність відносно вимог обох партнерів.

Після того, як буде обчислено дві оцінки схожості текстових описів, обчислюватиметься загальна оцінка сумісності, шляхом обрахунку середнього арифметичного отриманих чисел. Таким чином, буде отримано загальний відсоток сумісності для кожної аналізованої пари анкет, який можна відсортувати та відобразити користувачеві у вигляді переліку із п'яти партнерів із найбільшим відсотком сумісності.

Отже, було розроблено метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, який обчислює загальний відсоток сумісності пар, враховуючи вимоги до бажаних партнерів від обох сторін.

Перелік посилань

1. Потапчук Є. М., Балуюка А. І. Чинники міжособистісної сумісності майбутніх шлюбних партнерів. Причорноморський науково-дослідний інститут економіки та інновацій. Одеса, 2021. С167–171.
2. Потапчук Н., Чорноморець Д. Життєдіяльність сімей військовослужбовців в умовах війни як психологічна проблема. *Psychology Travelogs*, 2024. С 241–250.
3. Мазурець О.В., Віт Р.В. Дослідження ефективності методу виявлення цільових об'єктів предметної області. Інформаційні технології і автоматизація. Матеріали XVII міжнародної науково-практичної конференції. 31 жовтня – 1 листопада 2024 р. Одеса, ОНТУ. 2024. С.650-653.
4. Mazurets O., Tymofiiiev I., Dydo R. Approach for Using Neural Network BERT-GPT2 Dual Transformer Architecture for Detecting Persons Depressive State. *Ricerche scientifiche e metodi*

- della loro realizzazione: esperienza mondiale e realtà domestiche. Raccolta di articoli scientifici con gli atti della VI Conferenza scientifica e pratica internazionale. 15 novembre, 2024. Bologna, Repubblica Italiana. 2024. Pp. 147-151.
5. Blazhuk V., Mazurets O., Zalutska O. An Approach to Using the mBERT Deep Learning Neural Network Model for Identifying Emotional Components and Communication Intentions. The Impact of Scientific Research on the Development of the Modern World. Proceedings of the XLIV International scientific and practical conference. October 23-25, 2024. Dubrovnik, Croatia. 2024. Pp. 79-84.
 6. Мазурець О.В., Молчанова М.О., Кліменко В.І., Собко О.В., Супрун П.К. Даталогічна модель бази даних для виявлення гендерної приналежності за SVM-аналізом дописів інтернет-мереж з використанням об'єктно-орієнтованого проектування. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2024. №3, Т.2 (337). С. 197-204.
 7. Mazurets O., Molchanova M., Klimenko V., Prosvitliuk M Practice Implementation of Neural Network Model BART-Large-CNN for Text Annotation. Prospects of Scientific Research in the Conditions of the Modern World. Proceedings of XXVII International scientific and practical conference. June 12-14, 2024. Rotterdam, Netherlands. 2024. Pp. 97-102.
 8. Sobko O., Mazurets O., Didur V., Chervonchuk I. Recurrent Neural Network Model Architecture for Detecting a Tendency to Atypical Behavior Of Individuals by Text Posts. Theoretical and Practical Aspects of Modern Research. Proceedings of XXVI International scientific and practical conference. June 5-7, 2024. International Scientific Unity. Ottawa, Canada. 2024. Pp. 113-117.
 9. Mazurets O.V., Sobko O.V., Molchanova M.O., Zalutska O.O., Yurchak A.V. Practical Implementation of Neural Network Method for Stress Features Detection by Social Internet Networks Posts. Global Science: Prospects and Innovations. Proceedings of the II International Scientific and Theoretical Conference «Scientific Review of the Actual Events, Achievements and Problems». May 31, 2024. Berlin, Federal Republic of Germany: International Center of Scientific Research. 2024. Pp. 160-167.
 10. Molchanova M., Mazurets O., Sobko O., Boiarchuk I. Object-Oriented Approach for Ethnic Enmity Detection in Text Messages by NLP. Proceedings of XXI International Scientific and Practical Conference «Scientific Achievements and Innovations as a Way to Success». May 1-3, 2024. Vilnius, Lithuania. 2024. Pp. 73-77.
 11. Молчанова М.О., Мазурець О.В., Собко О.В., Кліменко В.І., Андрощук В.І. Метод нейромережевого виявлення кібербулінгу з використанням хмарних сервісів та об'єктно-орієнтованої моделі. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2024. №2 (333). С. 200-206.
 12. Залуцька О.О., Молчанова М.О., Біг Р.В., Мазурець О.В. Конфігурування нейронної мережі для класифікації емоційної тональності текстової інформації за показниками семантичної зв'язності. Збірник наукових праць за матеріалами XV Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2023». Хмельницький, 2023. с. 102-107.

Довідка: ВХНУ ТН 4-12/2024

Видання: Herald of Khmelnytskyi National University. Technical Sciences (Вісник Хмельницького національного університету. Технічні науки)

Категорія фаховості видання: затверджено як наукове фахове видання України, у якому можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора наук, кандидата наук та ступеня доктора філософії, категорії «Б» (наказ МОН №1643 від 28.12.2019, наказ МОН №409 від 17.03.2020).

Напрямок – технічні науки за спеціальностями – 101, 121, 122, 123, 124, 125, 141, 151, 161, 172, 181, 182 (28.12.2019), спеціальності – 131, 132, 133 (17.03.2020).

Назва статті: РЕЛЯЦІЙНА ПЕРСОНІФІКОВАНА ДАТАЛОГІЧНА МОДЕЛЬ ДАНИХ ДЛЯ ВИЗНАЧЕННЯ СУМІСНОСТІ АНКЕТНИХ ДАНИХ НЕЙРОМЕРЕЖЕВИМИ ЗАСОБАМИ

Автори: Жук Д.І., Мазурець О.В., Залуцька О.О., Денисенко Б.О.
Хмельницький національний університет

Номер, у який прийнято статтю: №1 за 2025 рік.

04.12.2024

Начальника відділу
інтелектуальної власності та трансферу технологій _____ Ю.В.Кравчик



[Handwritten signature]
[Handwritten signature]

УДК 004.8

ЖУК ДМИТРО

Хмельницький національний університет

e-mail: 51423dima@gmail.com**МАЗУРЕЦЬ ОЛЕКСАНДР**

Хмельницький національний університет

<https://orcid.org/0000-0002-8900-0650>e-mail: exe_chong@gmail.com**ЗАЛУЦЬКА ОЛЬГА**

Хмельницький національний університет

<https://orcid.org/0000-0003-1242-3548>e-mail: zalutsk.olha@gmail.com**ДЕНИСЕНКО БОГДАН**

Хмельницький національний університет

e-mail: bohbandenyzenko@ukr.net

РЕЛЯЦІЙНА ПЕРСОНІФІКОВАНА ДАТАЛОГІЧНА МОДЕЛЬ ДАНИХ ДЛЯ ВИЗНАЧЕННЯ СУМІСНОСТІ АНКЕТНИХ ДАНИХ НЕЙРОМЕРЕЖЕВИМИ ЗАСОБАМИ

У статті запропоновано підхід до визначення сумісності анкетних даних нейромережевими засобами. Підхід базується на двох методах, а саме порівнянні анкетних даних клієнтів і кандидатів через фреймворк модель подання та текстовий аналіз описів.

Реалізація підходу була здійснена через веб-сайт шлюбного агентства, де для збереження анкет клієнтів та допоміжної інформації розроблено персоналізовану реляційну даталогічну модель, що включає 23 взаємопов'язані таблиці. Основними елементами цієї моделі є дані про клієнтів та вимоги до потенційного партнера, що використовуються для пошуку сумісних кандидатів.

Результати дослідження показали, що зі збільшенням кількості кандидатів у результуючих множинах обох методів підходу, кількість сумісних пар зростає. Паралельне використання методу порівняння анкет за фреймворк моделлю та нейромережевого аналізу текстових описів дозволило значно підвищити ефективність визначення сумісності анкетних даних. Розроблений підхід забезпечує високу точність у визначенні сумісних пар клієнтів, покращуючи процес пошуку за допомогою технологій штучного інтелекту.

Ключові слова: база даних, анкети користувачів, сумісність, нейромережеві засоби

ZHUK DMYTRO, MAZURETS OLEKSANDR, ZALUTSKA OLHA, DENYSENKO BOHDAN
Khmelnyskyi National University

RELATIONAL PERSONALIZED DATALOGICAL DATA MODEL FOR DETERMINING THE COMPATIBILITY OF SURVEY DATA USING NEURAL NETWORK MEANS

The article presents an approach to determining the compatibility of questionnaire data using neural network-based methods. The method is based on comparing the questionnaire data of clients and candidates through a frame-based representation model and textual analysis of descriptions.

The implementation of the approach was carried out through a marriage agency website, where a personalized relational data model was developed for storing client questionnaires and auxiliary information. This model consists of 23 interconnected tables. The main components of this model include data about clients and their requirements for a potential partner, which are used to find compatible candidates.

The results of the study showed that as the number of candidates in the database increases, the number of compatible pairs also grows. The parallel use of the frame-based model for comparing questionnaires and neural

network-based analysis of textual descriptions significantly increased the effectiveness of determining the compatibility of questionnaire data. The developed approach ensures high accuracy in identifying compatible client pairs, enhancing the search process through the use of artificial intelligence technologies.

Furthermore, the application of this method allows the system to process large volumes of data efficiently, providing more reliable and accurate compatibility assessments. By combining both frame-based comparison and neural network analysis, the system can account for not only the structural features of the data but also the semantic content of the descriptions, which is essential for assessing compatibility in a deeper context. This dual approach ensures a more holistic understanding of potential matches, improving the user experience and outcomes in the matchmaking process. The proposed method opens new possibilities for personalized matchmaking systems, offering a scalable solution for web-based applications. Through the integration of AI technologies, this approach could be adapted and expanded for use in various domains where compatibility analysis is needed, making it a versatile tool for future applications in other fields beyond matchmaking.

Keywords: database, user questionnaires, compatibility, neural network tools

Аналіз предметної області

Шлюбні агентства відіграють важливу роль у вирішенні соціально значущих задач, пов'язаних із побудовою міжособистісних стосунків. Їх актуальність зумовлена сучасними соціальними процесами, зокрема зростаючою зайнятістю населення, підвищеними вимогами до потенційних партнерів і обмеженнями, викликаними географічними та демографічними факторами [1]. Для підвищення ефективності роботи таких агентств необхідна інтеграція сучасних інформаційних технологій, здатних обробляти великі обсяги анкетних даних і забезпечувати високу точність прогнозування сумісності між клієнтами.

Для визначення сумісності анкетних даних доцільно використовувати поєднання кількох підходів, які дозволяють не лише оптимізувати процес інтелектуального аналізу анкетних даних, а й забезпечити більш точні результати щодо сумісності клієнтів. Зокрема, статистичний аналіз числових і категоріальних даних є ефективним на етапі первинної фільтрації анкет, що дозволяє сформувати групу потенційних партнерів, які відповідають базовим критеріям. Подальша деталізація можлива за рахунок використання систем штучного інтелекту, зокрема моделей глибокого навчання, таких як BERT, для аналізу текстових описів [2].

Реалізація таких методів значною мірою залежить від ефективної організації баз даних. У контексті шлюбних агентств доцільно використовувати реляційні бази даних, які забезпечують структуроване зберігання анкетних даних із чіткою фіксацією взаємозв'язків між полями [3, 4]. Наприклад, таблиці можуть містити демографічну інформацію, психологічні профілі, уподобання клієнтів, а також результати проміжних розрахунків, зокрема рівня сумісності. Такі моделі дозволяють формалізувати процес представлення анкетних даних шляхом створення зв'язків між параметрами, що визначають персональні характеристики та вподобання клієнтів [4]. Використання цих моделей у поєднанні з нейромережевими засобами забезпечує можливість автоматизованого визначення рівня сумісності [5].

Таким чином, інтеграція статистичного аналізу, моделей глибокого навчання та ефективно структурованих баз даних створює комплексну систему, здатну забезпечити високоточне визначення сумісності анкет клієнтів шлюбних агентств.

Сучасний стан досліджень

Для визначення сумісності анкетних даних нейромережевими засобами необхідно використовувати сучасні методи інтелектуального аналізу даних (ІАД), нейромережеві засоби у поєднанні з ефективно організованими базами даних. Організація баз даних є також важливою для забезпечення ефективної роботи системи аналізу. Реляційні бази даних дозволяють зберігати структуровану інформацію, таку як демографічні характеристики, уподобання клієнтів і результати проміжних обчислень, забезпечуючи можливість швидкого доступу до цих даних для подальшої обробки [6]. Далі наведено аналіз публікацій присвячених аналізу та визначенню сумісності анкетних даних нейромережевими засобами.

У дослідженні [16] запропоновано метод нечіткого пошуку з використанням таблиці подібності

символів, що сприяє підвищенню швидкості та точності пошуку релевантних документів за заданою фразою. Цей метод враховує орфографічні помилки, синоніми, скорочення та альтернативні форми слів, використовуючи таблицю подібності символів для визначення різних рівнів схожості.

У дослідженні [17] розглянуто використання сіамських нейронних мереж (SNN) для аналізу схожості між об'єктами. Цей підхід дозволяє вимірювати схожість, порівнюючи вектори ознак об'єктів. Хоча він використовується для рекомендацій на основі уподобань користувачів, його можна застосувати для створення списку сумісних партнерів, порівнюючи анкети клієнтів.

Дослідження [18] вивчає використання нейронних мереж для визначення схожості текстів українською мовою, розглядаючи моделі LSTM та CNN. Порівняння результатів включало аналіз текстів з лематизацією та без неї. Експеримент показав, що лематизація покращує точність аналізу: для лематизованих текстів точність складала 64%, для нелематизованих – 61%.

Отже, проведено аналіз наукових публікацій з аналізу та визначення сумісності анкетних даних користувачів. Розглянуто підходи до порівняння текстових даних, зокрема методи та системи штучного інтелекту, включаючи моделі глибокого навчання, які демонструють високу точність у роботі з текстами. Зважаючи на актуальність даної теми, а також недостатньо високі показники точності визначення відсотку сумісності анкет, виникає потреба у подальших дослідженнях.

Мета дослідження полягає в розробці підходу для визначення сумісності анкетних даних нейромережевими засобами, проєктуванні реляційної персоніфікованої датовісочної моделі даних та розробка програмної реалізації веб-застосунку для визначення сумісності анкетних даних нейромережевими засобами.

Основна частина

Підхід для визначення сумісності анкетних даних нейромережевими засобами передбачає одночасне використання двох методів (рисунок 1).

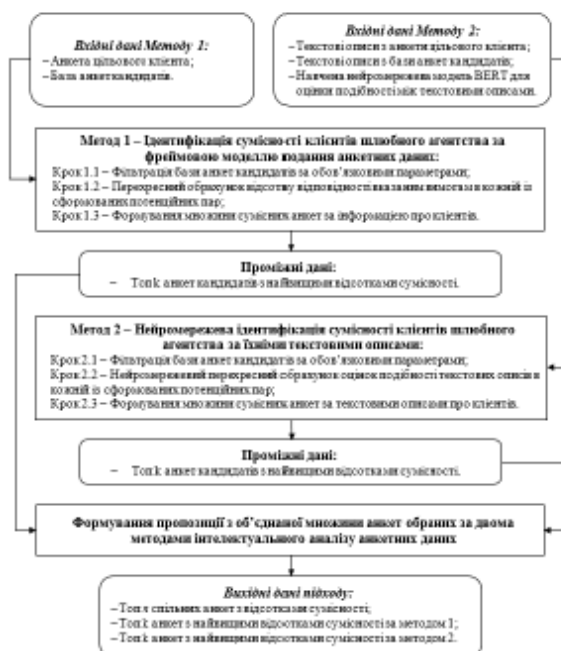


Рис. 1. Схема підходу визначення сумісності анкет клієнтів шлюбних агентств

Перший метод передбачає ідентифікацію сумісності анкет клієнтів шлюбного агентства на основі

фреймової моделі подання анкетних даних, тоді як другий метод здійснюється за допомогою нейромережевої ідентифікації сумісності на основі текстових описів клієнтів. У першому методі вхідними даними є анкета цільового клієнта шлюбного агентства та база анкет кандидатів. Анкета цільового клієнта є документом, на основі якого здійснюється підбір партнерів з високою сумісністю. База анкет кандидатів представляє собою клієнтську базу шлюбного агентства, що містить перелік зареєстрованих клієнтів.

Кожна анкета клієнта містить вимоги до бажаного партнера, тому на першому етапі обох методів відбувається фільтрація кандидатів за шми обов'язковими параметрами. Це дозволяє зменшити обсяг даних та відсіяти кандидатів, які не відповідають вимогам, наприклад, якщо клієнт шукає партнера протилежної статі.

Далі здійснюється перехресний обрахунок відсотка відповідності вимогам у кожній потенційній парі. Кожен кандидат, який пройшов фільтрацію, перетворюється на можливого партнера для цільового клієнта. Формується кілька потенційних пар, для яких обчислюється відсоток відповідності як для кандидата, так і для цільового клієнта. Отримане середнє значення є загальним відсотком сумісності пари. Якщо він перевищує 50%, пара вважається сумісною.

Використовуючи цей метод, формується множина сумісних анкет, з яких обираються найкращі для подальших досліджень. Одночасно з першим методом, реалізується нейромережева ідентифікація сумісності за текстовими описами. Вхідними даними є текстові описи цільового клієнта та кандидатів, а також навчена модель BERT для оцінки подібності текстів. Текстові описи клієнтів містять розповідь про себе та бажаного партнера. Аналогічні описи є і в кандидатів. Спочатку здійснюється фільтрація за обов'язковими параметрами, після чого обчислюються оцінки подібності текстів між кожною парою за допомогою BERT. Для кожної пари обчислюються дві оцінки подібності, і на основі середнього арифметичного визначається загальний відсоток сумісності.

Після виконання обох методів буде сформовано два списки анкет з найвищими відсотками сумісності (наприклад, топ 10 за кожним методом). З цих списків створюється об'єднана пропозиція, що включає лише анкети, які потрапили до результатів обох методів. Таким чином, вхідними даними запропонованого підходу є: топ n спільних анкет із відсотками сумісності; топ k анкет з найвищими відсотками сумісності за методом ідентифікації сумісності клієнтів шлюбного агентства на основі фреймової моделі подання анкетних даних; топ k анкет з найвищими відсотками сумісності за методом нейромережевої ідентифікації сумісності клієнтів за їхніми текстовими описами.

Структура інформаційної системи для визначення сумісності клієнтів шлюбних агентств складається з трьох підсистем і бази даних (рисунок 2). Перша підсистема відповідає за функціональність, пов'язану з поповненням клієнтської бази шлюбного агентства. Друга підсистема є прикладною та реалізується у вигляді веб-сайту шлюбного агентства, що використовує запропонований підхід для визначення сумісності анкет. Третя підсистема реалізує метод визначення сумісності клієнтів через інтелектуальний аналіз анкетних даних. База даних слугує для зберігання клієнтських анкет та іншої інформації, необхідної для роботи веб-сайту.



Рис. 2. Структура інформаційної системи

Була спроектована реляційна персоніфікована даталогічна модель даних для визначення сумісності анкетних даних нейромережевими засобами (Рисунок 3.2). База даних включає двадцять три таблиці: Users, Profiles, Ideal_profiles, Speaks, Languages, Signs, Religions, Pets, Offspring, Ethnicities, Educations, Smokes, Drugs, Drinks, Diets, Body_types, Orientations, Sex, Statuses, Jobs, Cities, Hobbies, Hobbies_lists. Далі описано основні таблиці даталогічна модель даних для визначення сумісності анкетних даних нейромережевими засобами.

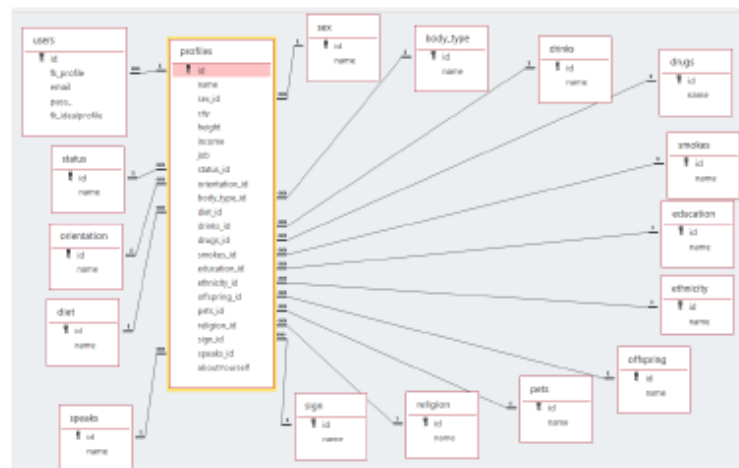


Рис. 3. Реляційна персоніфікована даталогічна модель даних для визначення сумісності анкетних даних нейромережевими засобами

Таблиця «Users» призначена для зберігання інформації про клієнтів шлюбного агентства. Вона включає

такі поля: унікальний ідентифікатор клієнта, особисті дані, електронну пошту, пароль, профіль користувача та вимоги до ідеального партнера. Атрибути таблиці: унікальний ідентифікатор клієнта «ID» (INT), ім'я клієнта «Name» (TEXT), прізвище клієнта «Surname» (TEXT), електронна пошта «Email» (TEXT), пароль клієнта «Password» (TEXT), профіль користувача «FK_profile» (INT) та вимоги до ідеального партнера «FK_idealprofile» (INT).

Таблиця «Profiles» призначена для зберігання детальної інформації про клієнтів шлюбного агентства. Вона включає такі поля: унікальний ідентифікатор анкети, особисті характеристики, такі як вік, зріст, вага, дохід, професія, сімейний статус, орієнтація, тип статури, освіта та етнічна приналежність. Додатково вказуються особливі звички та уподобання клієнтів, включно з дієтою, вживанням алкоголю чи наркотичних речовин, ставленням до куріння, наявністю дітей або домашніх тварин, а також релігійні погляди й знак зодіаку. Атрибути таблиці: унікальний ідентифікатор анкети «ID» (INT), вік клієнта «Age» (INT), стать «FK_sex» (INT), місто проживання «FK_City» (INT), зріст «Height» (INT), вага «Weight» (INT), дохід «Income» (INT), професія «FK_job» (INT), сімейний статус «FK_status» (INT), орієнтація «FK_orientation» (INT), тип статури «FK_body_type» (INT), дієта «FK_diet» (INT), вживання алкоголю «FK_drinks» (INT), вживання наркотиків «FK_drugs» (INT), ставлення до куріння «FK_smokes» (INT), освіта «FK_education» (INT), етнічна приналежність «FK_ethnicity» (INT), наявність дітей «FK_offspring» (INT), наявність домашніх тварин «FK_pets» (INT), релігія «FK_religion» (INT), знак зодіаку «FK_sign» (INT) та текстовий опис клієнта «About_yourself» (TEXT).

Допоміжні таблиці включають «sex» (стать), «status» (статус), «orientation» (орієнтація), «body_type» (тип статури), «diet» (дієта), «drinks» (звички щодо вживання алкоголю), «drugs» (вживання наркотиків), «smokes» (звички куріння), «education» (рівень освіти), «ethnicity» (етнічна приналежність), «offspring» (ставлення до дітей), «pets» (наявність домашніх тварин), «religion» (релігійні погляди), «sign» (знак зодіаку) та «speak» (мови спілкування). Кожна з цих таблиць має типову структуру, що включає унікальний ідентифікатор (id, тип INT) та атрибут значення (name, тип VARCHAR/TEXT). Загальна структура забезпечує зв'язки за типом «один-до-багатьох» між головною таблицею профілів та допоміжними таблицями, що дозволяє ефективно зберігати й аналізувати персоналізовані анкетні дані.

Для практичного застосування підходу підходу визначення сумісності анкет клієнтів шлюбних агентств було реалізовано інформаційну систему у форматі веб-сайту шлюбного агентства (рисунок 4).



Рис. 4. Відображення результатів визначення сумісності анкетних даних нейронмережевими засобами

Створений веб-сайт шлюбного агентства дозволяє користувачам пройти реєстрацію та заповнити анкету, вказавши особисту інформацію. На основі введених даних та критеріїв щодо бажаного партнера клієнти можуть здійснювати пошук потенційних кандидатів за допомогою підходу визначення сумісності анкетних даних нейромережевими засобами.

Експерименти та дискусія

Аналіз ефективності підходу визначення сумісності анкетних даних нейромережевими засобами, полягає у вивченні залежності кількості сумісних партнерів у об'єднаній множині від кількості партнерів у сформованих множинах. Це досягається шляхом зміни параметра k , що визначає кількість партнерів з найвищим відсотком сумісності, які відбираються до множин за результатами обох методів. Зі зміною параметра k змінюється й кількість партнерів n , що потрапляють до фінальної об'єднаної множини.

Для перевірки ефективності запропонованого підходу було проведено чотири варіанти пошуку сумісних партнерів для одного клієнта. У кожному варіанті застосовували різні значення параметра k (таблиця 1). Важливо зазначити, що значення параметра k завжди однакове для обох методів.

Таблиця 1

Вплив параметру k на кількість анкет в об'єднаній множині

Кількість анкет, які будуть відібрані в множині результатів (k)	Кількість анкет, які потрапили в об'єднану множину (n)
5	0
10	2
15	6
20	9

Таким чином, результати проведеного експерименту показують, що зі збільшенням розміру множин сумісних партнерів зростає кількість анкет, які потрапляють до фінальної об'єднаної множини. Водночас при низьких значеннях параметра k існує ймовірність, що жодна анкета не буде включена до об'єднаної множини.

Експеримент було проведено повторно для 10 різних анкет. Для кожного значення параметра k визначали середнє значення параметра n . Узагальнені результати дослідження представлені у вигляді діаграми на рисунку 5.

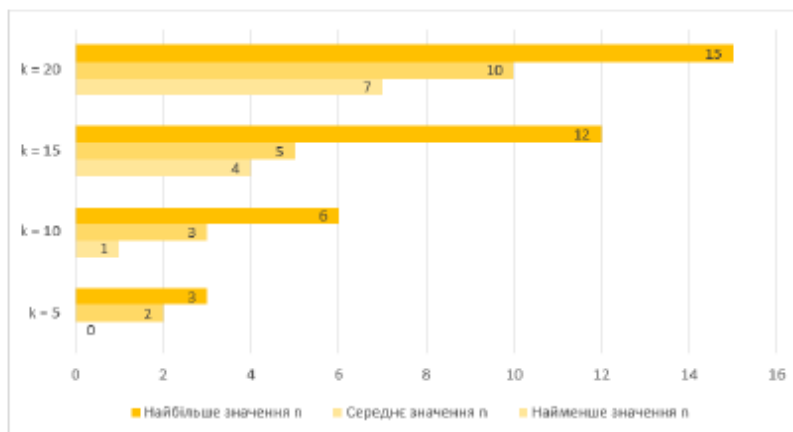


Рис. 6. Діаграма залежності кількості сумісних партнерів в об'єднаній множині від кількості сумісних партнерів в результативних множинах обох методів

Результати проведеного дослідження підтверджують, що зі збільшенням кількості сумісних анкет, які потрапляють до результативних множин обраних методів, зростає й кількість анкет у фінальній об'єднаній

множині. Це, у свою чергу, надає клієнту шлюбного агентства можливість переглянути ширший перелік кандидатів та обрати серед них партнера з найбільш відповідними якостями.

Висновки

Реалізовано підхід до визначення сумісності анкетних даних нейромережевими засобами, що дозволяє на основі вхідних даних у формі анкети клієнта та бази анкет кандидатів розраховувати відсоток сумісності. Процес ґрунтується на порівнянні анкетних даних через фреймову модель подання та текстовий аналіз описів. Практичне використання методу впроваджено через веб-сайт шлюбного агентства.

Для збереження анкет клієнтів та допоміжної інформації було розроблено реляційну персоніфіковану даталогічну модель даних, що складається з 23 взаємопов'язаних таблиць. Основні таблиці містять дані про клієнта та перелік вимог до потенційного партнера. Саме ці дані є основою для пошуку сумісних кандидатів за запропонованим методом аналізу анкетних даних.

Проведене дослідження ефективності роботи запропонованого підходу показало, що зі збільшенням кількості кандидатів у результуючих множинах обох методів, кількість кандидатів в об'єднаній множині також зростає.

Таким чином наведений підхід визначення сумісності анкетних даних нейромережевими засобами, спроектована реляційна персоніфікована даталогічна модель даних та програмна реалізація веб-застосунок для визначення сумісності анкетних даних нейромережевими засобами дозволили якісно покращити визначення сумісності анкетних даних. Розроблений підхід працює шляхом паралельного використання методу порівняння анкет за фреймовою моделлю та нейромережевого визначення оцінок схожості текстових описів клієнтів, було отримано відсотки сумісності між клієнтськими парами.

Література

1. Zhuk D. Datalogic Model for Determining the Clients Compatibility Based on Questionnaires Data Analysis by Artificial Intelligence Means / D. Zhuk, O. Mazurets, O. Sobko, V. Klimenko // *Innovative Solutions in Science: Balancing Theory and Practice. Proceedings of the 1st International scientific and practical conference. October 28-30, 2024. San Francisco, USA. 2024. Pp. 80-85.*
2. Жук Д.І., Мазурець О.В., Кадиська В.Д., Тищенко О.О. Підхід до визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних / Д.І. Жук, О.В. Мазурець, В.Д. Кадиська, О.О. Тищенко // *Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. – С. 208-214.*
3. Mazurets O.V. Datalogic projecting for applied adaptive semantic testing / O.V. Mazurets // *Heritage of European Science: international collective monograph. – Karlsruhe, Germany: Monographic series «European Science». – 2024. – Book 27. Part 3. Chapter 8. – P. 147-156.*
4. Mazurets O. Relation datalogic model for determining the diagnosis based on intellectual NLP-analysis of symptom description / O. Mazurets, O. Sobko, V. Klimenko, Y. Kozenko // *Innovative Development: Synthesis of Scientific Approaches in Various Fields of Research: proceedings of XV International Scientific and Practical Conference. – Tallinn, Estonia, 20-22 березня 2024. – 2024. – P. 61-66.*
5. Mazurets O. Datalogic model for image recognition by convolutional neural network using cloud services / O. Mazurets, M. Molchanova, V. Klimenko, D. Klopotivskiy // *Modern Scientific Research: Theoretical and Practical Aspects: proceedings of XXII International Scientific and Practical Conference. – Oslo, Norway, 8-10 травня 2024. – 2024. – Pp. 64-68.*
6. Мазурець О.В. Даталогічна модель бази даних для виявлення гендерної приналежності за SVM-аналізом дописів інтернет-мереж з використанням об'єктно-орієнтованого проєктування / О.В. Мазурець, М.О. Молчанова, В.І. Кліменко, О.В. Собко, П.К. Супрун // *Вісник Хмельницького національного університету. Серія: Технічні науки. – 2024. – №3, Т.2 (337). – С. 197-204.*

7. Клещ К.О. Підвищення ефективності алгоритмів нечіткого пошуку з використанням таблиці подібності символів / К.О. Клещ. – Київ, 2024. – 137 с.

8. Брус А. Використання нейронних мереж для визначення схожості текстів українською мовою / А. Брус. – Міністерство освіти і науки України, національний університет «Києво-Могилянська академія». – Київ, 2020. – 35 с.

References

1. Zhuk D. Datalogic Model for Determining the Clients Compatibility Based on Questionnaires Data Analysis by Artificial Intelligence Means / D. Zhuk, O. Mazurets, O. Sobko, V. Klimenko // *Innovative Solutions in Science: Balancing Theory and Practice. Proceedings of the 1st International scientific and practical conference. October 28-30, 2024. San Francisco, USA. 2024. Pp. 80-85.*

2. Zhuk D.I., Mazurets O.V., Kadynska V.D., Tyshchenko O.O. Pidkhdid do vyznachennia sumisnosti klientiv shliubnykh ahentstv za intelektualnym analizom anketnykh danykh / D.I. Zhuk, O.V. Mazurets, V.D. Kadynska, O.O. Tyshchenko // *Zbirnyk naukovykh prats za materialamy XVI Vseukrainskoi naukovo-praktychnoi konferentsii «Aktualni problemy kompiuternykh nauk APKN-2024». 15-16 lystopada 2024. Khmelnytskyi, 2024. – S. 208-214.*

3. Mazurets O.V. Datalogic projecting for applied adaptive semantic testing / O.V. Mazurets // *Heritage of European Science: international collective monograph. – Karlsruhe, Germany: Monographic series «European Science». – 2024. – Book 27. Part 3. Chapter 8. – P. 147-156.*

4. Mazurets O. Relation datalogic model for determining the diagnosis based on intellectual NLP-analysis of symptom description / O. Mazurets, O. Sobko, V. Klimenko, Y. Kozenko // *Innovative Development: Synthesis of Scientific Approaches in Various Fields of Research: proceedings of XV International Scientific and Practical Conference. – Tallinn, Estonia, 20-22 bereznia 2024. – 2024. – P. 61-66.*

5. Mazurets O. Datalogic model for image recognition by convolutional neural network using cloud services / O. Mazurets, M. Molchanova, V. Klimenko, D. Klopotivskyi // *Modern Scientific Research: Theoretical and Practical Aspects: proceedings of XXII International Scientific and Practical Conference. – Oslo, Norway, 8-10 travnia 2024. – 2024. – Pp. 64-68.*

6. Mazurets O.V. Datalogichna model bazy danykh dla vyiavlennia hendernoï pry налеzhnosti za SVM-analizom dopysiv internet-merezh z vykorystanniam obiektno-orientovanoho proektuvannia / O.V. Mazurets, M.O. Molchanova, V.I. Klimenko, O.V. Sobko, P.K. Suprun // *Visnyk Khmelnytskoho natsionalnoho universytetu. Seria: Tekhnichni nauky. – 2024. – №3, T.2 (337). – S. 197-204.*

7. Kleshch K.O. Pidvyshchennia efektyvnosti alhorytmiv nechitkoho poshuku z vykorystanniam tablytsi podobnosti symboliv / K.O. Kleshch. – Kyiv, 2024. – 137 s.

8. Brus A. Vykorystannia neuronnykh merezh dla vyznachennia skhozhosti tekstiv ukrainskoïu movoïu / A. Brus. – Ministerstvo osvity i nauky Ukrainy, natsionalnyi universytet «Kyievo-Mohylianska akademiia». – Kyiv, 2020. – 35 s.

Додаток Е

Презентаційний матеріал

Кваліфікаційна робота магістра

МЕТОД ВИЗНАЧЕННЯ СУМІСНОСТІ КЛІЄНТІВ ШЛЮБНИХ АГЕНТСТВ ЗА ІНТЕЛЕКТУАЛЬНИМ АНАЛІЗОМ АНКЕТНИХ ДАНИХ

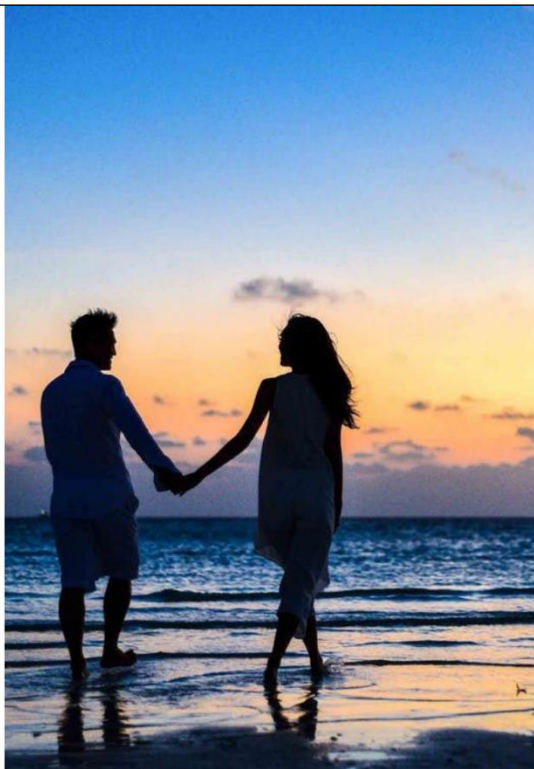
Виконав: студент 2 курсу, групи КНм-23-1 Жук Дмитро Іванович
Науковий керівник: доцент кафедри КН Мазурець Олександр Вікторович

МЕТА ТА ЗАВДАННЯ РОБОТИ

Метою роботи є покращення визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

Для досягнення поставленої мети необхідно вирішити такі задачі:

- провести аналіз предметної області, пов'язаної із діяльністю шлюбних агентств;
- розробити метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних;
- обрати датасет для формування клієнтської бази;
- створити набір промптів для автоматизованої генерації текстових описів клієнтів через ChatGPT;
- спроектувати структуру веб-застосунку для пошуку партнерів;
- спроектувати структуру даталогічної моделі бази даних веб-застосунку для пошуку партнерів;
- виконати програмну реалізацію веб-застосунку для пошуку партнерів;
- провести тестування розробленого веб-застосунку для пошуку партнерів;
- виконати дослідження ефективності запропонованого методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних використовуючи реалізований веб-застосунок.



ОСОБЛИВОСТІ ДІЯЛЬНОСТІ ШЛЮБНИХ АГЕНТСТВ

Шлюбні агентства – це спеціалізовані організації, які надають послуги пошуку партнерів для створення сімейних зв'язків. Основна мета шлюбних агентств – допомога клієнту в пошуку ідеального партнера з високим рівнем сумісності.

Актуальність існування шлюбних агентств зумовлене рядом соціальних проблем, пов'язаних із пошуком партнерів та створенням сім'ї. Основними проблемами сучасності, які заважають знайти підходящого партнера є: зайнятість і брак часу, підвищенні вимоги до партнера, географічне розташування.



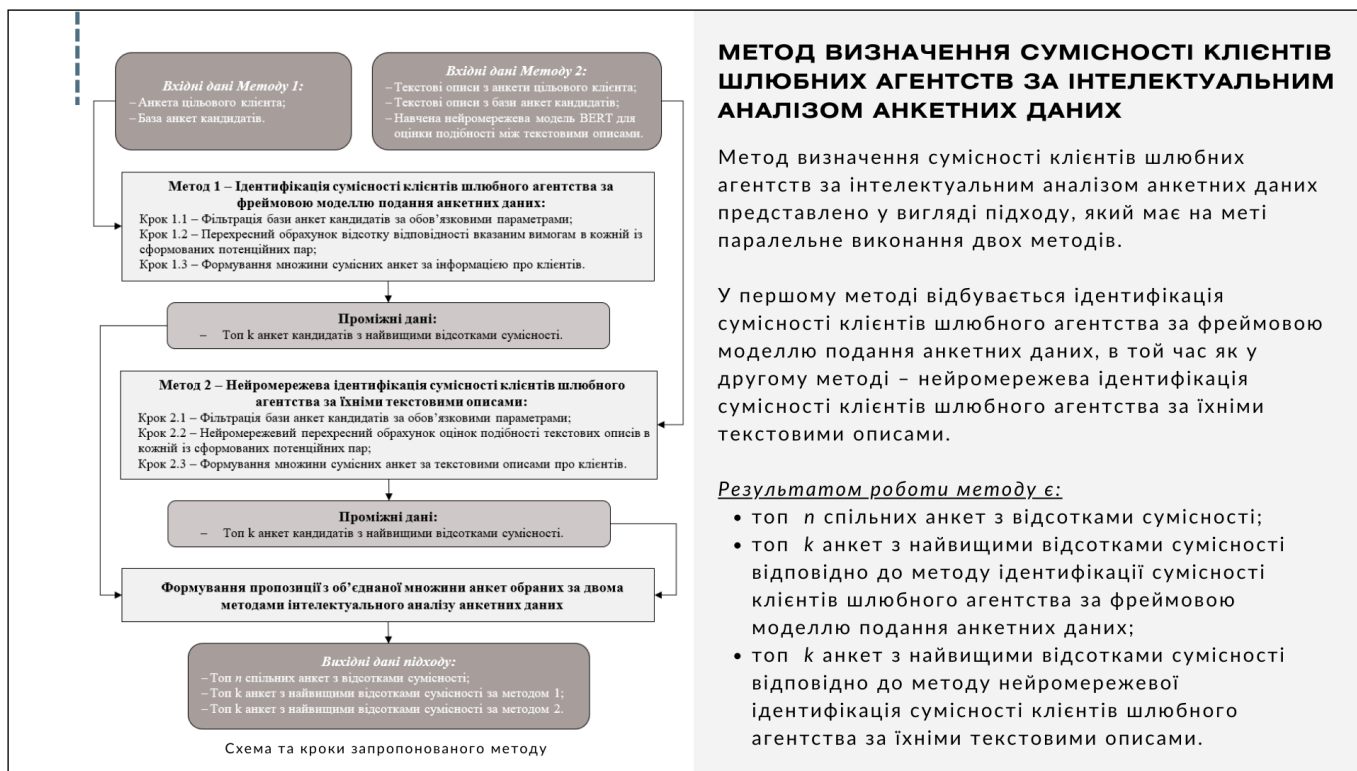
ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ

ВИЗНАЧЕННЯ

Інтелектуальний аналіз даних – це процес виявлення прихованих, раніше невідомих або неочевидних закономірностей, знань або моделей у великих об'ємах даних, використовуючи математичні, статистичні та засновані на алгоритмах штучного інтелекту методи.

ЗВ'ЯЗОК ІАД З ПРЕДМЕТНОЮ ОБЛАСТЮ

Інтелектуальний аналіз даних застосовано для визначення сумісності клієнтів шлюбних агенств шляхом порівняння їхніх анкетних даних, використовуючи як статистичні методи, так і системи штучного інтелекту.



МЕТОД ВИЗНАЧЕННЯ СУМІСНОСТІ КЛІЄНТІВ ШЛЮБНИХ АГЕНТСТВ ЗА ІНТЕЛЕКТУАЛЬНИМ АНАЛІЗОМ АНКЕТНИХ ДАНИХ

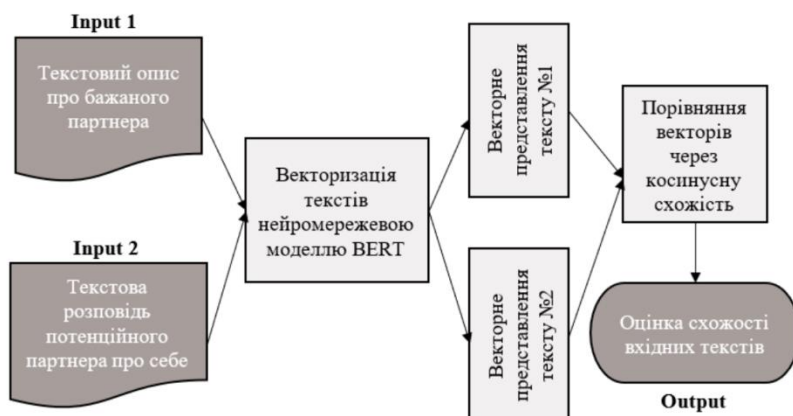
Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних представлено у вигляді підходу, який має на меті паралельне виконання двох методів.

У першому методі відбувається ідентифікація сумісності клієнтів шлюбного агентства за фреймовою моделлю подання анкетних даних, в той час як у другому методі – нейромережева ідентифікація сумісності клієнтів шлюбного агентства за їхніми текстовими описами.

Результатом роботи методу є:

- топ n спільних анкет з відсотками сумісності;
- топ k анкет з найвищими відсотками сумісності відповідно до методу ідентифікації сумісності клієнтів шлюбного агентства за фреймовою моделлю подання анкетних даних;
- топ k анкет з найвищими відсотками сумісності відповідно до методу нейромережевої ідентифікації сумісності клієнтів шлюбного агентства за їхніми текстовими описами.

СХЕМА МЕТОДУ ОБЧИСЛЕННЯ ОЦІНКИ СХОЖОСТІ ДЛЯ ОДНІЄЇ ПАРИ ТЕКСТОВИХ ОПИСІВ



Метод нейромережевої ідентифікації сумісності клієнтів шлюбного агентства за їхніми текстовими описами має на меті обчислення оцінки схожості текстових описів клієнтів на основі порівняння їхніх векторних представлень для чого використовується нейромережева модель BERT.

Обчислення оцінки схожості векторизованих текстових описів відбувається з використанням функції порівняння. Для цього, використовується косинусна схожість (Cosine Similarity), яка дозволяє обчислювати подібність між двома векторами обчислюючи косинус кута між ними

ОСОБЛИВОСТІ ФОРМУВАННЯ КЛІЄНТСЬКОЇ БАЗИ

Для забезпечення роботи методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних було сформовано базу анкет кандидатів, тобто клієнтську базу шлюбного агентства.

1

OKCUPID PROFILES

Датасет з інформацією про профілі користувачів платформи для знайомств "OkCupid".

2

FAKER

Спеціалізований програмний інструмент для генерації несправжніх даних, які відсутні у датасеті.

3

ПРОМПТИ

Текстові запити, які служать інструкціями для моделі ШІ, з метою генерації текстових описів за анкетними даними.



СХЕМА СТРУКТУРИ ІНФОРМАЦІЙНОЇ СИСТЕМИ ВИЗНАЧЕННЯ СУМІСНОСТІ КЛІЄНТІВ ШЛЮБНИХ АГЕНТСТВ



Структуру інформаційної системи визначення сумісності клієнтів шлюбних агентств поділено на три підсистеми та базу даних.

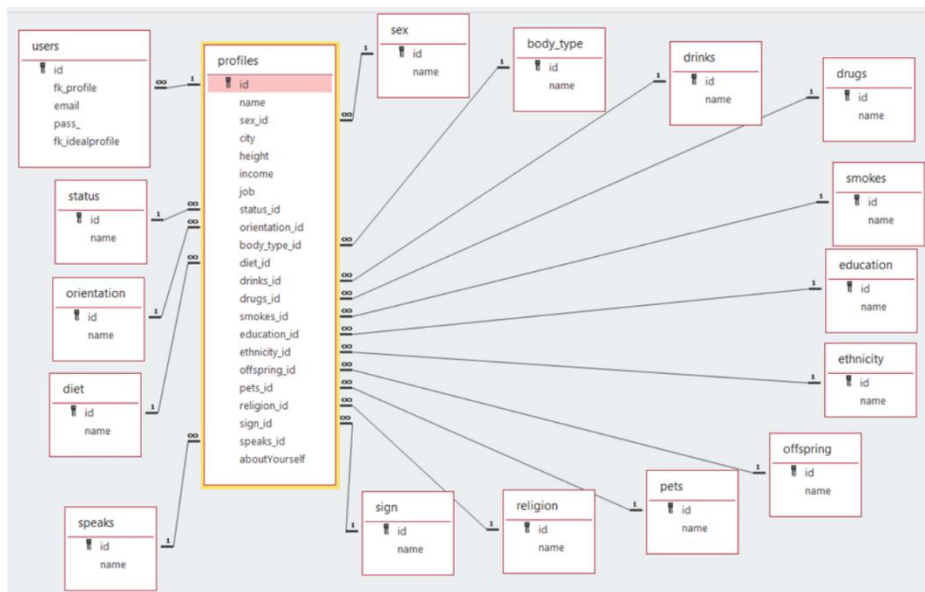
Перша підсистема необхідна для реалізації функціоналу пов'язаного із наповненням клієнтської бази шлюбного агентства.

Друга підсистема – це прикладна підсистема у вигляді веб-сайту шлюбного агентства, який використовуватиме запропонований метод для підбору партнерів.

Третя підсистема – це, власне, реалізація методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

База даних необхідна, в першу чергу, для збереження клієнтських анкет та всієї допоміжної інформації, необхідної для функціонування веб-сайту.

ДАТАЛОГІЧНА МОДЕЛЬ БАЗИ ДАНИХ



ПРИКЛАД РОБОТИ ВЕБ-САЙТУ ШЛЮБНОГО АГЕНТСТВА

The screenshot shows a web interface for a dating agency. The header includes the site name 'Шлюбне Агентство' and navigation links 'Головна', 'Послуги', and 'Від'. The main content is organized into three sections:

- Найкращі кандидати за двома методами:** Displays three profile cards for Amy Mooney, Alyssa Wise, and Robert Dixon, each with a 'Порівняти профіль' button.
- Співпадіння за методом порівняння фреймів:** Displays three pairs of profile cards. The first pair is Amy Mooney (73.5%) and David Lane (77.5%). The second pair is Alyssa Wise and Robert Dixon. The third pair is Erika Sanchez (72.75%) and Richard Fox.
- Порівняння за текстовими описами:** Displays three pairs of profile cards. The first pair is Jessica Wolfe (73.000%) and Barry Strong (69.910%). The second pair is Robert Dixon (66.000%) and Jessica Zhang (64.000%). The third pair is Christopher Fields (67.000%) and Edwin Gardner (67.000%).

На скріншоті продемонстровано отримання результатів роботи методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Результати відображаються у вигляді трьох переліків з обчисленими відсотками сумісності.

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ РОБОТИ МЕТОДУ

Для перевірки ефективності створеного методу та дослідження коректності його роботи було проведено порівняння між роботою методу та експертним висновком від моделі GPT-4o, доступ до якої здійснювався через ChatGPT.

В ході дослідження порівнювались текстові описи клієнтів із бази даних шлюбного агентства, з метою обрахунку оцінки подібності текстів. Також, було встановлено допустиме відхилення від експертного висновку в межах 5%. Крім цього, було встановлено поріг сумісності та несумісності клієнтів, відповідно до якого, якщо оцінка подібності текстових описів менша 50% – пара вважається несумісною.

Відповідно до проведених досліджень було встановлено, що в порівнянні з моделлю GPT-4o, точність визначення оцінок подібності текстових описів у випадку сумісних пар становить 86%, в той час як для несумісних пар – 78%. Отже, загальна точність методу порівняння текстових описів становить 82%.

	Сумісні	Несумісні
Сумісні	47	6
Несумісні	3	44

Матриця плутанини

■ ВИСНОВКИ

01

Проведено аналіз предметної області, пов'язаної із дослідженням діяльності шлюбних агентств та методами інтелектуального аналізу анкетних даних клієнтів шлюбних агентств.

02

Розроблено метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, у вигляді підходу, що має на меті паралельне використання двох методів.

03

Реалізовано прикладне застосування методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних у вигляді веб-сайту шлюбного агентства.

04

Досліджено ефективність роботи методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. ■

Anti-Plagiarism v-15.257

Максимальне співпадіння з одним документом 3.0%

Словники перевірки: en_US, ru_RU, ua_UA. **Помилоч в документах: 9%**

ID: 159992 Назва: КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА на тему Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних Додано в БД: 2024-12-16 Автора: Дмитро ЖУК Керівники: Олександр МАЗУРЕЦЬ Консультанти: Опоненти:	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	112955	1652	6270 (6%)	84 (5%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

Автор: Дмитро ЖУК

Співавтор:

Назва: Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних

Науковий керівник: Олександр МАЗУРЕЦЬ, к.т.н., доцент

Підрозділ: Кафедра комп'ютерних наук

Коефіцієнт подібності 1: 12%

Коефіцієнт подібності 2: 5%

Мікропробіли: 0

Заміна букв: 9

Інтервали: 0

Білі знаки: 48

Дата створення звіту: 2024-12-16 07:03:38.0

Після аналізу Звіту подібності констатую наступне:

Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.

Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.

Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедур. Таким чином робота не приймається.

Обґрунтування:

Дата 16.12.2024

експерт

Летровецький Р.Р.

РІШЕННЯ ЕКСПЕРТНОЇ КОМІСІЇ
КАФЕДРИ КОМП'ЮТЕРНИХ НАУК
ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ МАГІСТРА ДО ЗАХИСТУ
ЗА РЕЗУЛЬТАТАМИ АНАЛІЗУ ЗВІТУ ПОДІБНОСТІ

Підтверджуємо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних

Автор: Дмитро ЖУК

Спеціальність: 122 – Комп'ютерні науки

Освітня програма: освітньо-професійна

Науковий керівник: к.т.н., доц. каф. КН Олександр МАЗУРЕЦЬ

Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом. Робота приймається до захисту.	відповідає
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи	—
3	Виявлені запозичення не є плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. В зв'язку з цим мета роботи та поставлені завдання не були досягнені. Робота може бути допущена до захисту (наступного року) після того як буде відкоригована та допрацьована і успішно пройде повторну перевірку на академічний плагіат.	—
4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	—

Підтвердження:

Запозичення, виявлені в роботі, є законними і не є плагіатом, оскільки:

- 1) за програмою Anti-Plagiarism виявлені 3%, схожість виявлена зі звітом автора з науково-дослідної практики.
- 2) за програмою StrikePlagiarism КП1 12%, КЦ 5%,

які містять матеріали огляду предметної області; інші схожості є фрагментарними – містять поширені конструкції, загальновідомі терміни, скорочення та визначення, що, з урахуванням наведених обґрунтувань, відповідає характеру наукового дослідження і свідчить на користь кваліфікаційної роботи. КЦ 5% пояснюється переважним чином збігом із власними публікаціями автора за темою кваліфікаційної роботи магістра. Запозичення, виявлені в роботі є законними і не є плагіатом.

Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.

Керівник роботи

Олександр МАЗУРЕЦЬ

Гарант ОП

Руслан БАГРІЙ

Завідувач кафедри КН

Олександр БАРМАК



ВІДГУК НАУКОВОГО КЕРІВНИКА

на кваліфікаційну роботу магістра

гр. КНМ-23-1 Дмитра Жука за темою: *Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних*

1. Актуальність обраної теми

У сучасному світі соціальні аспекти міжособистісних взаємин є одними з ключових аспектів побудови стійкого та успішного суспільства. При цьому, пошук партнера для серйозних стосунків ускладнюється через швидкий темп життя, соціальні вимоги та високий рівень міграції. Традиційні методи роботи шлюбних агентств втрачають ефективність через неможливість ручного аналізу великої кількості анкет. Автоматизація цього процесу за допомогою інтелектуального аналізу даних дозволяє швидко та об'єктивно підбирати сумісних партнерів, обробляючи значні обсяги інформації. Таким чином, впровадження методів інтелектуального аналізу є актуальним, й допоможе шлюбним агентствам розширити базу клієнтів і підвищити ефективність підбору партнерів.

2. Відповідність роботи предметній області спеціальності 122 Комп'ютерні науки та загальним вимогам до наукових робіт

Кваліфікаційна робота магістра Дмитра Жука на тему «Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних» повною мірою відповідає предметній області спеціальності 122 «Комп'ютерні науки» та вимогам до кваліфікаційної роботи.

3. Професійні та особистісні якості магістранта

Під час виконання кваліфікаційної роботи магістра Дмитро Жук зарекомендував себе як відповідальний і наполегливий студент із високим рівнем організованості та професійного підходу. Він демонстрував здатність чітко планувати свою діяльність, ефективно розподіляти час і ресурси, а також дотримуватись усіх установлених вимог і дедлайнів. Студент проявив ініціативність у вирішенні складних завдань і прагненні досягти високих результатів, що свідчить про його цілеспрямованість та бажання розвиватися як фахівець.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Результати, отримані в результаті виконання кваліфікаційної роботи магістра, є результатом самостійної діяльності студента. Отримані положення наукової новизни та

інновації, описані в роботі, дозволили покращити існуючі методи в сфері визначення сумісності клієнтів шлюбних агентств.

5. Наукова новизна та оригінальність запропонованих підходів

Результати виконання кваліфікаційної роботи магістра містять інновації та наукову новизну, зокрема було якісно покращено визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

6. Ступінь оволодіння методами дослідження

Магістрант виявив високий ступінь оволодіння необхідними методами дослідження.

7. Повнота та якість розкриття теми роботи

Тема роботи в повній мірі обґрунтована й розкрита, проведено аналіз актуальності та відомих досліджень в межах обраної теми, поставлені завдання у роботі виконані, а також проведено аналіз результатів прикладного застосування запропонованих засобів методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

8. Логічність, послідовність, аргументованість, літературна грамотність викладу матеріалу

Структура роботи й послідовність викладення логічні та відповідні поставленій меті. Викладення матеріалу грамотне та виявляє високий ступінь відповідності стилю.

9. Можливість практичного застосування кваліфікаційної роботи, окремих її частин

Було створено інформаційну систему для визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Розроблено інформаційну систему визначення сумісності клієнтів шлюбних агентств, яка реалізує метод аналізу анкетних даних. Система дозволяє користувачам реєструватися, заповнювати анкети та знаходити сумісних партнерів. Вона складається з трьох підсистем: формування клієнтської бази, веб-сайту агентства та визначення сумісності, а також бази даних. Точність методу: 86% для сумісних пар, 78% для несумісних, із загальною точністю 82%.

10. Висновок про можливість допуску кваліфікаційної роботи до захисту, на яку оцінку заслуговує робота

Враховуючи високий рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «відмінно».

Науковий керівник

к.т.н., доц. каф. КН Олександр МАЗУРЕЦЬ



ВІДГУК ОПОНЕНТА

на кваліфікаційну роботу магістра

гр. КНМ-23-1 Дмитра Жука за темою: Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних

1. Актуальність обраної теми

У сучасних умовах пошук партнера для серйозних стосунків є одним із базових аспектів побудови успішного й само відтворюваного громадянського суспільства, проте цей процес стає дедалі складнішим через стрімкий темп життя, зростання соціальних вимог і високий рівень міграції. Необхідність існування шлюбних агентств обумовлена низкою соціальних проблем, які ускладнюють процес пошуку партнера з яким можна побудувати міцні сімейні зв'язки. Серед ключових факторів, які стають на заваді при пошуку партнера, є зайнятість на роботі, високі вимоги до партнера та місце проживання. В епоху цифровізації та постійного збільшення обсягів інформації та цифрових даних, покращення визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних є актуальним завданням. Це дасть змогу шлюбним агентствам обслуговувати більшу кількість клієнтів та значно розширити базу кандидатів для пошуку сумісностей.

2. Відповідність роботи предметній області спеціальності 122 Комп'ютерні науки та загальним вимогам до наукових робіт

Обрана тема «Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних», в межах якої виконані поставлені задачі, повною мірою відповідає предметній області спеціальності 122 «Комп'ютерні науки» та вимогам до кваліфікаційної роботи магістра.

3. Повнота розкриття мети та завдань дослідження

В роботі автор повністю розкриває мету дослідження та поставленні в межах теми завдання.

4. Наявність наукової новизни

Результати виконання кваліфікаційної роботи магістра містять наукову новизну, зокрема було покращено визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Розроблений метод працює шляхом використання методу порівняння анкет за фреймовою моделлю та нейромережевого визначення оцінок схожості текстових описів клієнтів, було отримано відсотки сумісності між клієнтськими парами.

5. Зміст кожного розділу роботи

Робота містить чотири розділи: у першому розділі виконано дослідження предметної області визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. Другий розділ присвячено розробці методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних. У третьому розділі виконано проектування інформаційної системи визначення сумісності клієнтів шлюбних агентств. У четвертому розділі виконано дослідження ефективності методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

6. Ступінь розкриття теми роботи

Тема кваліфікаційної роботи повною мірою розкрита та обґрунтована, проведено аналіз актуальності та відомих досліджень в межах обраної теми, поставлені завдання у роботі виконані, та проведено аналіз результатів прикладного застосування запропонованих методу і засобів.

7. Якість оформлення кваліфікаційної роботи

Оформлення роботи відповідає необхідним нормам та вимогам, які ставляться до оформлення кваліфікаційних робіт.

8. Недоліки кваліфікаційної роботи

Кваліфікаційна робота магістра виконана на високому рівні, проте є зауваження до переліку посилань, а також в тексті роботи присутні незначні граматичні помилки.

9. Загальний висновок (допускається чи не допускається до захисту), якої оцінки заслуговує кваліфікаційна робота

Враховуючи високий рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «відмінно».

Опонент (прізвище, ім'я, по батькові, посада, місце роботи)

Гнатюк Єлизавета Геннадіївна, д.т.н., доцент
доцент каф. комп'ютерної інженерії та інф.
систем

«16» 12 2024 р

підпис