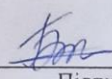



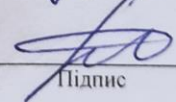
КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

на тему Метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом

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: студентка 4 курсу, група КН-19-1  В.Д. Блажук
Курс, група виконавця Підпис Ініціали, прізвище

Керівник: к.т.н., доцент кафедри КН  О.В. Мазурець
Науковий ступінь, посада Підпис Ініціали, прізвище

Нормоконтроль: к.т.н., доцент кафедри КН  Р.О. Багрій
Науковий ступінь, посада Підпис Ініціали, прізвище

До захисту допускаю:

Зав. кафедри КН, д.т.н., професор  О.В. Бармак
Підпис Ініціали, прізвище

01 06 2023 р.

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет інформаційних технологій

Кафедра комп'ютерних наук

Освітній ступінь бакалавр

Галузь знань 12 – Інформаційні технології

Спеціальність 122 – Комп'ютерні науки

Освітня програма освітньо-професійна програма підготовки бакалавра

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних наук


(підпис)

д.т.н., професор О.В. Бармак

« 06 » 03 2023 року

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА**

1. Тема кваліфікаційної роботи бакалавра: «Метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом»

2. Завдання видано студентці Блажук Вікторії Дмитрівні
(прізвище, ім'я, по батькові)

3. Керівник роботи доцент кафедри КН Мазурець Олександр Вікторович
(посада, прізвище, ім'я, по батькові)

4. Затверджено наказом університету від « 01 » 03 2023 р. № 5

5. Дата видачі завдання студенту: « 03 » 03 2023 р.

6. Зміст пояснювальної записки (перелік задач) та вихідні дані:

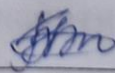
Провести аналіз предметної області визначення емоційного забарвлення цифрових текстів за семантичним аналізом. Виконати аналіз методів пошуку ключових слів. Виконати аналіз існуючих рішень щодо подібних задач. Розробити й програмно реалізувати метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом. Виконати дослідження ефективності розробленого методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом.

7. Календарний план виконання кваліфікаційної роботи бакалавра:

№	Назва етапів (розділів) кваліфікаційної роботи бакалавра	Термін виконання	Примітка
1	Вибір напрямку дослідження та узгодження тематики кваліфікаційної роботи бакалавра з керівником	грудень 2022	виконано
2	Ознайомлення з предметною областю, формулювання мети та задач дослідження, визначення об'єкта та предмета дослідження	січень 2023	виконано
3	Робота над розділом 1 – Характеристика предметної області та постановка задачі	січень 2023	виконано
4	Робота над розділом 2 – Метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом	березень 2023	виконано
5	Робота над розділом 3 – Програмна реалізація методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів	квітень 2023	виконано
6	Оформлення пояснювальної записки згідно вимог	травень 2023	виконано
7	Попередній захист кваліфікаційної роботи бакалавра	травень 2023	виконано
8	Захист кваліфікаційної роботи бакалавра на засіданні Екзаменаційної комісії	червень 2023	виконано

Виконавець:

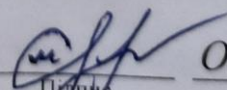
студентка 4 курсу, група КН-19-1
Курс, група виконавця


Підпис

В.Д. Блажук
Ініціали, прізвище

Керівник:

к.т.н., доцент кафедри КН
Науковий ступінь, посада


Підпис

О.В. Мазурець
Ініціали, прізвище

Анотація

Тема кваліфікаційної роботи бакалавра: Метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом

Виконавець кваліфікаційної роботи бакалавра: студентка групи КН-19-1 Блажук Вікторія Дмитрівна

Керівник кваліфікаційної роботи бакалавра: к.т.н., доцент кафедри КН Мазурець Олександр Вікторович

Кваліфікаційна робота бакалавра містить:

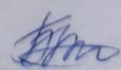
Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
74	31	8	30	4

Кваліфікаційна робота бакалавра присвячена розробці методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом та відповідної інформаційної системи у вигляді соціально орієнтованого вебресурсу, що дозволяє за текстовим контентом допису визначати його емоційне забарвлення.

Результатом виконання кваліфікаційної роботи бакалавра є соціально орієнтований вебресурс у вигляді блогу книголюбів. Розроблена інформаційна система дає можливість авторизації користувача та створення особистого кабінету з подальшою публікацією відгуків на книги. Розроблений метод визначення емоційного забарвлення дає змогу визначити категорію допису за його емоційним забарвленням та перелік емоційно значущих слів допису.

Ключові слова: соціально орієнтований вебресурс, емоційне забарвлення тексту, ключові слова, емоції, семантичний аналіз, допис, інформаційна система.

Виконавець: студентка 4 курсу, група КН-19-1
Курс, група виконавця


Підпис

В.Д. Блажук
Ініціали, прізвище

Зміст

Перелік скорочень	4
Вступ.....	5
Розділ 1 Характеристика предметної області і постановка задачі.....	7
1.1 Аналіз предметної області визначення емоційного забарвлення цифрових текстів за семантичним аналізом.....	7
1.2 Особливості семантичного аналізу контенту соціально орієнтованих вебресурсів.....	10
1.3 Аналіз методів пошуку ключових слів	14
1.4 Аналіз існуючих рішень для подібних задач	16
1.5 Мета, задачі та вимоги до реалізації програмної системи.....	19
1.6 Висновки до розділу 1	20
Розділ 2 Метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом	22
2.1 Опис і кроки методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів	22
2.2 Застосування Баєсового класифікатора та методу TF-IDF для визначення емоційного забарвлення дописів	27
2.3 Формування навчальних множин дописів для аналізу емоційного забарвлення.....	30
2.4 Проектування структури соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів	32
2.5 Даталогічна модель бази даних соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів	38
2.6 Вибір засобів розробки соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів	41
2.7 Висновки до розділу 2	42
Розділ 3 Програмна реалізація соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів	44

3.1 Структура модулів соціально орієнтованого вебресурсу	44
3.2 Особливості реалізації соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів	47
3.3 Тестування функціональних можливостей соціально орієнтованого вебресурсу.....	56
3.4 Дослідження ефективності методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом.....	63
3.5 Висновки до розділу 3	67
Висновки	69
Перелік посилань.....	72
Додатки	

Перелік скорочень

Скорочення, термін, позначення	Пояснення
NLP	Natural Language Processing
TF	Term Frequency
IDF	Inverse Document Frequency
SVM	Support Vector Machines
PHP	Hypertext Preprocessor
ID	Identity Documen
FK	Foreign Key
SQL	Structured query language
HTML	Hypertext Markup Language
ML	Machine learning
MVC	Model-View-Controller
БД	База даних
СКБД	Система керування базами даних
КРБ	Кваліфікаційна робота бакалавра
КН	Комп'ютерні науки
ІІ	Штучний інтелект

Вступ

Кваліфікаційна робота бакалавра присвячена розробці методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом та відповідної інформаційної системи у вигляді соціально орієнтованого вебресурсу, що дозволяє за текстовим контентом допису визначати його емоційне забарвлення.

Актуальність. Розвиток та активне використання соціально орієнтованих вебресурсів призводять до щоденного збільшення кількості контенту в мережі. Кожен написаний та опублікований допис несе в собі певне емоційне забарвлення, і як результат, може впливати на читачів. Тому, аналіз дописів допомагає відстежувати думки та настрої соціуму. Проблема визначення емоційного забарвлення дописів стає особливо актуальною під час значних змін у суспільстві. Саме тому, дослідження даної проблеми є актуальним в наш час.

Об'єкт дослідження – процес визначення емоційного забарвлення цифрових текстів за семантичним аналізом.

Предмет дослідження – моделі, методи, алгоритми та засоби для семантичного аналізу цифрових текстів та визначення їх емоційного забарвлення.

Мета кваліфікаційної роботи бакалавра – розробка й програмна реалізація методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом.

Завдання кваліфікаційної роботи бакалавра – Провести аналіз предметної області визначення емоційного забарвлення цифрових текстів за семантичним аналізом. Виконати аналіз методів пошуку ключових слів. Визначити особливості семантичного аналізу контенту соціально орієнтованого вебресурсу. Виконати аналіз наявних рішень щодо подібних задач. Розробити метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом. Спроекувати структуру прикладного соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного

забарвлення дописів. Спроекувати структуру бази даних соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів. Виконати вибір засобів розробки соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів. Виконати програмну реалізацію соціально орієнтованого вебресурсу. Провести тестування розробленого соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів. Виконати дослідження ефективності розробленого методу визначення емоційного забарвлення дописів за семантичним аналізом.

Розділ 1 Характеристика предметної області і постановка задачі

1.1 Аналіз предметної області визначення емоційного забарвлення цифрових текстів за семантичним аналізом

Для реалізації методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом, необхідно провести детальний аналіз предметної області, зокрема, дослідити сучасні методи визначення емоційного забарвлення цифрових текстів за їх семантичним аналізом.

Семантичний аналіз текстів є важливою і поширеною областю досліджень у сфері інформаційних технологій. З точки зору мовознавства, поняття семантики означає тлумачення значень мовних одиниць, тобто, окремих слів та їх складових частин, словосполучень, сталих виразів та фразеологізмів [1].

Відповідно до цілей семантики, дану науку можна поділити на дві основні концепції: лексичну семантику та фразову семантику [2]. У першому випадку розглядаються лише буквальні значення відокремлених слів незалежно від контексту. Фразова семантика досліджує значення слів відповідно до контексту в якому вони вживаються. Таким чином, смислові межі слова значно розширюються. З цього випливає, що одне слово може нести в собі різний сенс. Такий підхід дозволяє краще зрозуміти значення слова не лише в його буквальному розумінні, а й з точки зору використання слова як метафори, алегорії чи порівняння.

Під час проведення семантичного аналізу тексту, текст розглядається як сукупність взаємопов'язаних мовних одиниць. Семантичний аналіз тексту передбачає необхідність проведення оцінки різних показників тексту, таких як [3]:

- загальна кількість символів, які містяться в тексті;
- загальна кількість слів;
- кількість речень, з яких складається текст;

- кількість та частота зустрічання слів у тексті (для кожного слова);
- кількість помилок у тексті (лексичних, орфографічних, пунктуаційних);
- семантичне ядро (набір слів, які формують тематику тексту).

Семантичний аналіз текстової інформації з використанням інформаційних технологій перш за все ґрунтується на обробці природної мови. Природною мовою називається мова, яка не створена за допомогою будь-яких інформаційних технологій, тобто, мова спілкування між людьми [4].

Обробка природної мови (Natural Language Processing – NLP) – це аналіз текстової інформації представленої природною мовою з використанням комп’ютерних технологій, зокрема штучного інтелекту та лінгвістичних методів аналізу мови [5].

Процес обробки природної мови можна розділити на декілька етапів [6]:

- розпізнавання мовлення (використовується у випадку, коли потрібно перетворити голосовий запис або зображення на текст для подальшого аналізу);
- визначення частин мови;
- усунення двозначності (вибір одного значення слова);
- розпізнавання іменованих сутностей (пошук власних назв);
- аналіз емоційного забарвлення (виділення з тексту певних суб’єктивних ознак – емоцій).

Емоційно-смісловий аналіз текстів – це набір методів для автоматизованого пошуку емоцій у природній мові [7]. Для проведення дослідження емоційного забарвлення текстів, необхідно чітко визначити яким чином емоції проявляються в тексті, та власне, що таке емоції.

Емоції – це реакції людини на певні події та ситуації, які з нею відбуваються. Емоції можна розділити на позитивні, нейтральні та негативні [8].

До позитивних емоцій відносять радість, любов, захоплення, задоволення, захват, симпатію. Вони характеризуються піднесеним емоційним станом та викликають почуття щастя та задоволення. Проявом негативних емоцій може бути страх, злість, лють, роздратованість, сум, розпач, тривога, обурення, відраза. Часто, негативні емоції можуть супроводжуватись агресивною

поведінкою та, навіть, нести шкоду оточенню. Нейтральні емоції – це щось середнє між позитивними та негативними емоціями. Нейтральні емоції можна описати як байдуже ставлення людини до чогось, що характеризується бездіяльністю людини до певної проблеми [9].

Основний інструмент прояву емоцій через текст це ключові слова, які мають чітке емоційне навантаження (добре, погано, сумно, страшно). Присутність подібних слів у тексті дозволяє читачеві відчувати емоційний стан оповідача. Отже, ключові слова відіграють важливу роль під час визначення емоційного забарвлення текстів. Ключовим словом вважається слово, або сталий вираз, яке має конкретне значення та смислове навантаження [10].

Існує два основних шляхи для реалізації системи автоматизованого аналізу емоційного забарвлення текстів: використання методів та систем штучного інтелекту, або, методів побудованих на формуванні словникової бази та певних правил.

Якщо використовувати ШІ, то визначення емоційного забарвлення дописів зводиться до задачі класифікації. Задача класифікації полягає у формуванні заздалегідь продуманих класів з відповідною характеристикою. У випадку емоційно-смиислового аналізу, класами будуть позитивні та негативні емоції [11].

Для того щоб процес був автоматизований, спочатку необхідно навчити класифікатор на підготовленій тестовій вибірці. Ця вибірка повинна включати набір текстів, які заздалегідь розподілені до певного класу емоційної забарвленості. Після навчання модель буде здатна самостійно класифікувати тексти. Чим краще навчена модель, тим точнішою буде її подальша класифікація. Такий підхід потребує великої вибірки тренувальних даних, а їхня підготовка може бути досить трудомісткою та довгою.

Згідно з другим варіантом, семантичний аналіз емоційного забарвлення тексту здійснюється відповідно до сформованого словника тональності та правил [12]. Словник тональності формується на основі ключових слів, які можуть виражати певний спектр емоцій. Кожному такому слову надається певна

оцінка емоційності, яка відносить його до певного класу емоцій (позитивні, нейтральні та негативні). Емоційно-смісловий аналіз тексту відбувається шляхом пошуку ключових слів із тексту, та зіставлення їх із відповідними словами зі словника тональності.

Визначення емоційного забарвлення текстової інформації досить новий напрямок, який постійно розвивається. Прояв емоцій через текст відбувається завдяки додаванню в текст ключових слів, які характеризують певну емоцію. Тому, для реалізації методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом необхідно визначити ключові слова, які мають емоційне забарвлення.

1.2 Особливості семантичного аналізу контенту соціально орієнтованих вебресурсів

Розвиток комп'ютерних технологій та стрімке розширення інтернет покриття, призвело до виникнення безлічі соціально орієнтованих вебресурсів. Це стало причиною переходу частини соціального життя людей у віртуальний світ месенджерів, соціальних мереж та різних платформ для спілкування. Крім цього, це допомогло бізнесу вийти на нову сходинку розвитку, адже стало можливе просування товарів та послуг через інтернет.

Отже, соціально орієнтовані вебресурси – це інформаційні технології, метою яких є створення та поширення інформації між користувачами через всесвітню мережу «Інтернет» [13].

Використання вебресурсів надає користувачеві безліч можливостей, серед яких можна виділити найкорисніші:

- вільний доступ до інформаційних ресурсів;
- можливість створення та розміщення власного контенту;
- обмін досвідом з іншими користувачами;
- свобода самовираження;
- участь в обговореннях.

Більшу частину часу використання соціально орієнтованих вебресурсів користувачі витрачають на обмін, створення та поглинання інформації. У цьому й полягає основна мета таких сервісів, де користувачі одночасно є авторами й читачами.

Соціально орієнтовані вебресурси можна класифікувати на декілька груп:

- соціальні мережі (Instagram, Facebook, TikTok);
- форуми;
- блоги;
- чати для спілкування;
- сервіси відеохостингу (YouTube);
- поштові системи (Gmail, Outlook);
- месенджери (Telegram, Viber, Facebook Messenger).

Найпопулярнішими серед соціально орієнтованих вебресурсів, однозначно, можна вважати соціальні мережі. Соціальна мережа – це інформаційна система для великої кількості користувачів, метою якої є пошук однодумців, створення власного контенту, просування бізнесу та комунікація з іншими користувачами. Перевагою соціальних мереж є можливість персоналізації контенту [14].

З появою соціальних мереж та месенджерів, форуми та чати для спілкування значно втратили свою популярність. Це зумовлено тим, що соціальні мережі вміщують в собі більше різноманітного контенту, в той час як форуми та чати, здебільшого вузько направлені. Як результат, користувачів у форумах та чатах набагато менше. Відмінність між форумом і чатом полягає в тому, що форуми це довготривалі сервіси для комунікації на певні тематики. Користувачі форумів можуть надсилати та переглядати повідомлення в будь-який зручний для них час. На відмінну від форумів, чати розраховані на спілкування в реальному часі на різні теми, які цікавлять співрозмовників саме в момент комунікації [15].

Сервіси відеохостингу – це вебсайти, які дозволяють переглядати відео через спеціальний відео програвач. В більшості сервісів відеохостингу

передбачена можливість завантаження власного відео контенту. Крім цього, до кожного відео закріплений чат, в якому користувачі можуть залишати коментарі та вести обговорення відносно переглянутого відео [16].

Поштові системи – це сервіси передачі електронних повідомлень між користувачами. Здебільшого, послугами поштових систем користуються для отримання офіційних листів, які стосуються роботи, бізнесу та навчання. Також, широкого застосування набули розсилки листів з рекламою [17].

За останні декілька років, особливої популярності серед користувачів інтернету набули месенджери. Месенджер – це спеціально розроблена інформаційна система у вигляді додатка для смартфона або програми для ПК, яка дозволяє користувачам спілкуватись між собою та обмінюватись файлами [18].

В сучасних умовах життя, соціально орієнтовані вебресурси набули особливої популярності не тільки як інструмент комунікації, а і як спосіб формування суспільних та політичних течій, які здатні впливати на життя усього людства. Зростання популярності соціально орієнтованих вебресурсів зумовило їхнє розширення практично в усі сфери життєдіяльності. В результаті цього, соціальні мережі стали чудовим інструментом впливу на широкі маси, а також, відкрили можливість лідерам думок впливати на погляди та переконання своєї аудиторії. Для творців контенту важлива не лише боротьба за увагу аудиторії, а й можливість впливати на неї. Тому, формування громадської думки це одне з найважливіших завдань для творців контенту в мережі [19].

Велика кількість авторів зумовлює жорстку конкуренцію за увагу споживачів інформації [19]. Тому, часто, автори вдаються до не найкращих методів створення контенту, таких як, публікація неправдивої інформації з метою збільшення аудиторії. Як наслідок, це може призвести до неправильної громадської думки та хибних суджень, що своєю чергою, може стати поштовхом для формування певних соціальних та політичних течій. Ось чому, на творцях контенту лежить велика відповідальність за формування думки своєї аудиторії.

Для відстеження соціальних процесів, які відбуваються в мережевому суспільстві залучено безліч комп'ютерних технологій, які допомогли автоматизувати цей процес. Адже, через велику кількість інформації, яка міститься в інтернеті, для людини фізично не можливо відстежувати кожне опубліковане відео чи статтю в мережі. Саме тому, на допомогу приходять автоматизовані системи аналізу текстових, аудіо або відео матеріалів та зображень. Це дає змогу швидко відстежувати дописи, які можуть порушувати політику використання вебресурсів, публікують фейкову інформацію або несуть небезпеку для конкретної людини чи групи людей.

Найпоширеніша сфера використання методів визначення емоційного забарвлення текстів – це аналіз дописів у соціальних мережах, коментарів під дописами користувачів та відгуків на товари та послуги. Визначення емоційного забарвлення текстової інформації взятої із соціально орієнтованих вебресурсів може використовуватись для проведення соціальних досліджень. Також, це допомагає відстежувати судження, що містять агресію, заклики до конфліктів, образливу поведінку, цькування та фейкову інформацію. Для бізнесу така технологія теж важлива, адже це дозволяє відстежувати негативні відгуки та оперативно реагувати на проблеми.

Особливо актуальна ця технологія стає під час сильних змін у суспільстві, спричинених політичними конфліктами, війною, масовою непокорю суспільства владі, пандемією та іншими природними явищами, які загрожують значній частині суспільства. Тому, розробка методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом буде корисною при відстеженні масових соціальних рухів в значній частині суспільства, а також, для вирішення більш локальних проблем, які можуть містити агресію або образливу поведінку до певної людини або групи людей. Крім цього, з точки зору бізнесу, таку інформаційну систему можна використовувати для створення аналітики на основі відгуків і коментарів до послуг та товарів.

1.3 Аналіз методів пошуку ключових слів

Основою методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом є пошук ключових слів в тексті, які несуть емоційне забарвлення. Існує безліч методів та алгоритмів автоматизованого пошуку ключових слів у текстах. Починаючи з простих статистичних підходів, основою яких є підрахунок частотності слів в тексті, закінчуючи складними системами штучного інтелекту та машинного навчання.

Найпоширенішими статистичними методами пошуку ключових слів у текстах є частотна оцінка TF та оцінка TF-IDF.

Частотна оцінка або частота термінів TF (term frequency) визначає, як часто термін, тобто слово або словосполучення, зустрічається в тексті. Оскільки, одне і те ж слово матиме різну частотну оцінку, залежно від величини тексту, то для нормалізації цього визначення, прийнято ділити кількість повторів слова на загальну кількість слів у тексті [20].

Формула для обчислення TF матиме наступний вигляд:

$$TF_i = \frac{n_i}{\sum_k n_{ik}} \quad (1.1)$$

де TF_i – слово для якого обчислюється частотна оцінка, n_i – кількість входжень даного слова в текст, $\sum_k n_{ik}$ – загальна кількість слів в тексті.

Логічно можна припустити, що чим частіше слово з'являється в документі, тим краще воно характеризує даний текст. Проте, в такому підході є значний недолік, адже кожен текст складається з великої кількості загальних слів та сполучників, які не несуть жодного смислового навантаження, але при обчисленні частотної оцінки можуть потрапити в список ключових слів. Для розв'язання такої проблеми використовується список стоп-слів, таких як: і, або, але, проте, зокрема, зазвичай тощо. Перед обчисленням частотної оцінки з тексту видаляються слова, які містяться у списку стоп-слів. Це дозволяє зробити визначення ключових слів більш точнішим.

Інший спосіб виключення загальних слів полягає у використанні поняття зворотної документарної частоти слова IDF (inverse document frequency). IDF

зменшує значущість слів, які найбільше зустрічаються в тексті, та дає більшу вагу ключовим словам [21].

IDF обчислюється за формулою:

$$IDF_i = \log \frac{D}{d_i} \quad (1.2)$$

де IDF_i – слово, для якого обчислюється інверсна частота, D – кількість документів, які використовуються для аналізу, d_i – кількість документів у яких присутнє дане слово.

Таким чином, для зниження важливості загальних слів та збільшення ваги ключових слів, які характеризуватимуть текст, необхідно TF помножити на IDF. Даний підхід називається оцінка TF-IDF. Також, описаний алгоритм відмінно підходить для аналізу невеликих та неструктурованих текстових файлів, якими можуть бути дописи у соціальних мережах, відгуки користувачів на товари та послуги та інші види висловлювання особистої думки, щодо певної проблеми.

Крім статистичних методів для визначення ключових слів, використовують підходи на основі графів. Найпопулярнішим підходом на основі графів є модель TextRank. Модель застосовується не лише для пошуку ключових слів, а й для пошуку найважливіших речень в тексті.

У випадку знаходження ключових слів будується орієнтований граф, вершинами якого є слова з тексту. На основі порядку слів у реченні, будуються зв'язки графу. Вага ребра між двома словами буде більшою, якщо ці слова частіше зустрічаються в реченні одне після одного. Після побудови графу вимірюється ступінь вершини, тобто число, яке визначає важливість слова в тексті [22].

Ступінь вершини обчислюється за формулою:

$$D_v = (D_v^{in} + D_v^{out}) / (N - 1) \quad (1.3)$$

де D_v – ступінь вершини, (D_v^{in}) – кількість ребер, які потрапляють у вершину, D_v^{out} – кількість ребер, які виходять із вершини, N – кількість вершин у графі.

Таким чином, чим більший ступінь вершини, тим більше значення відіграє дане слово в тексті. Основою моделі TextRank є знаходження взаємозв'язків між словами в тексті. Головна ідея полягає в тому, що у зв'язному

тексті пов'язані слова утворюють мережу зав'язків, подібну до тої, яка створюється природною мовою. Отже, на основі аналізу зав'язків між текстовими одиницями можна визначити ключові слова.

Методи та системи штучного інтелекту та машинного навчання широко використовуються для задач обробки тексту, в тому числі й для пошуку ключових слів. Одним з популярних та часто використовуваних алгоритмів ШІ є метод опорних векторів (Support Vector Machines – SVM). Алгоритм SVM складається з трьох частин: побудова бази ключових слів, вибір навчальних зразків та навчання алгоритму [23]. Даний алгоритм застосовується для вилучення ключових слів із текстових документів, які містять у собі заголовки, анотацію, основну частину та висновки. Такий алгоритм не підходить для аналізу невеликих текстів, таких як дописи у соціальних мережах, адже там відсутні потрібні елементи. Отже, SVM варто застосовувати до статей, наукових публікацій та інших документів, які відповідають зазначеним вимогам.

Розробка методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом, передбачає аналіз невеликої за обсягом текстової інформації розміщеної в соціальних мережах. Для реалізації даного задуму необхідно обрати метод для визначення ключових слів, які містяться в дописах користувачів, які необхідно проаналізувати. Після проведеного аналізу наявних методів та алгоритмів для пошуку ключових слів, було вирішено використовувати алгоритм TF-IDF, який відмінно підходить для аналізу невеликих текстів.

1.4 Аналіз існуючих рішень для подібних задач

Семантичний аналіз текстів досить популярний напрямок у сфері комп'ютерної лінгвістики, який дозволяє зрозуміти зміст тексту, використовуючи інформаційні технології. Такий підхід широко використовується для задач машинного перекладу текстів, пошуку плагіату, розпізнавання та генерації природної мови, пошуку ключових слів та інших

ключових ознак у текстах. Кожна з таких задач має велику практичну цінність, а тому потребує постійного дослідження та вдосконалення. Як висновок, існує безліч прикладів програмного забезпечення, яке розв'язує задачу семантичного аналізу тексту.

Найпоширенішими прикладами існуючого програмного забезпечення для даної предметної області, є вебсайти призначенні для аналізу текстів. Зазвичай, такі сайти розболяться для копірайтерів та авторів різноманітних статей. Одним з популярних сайтів для аналізу текстів є онлайн-сервіс «Istio» [24].

Онлайн-сервіс «Istio» пропонує ряд послуг, серед яких основними є семантичний аналіз тексту та перевірка правопису. Семантичний аналіз тексту оцінює текст за такими ключовими параметрами: кількість слів у тексті, насиченість ключовими словами, щільність ключових слів, запам'ятованість, орфографічні та друкарські помилки.

Для проведення семантичного аналізу необхідно вставити текст в потрібне поле та натиснути кнопку «Аналіз». Для прикладу взято текст із розділу 1.1 (Рисунок 1.1). Як видно з проведеного аналізу, система знаходить помилки в тексті та наводить підказки як їх можна виправити. Відстежуються не лише граматичні помилки, а й пунктуаційні, такі як зайві пробіли та неправильні знаки пунктуації.

Отже, онлайн-сервіс «Istio» відмінно підходить для вирішення задачі семантичного аналізу тексту. З його допомогою можна легко перевірити текст на помилки та швидко виправити їх користуючись підказками. Також, система добре справляється із пошуком ключових слів.

Серед українських вебресурсів не було знайдено інформаційних систем, які можуть визначати емоційне забарвлення тексту. Проте, такий ресурс було знайдено та проаналізовано на одному із закордонних вебсайтів під назвою «MonkeyLearn» [25].

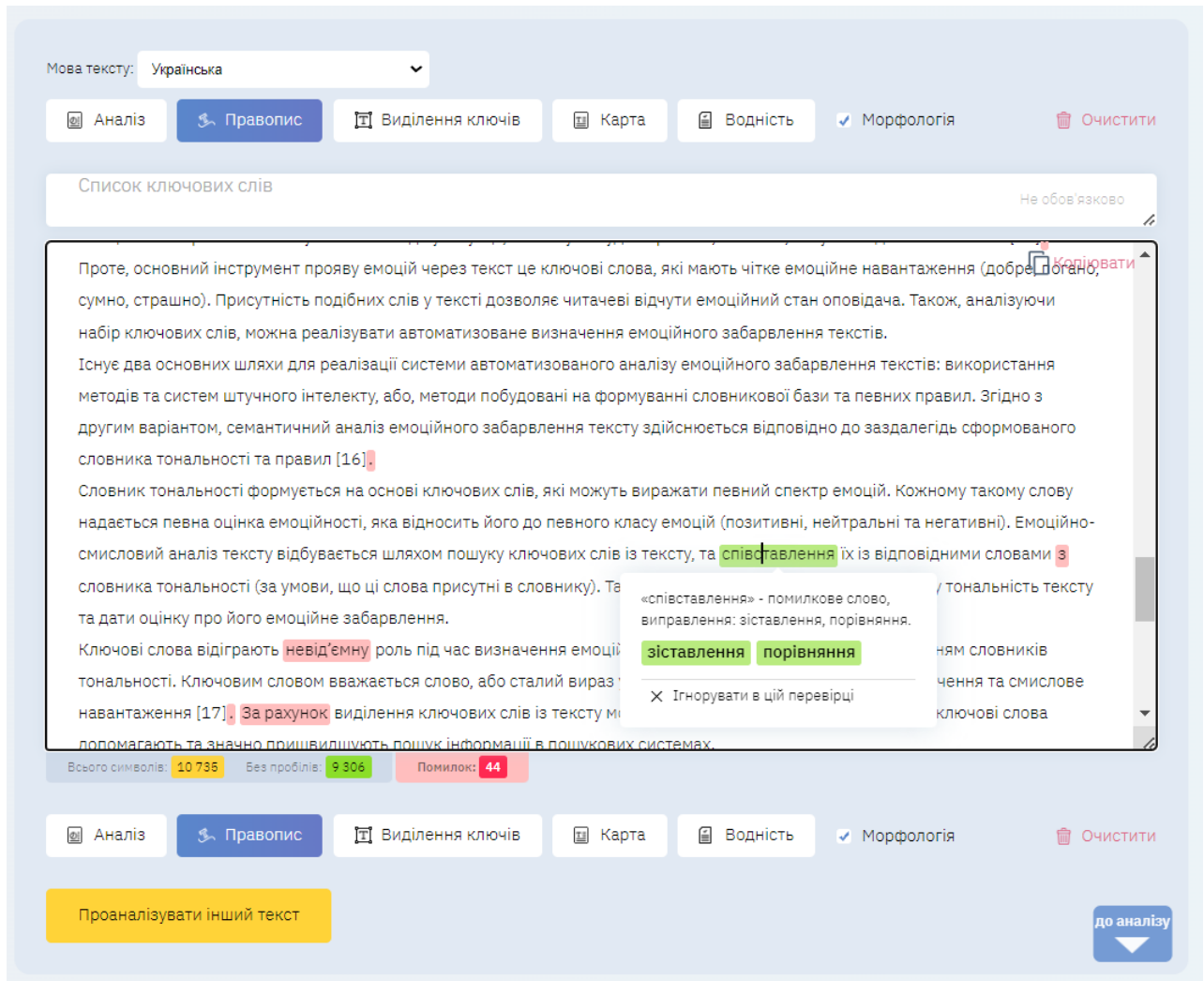


Рисунок 1.1 – Аналіз тексту онлайн-сервісом «Istio» [24]

Розглянутий сервіс аналізує введений текст та видає відсоткову оцінку негативності або позитивності емоційного забарвлення. Тобто, система виконує алгоритм класифікації тексту до одного з двох класів: позитивного або негативного. Як приклад, було введено короткий текст про опис дощової погоди, яка викликає негативні емоції (Рисунок 1.2). Результатом аналізу стала оцінка в майже 74% негативності емоційного забарвлення розглянутого тексту. Важко не погодитись з отриманим результатом, тому можна зробити висновок, що даний сервіс добре виконує завдання аналізу емоційного забарвлення в текстах.

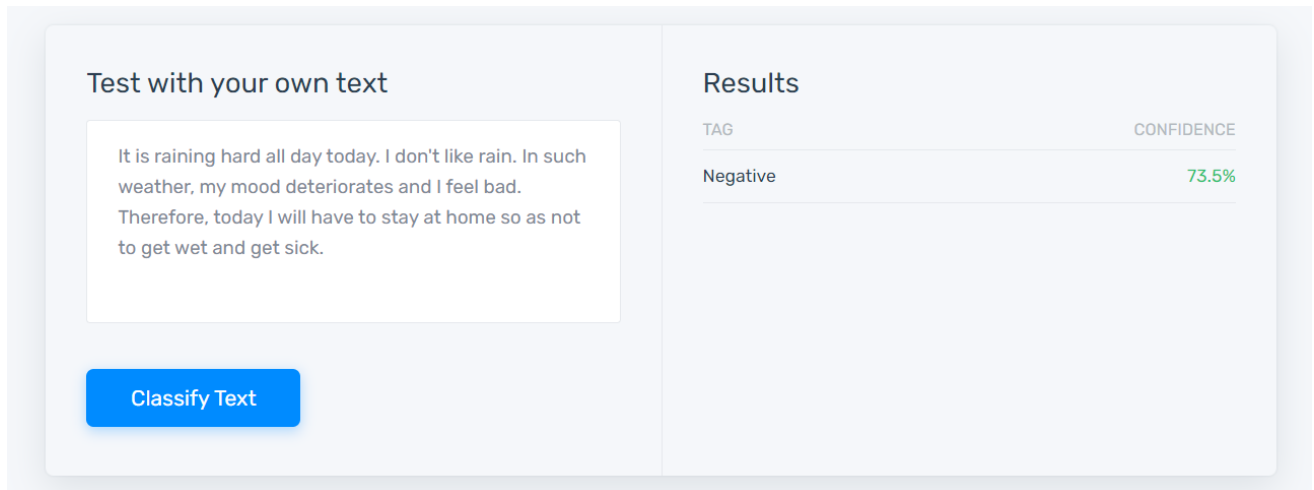


Рисунок 1.2 – Результати визначення емоційного забарвлення тексту сервісом «MonkeyLearn» [25]

Отже, проаналізувавши існуючі програмні рішення для вирішення подібних задач можна зробити висновок, що методи та системи семантичного аналізу текстів мають досить широке застосування. І, оскільки, ця область є досить актуальною, то її дослідження та вдосконалення існуючих методів обробки текстів, також є актуальним завданням. Крім цього, звернувши увагу на те, що не було знайдено українських сервісів для визначення емоційного забарвлення текстів, то завдання, яке передбачає створення такого сервісу є актуальним.

1.5 Мета, задачі та вимоги до реалізації програмної системи

Метою роботи є розробка й програмна реалізація методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом. Для досягнення поставленої мети необхідно вирішити такі задачі:

- розробити метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом;
- спроектувати структуру прикладного соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів;

- спроектувати структуру бази даних соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів;
- виконати вибір засобів розробки соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів;
- виконати програмну реалізацію соціально орієнтованого вебресурсу;
- провести тестування розробленого соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів;

Відповідно до мети роботи, основними функціями соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів є:

- аутентифікація користувачів;
- робота користувача з персональними даними;
- додавання власних дописів;
- перегляд особистих сторінок інших користувачів;
- перегляд журналу обраного користувача з усіма дописами;
- визначення емоційного забарвлення обраного допису;
- візуалізація емоційно значущих слів.

Результатом виконання поставлених задач та вимог має бути соціально орієнтований вебресурс з автоматизованим визначенням емоційного забарвлення дописів за семантичним аналізом, який класифікує дописи до певної категорії (позитивної або негативної).

1.6 Висновки до розділу 1

Семантичний аналіз текстів та обробка природної мови займають окреме місце в галузі інформаційних технологій. Адже, постійне збільшення текстової інформації в цифровому суспільстві, призвело до потреби автоматизованого аналізу цієї інформації з метою пошуку потрібних та важливих даних.

Розвиток цифрового суспільства та розширення інформаційного простору соціальних мереж призвело до збільшення їхнього впливу на життя суспільства.

Щодня люди публікують дуже багато дописів у яких висловлюють свою думку, почуття, емоції. Кожен з таких дописів має певне емоційне забарвлення та має певний вплив на людей, які його прочитають. Таким чином, через висловлення думки у соціальних мережах, месенджерах та інших соціально орієнтованих вебресурсах, творці контенту можуть нести вплив на величезні маси людей. Дослідження емоційного забарвлення дописів дає змогу відстежувати політичні, релігійні та соціальні погляди суспільства.

Основою для розробки системи автоматизованого визначення емоційного забарвлення дописів є пошук ключових слів у тексті. Це необхідно для того щоб знайти слова, які несуть емоційне навантаження для тексту, та класифікувати за цими словами емоційне забарвлення допису. Існує безліч методів та алгоритмів пошуку ключових слів, кожен з яких має свої переваги та недоліки. Для невеликих текстів найкраще підходить метод пошуку ключових слів TF-IDF, який формує список найбільш релевантних слів, які характеризують текст.

Численні дослідження даної області призвели до виникнення великої кількості вебресурсів, які вирішують завдання семантичного аналізу текстів, зокрема визначення їхнього емоційного забарвлення. Проте, досить мало таких ресурсів, які можуть проводити семантичний аналіз україномовних текстів. Саме тому, розробка та програмна реалізація методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом, є актуальним завданням.

Відповідно до проведеного аналізу предметної області й дослідження проблеми автоматизованого визначення емоційного забарвлення текстів за їх семантичним аналізом, було сформульовано мету роботи – розробку й програмну реалізацію методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом.

Розділ 2 Метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом

2.1 Опис і кроки методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів

Метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів призначений для одержання вихідних даних у вигляді категорії допису за його емоційним забарвленням та переліку емоційно значущих слів тестового допису за вхідними даними у вигляді позитивної й негативної навчальних множини дописів і тестового допису для аналізу емоційного забарвлення, для чого застосовуються формування векторів семантично важливих слів за категоріями емоцій та визначення емоційного забарвлення допису з використанням Баєсового класифікатора та методу TF-IDF.

Метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів поділяється на два окремі етапи. Перший етап базується на формуванні векторів семантично важливих слів за категоріями емоцій на основі завчасно підготованих тестових зразків. Другий етап – це, власне, визначення емоційного забарвлення допису, за допомогою раніше сформованих векторів семантично важливих слів та методу наївного Баєса (Рисунок 2.1).

За вхідні дані для роботи методу будуть братись множини навчальних дописів, розділені за категоріями відповідно до емоційного забарвлення, та тестовий допис для аналізу його емоційного забарвлення. Кожен з етапів методу визначення емоційного забарвлення дописів складається з послідовних кроків по роботі із текстовим контентом дописів. Відповідно до цього, вхідними даними для обох етапів будуть цифрові тексти.

Отже, вхідними даними першого етапу є множини навчальних дописів, які розділені на дві категорії: позитивні дописи та негативні. Чим більша кількість даних буде підготована для цього етапу, тим кращими будуть результати визначення емоційного забарвлення тексту на наступному етапі.

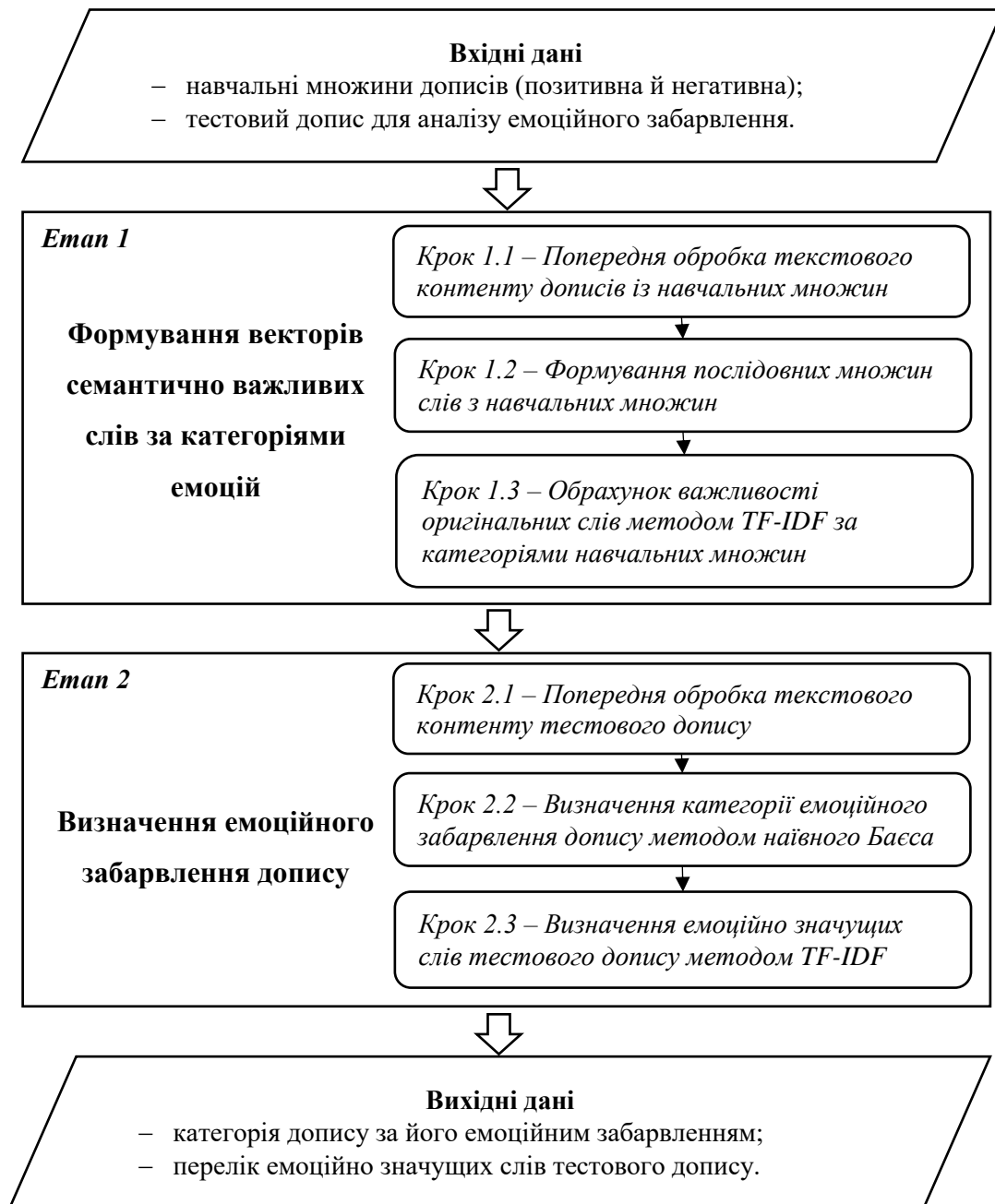


Рисунок 2.1 – Схема методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів

Перший етап методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів складається з таких кроків (Рисунок 2.2):

- попередня обробка текстового контенту дописів із навчальних множин;
- формування послідовних множин слів із навчальних множин;
- обрахунок важливості оригінальних слів методом TF-IDF за категоріями навчальних множин.

Попередня обробка тексту охоплює видалення розділових знаків та зміну усіх літер верхнього регістру на нижній. Видалення розділових знаків із тексту потрібне для того, щоб очистити текст від зайвих символів, які не впливають на його семантичний аналіз. Зміна регістру необхідна для того, щоб програма враховувала однакові слова незалежно від регістру в якому вони написані. Таким чином, попередня обробка текстового контенту дописів дає змогу підготувати вхідний текст до подальшого аналізу.

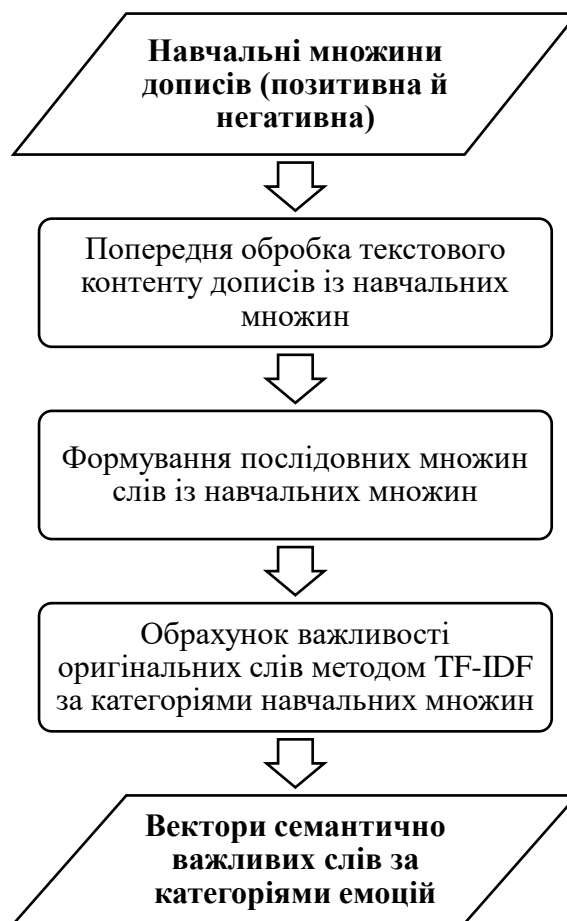


Рисунок 2.2 – Схема виконання першого етапу методу визначення емоційного забарвлення дописів

Під формуванням послідовних множин слів із навчальних множин мається на увазі формування двох множин, які включатимуть тексти з позитивним емоційним забарвленням та негативним. Таким чином, буде отримано дві окремі множини слів (позитивна та негативна).

На наступному кроці відбувається обрахунок важливості оригінальних слів методом TF-IDF, окремо для кожної категорії. Тобто, кожному слову надається певна вага, яка впливає на значущість даного слова. В контексті емоційного аналізу текстів, словам надається певна емоційна оцінка. Це допомагає при навчанні класифікатора, адже при визначенні емоційного забарвлення текстів важливо звертати увагу не на частоту появи слова в тексті, а на його емоційне навантаження.

Таким чином, результатами роботи першого етапу будуть вектори семантично важливих слів для кожної категорії емоційного забарвлення тексту.

Вхідними даними для другого етапу є сформовані раніше вектори семантично важливих слів за категоріями емоцій та тестовий допис для аналізу його емоційного забарвлення (Рисунок 2.3).

Перед тим, як провести аналіз тестового допису, необхідно повторити перший крок з попереднього етапу, тобто, потрібно провести попередню обробку вхідного тексту. Далі, оброблений текст передається класифікатору Баєса, який відносить його до однієї із двох емоційних категорій (позитивної чи негативної). Важливо відмітити, що вхідними даними для класифікатора є не просто слова з тексту, а їхні числові значення, які були обраховані на першому етапі методу.

Крім визначення категорії, до якої належить тестовий допис, за допомогою методу TF-IDF формується перелік емоційно значущих слів допису. Таким чином, якщо допис буде віднесено до позитивної категорії то для цього допису буде сформовано перелік позитивно забарвлених слів. І навпаки, якщо допис виявиться негативним, то перелік включатиме в себе негативно забарвлені слова.

Результати роботи другого етапу можна вважати результатами роботи всього методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів. Отриманими вихідними даними будуть передбачення класифікатора, який відноситиме тексти до негативної чи позитивної категорії

емоцій. Крім цього, буде сформовано список ключових слів із тексту, які мають емоційне забарвлення.



Рисунок 2.3 – Схема виконання другого етапу методу визначення емоційного забарвлення дописів

Аналіз схеми методу, яка наведена на рисунку 2.1, дозволяє краще зрозуміти, які кроки необхідно пройти щоб отримати бажаний результат, не заглиблюючись в роботу алгоритмів, які необхідно використати для досягнення цього результату. Таким чином, знаючи якими є вхідні дані, та яким має бути результат, можна спроектувати та продумати кроки, необхідні для подальшої розробки інформаційної системи.

Таким чином, було розроблено метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів, який призначений для одержання вихідних даних у вигляді категорії допису за його емоційним забарвленням та переліку емоційно значущих слів тестового допису за вхідними даними у вигляді

позитивної й негативної навчальних множини дописів і тестового допису для аналізу емоційного забарвлення, для чого застосовуються формування векторів семантично важливих слів за категоріями емоцій та визначення емоційного забарвлення допису.

2.2 Застосування Бассового класифікатора та методу TF-IDF для визначення емоційного забарвлення дописів

Розробка програмних продуктів не можлива без вибору необхідних методів та алгоритмів для вирішення поставлених задач. Зазвичай, для однієї й тієї ж задачі існує декілька методів її вирішення. Тому, важливо обрати саме той метод, який найкраще підходить в визначеній ситуації.

Відповідно до схеми методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів, наведеної в пункті 2.1, необхідно обрати та описати такі методи:

- метод для попередньої обробки тексту;
- метод для формування векторів семантично важливих слів;
- метод визначення емоційного забарвлення допису.

Оскільки, попередня обробка тексту означає видалення з тексту розділових знаків, то насамперед, метод попередньої обробки тексту, полягає в формуванні списку розділових знаків. З математичної точки зору, задачу видалення розділових знаків можна описати як різницю двох множин. За одну множину (A) приймається вхідний текст, за іншу (B) – список розділових знаків. Тож, різницею між цими двома множинами буде нова множина (C), яка містить усі елементи множини A, яких немає в множині B. Тобто, результатом проведеної операції буде текст, який не включатиме жодного розділового знаку із сформованого списку. Зміна регістру літер не потребує вибору конкретного методу, адже це можна легко реалізувати за допомогою вбудованих методів мови програмування PHP.

Для формування векторів семантично важливих слів використовується статистичний алгоритм TF-IDF описаний у пункті 1.3. Даний алгоритм дає змогу надати більшу вагу тим словам, які мають емоційне забарвлення. Значною перевагою обраного алгоритму є зниження важливості нейтральних та загальноживаних слів, адже ці слова не несуть жодного емоційного забарвлення. Такий підхід значно покращує результати класифікації текстів відповідно до емоційного забарвлення, адже більша увага приділяється саме емоційно забарвленим словам, а не словам, які частіше зустрічаються в тексті.

Основою методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом є метод класифікації. Класифікація належить до задач машинного навчання з вчителем. Навчання з вчителем передбачає наявність готових зразків з розміченими ознаками та категоріями. На основі підготованих зразків система навчатиметься за поданими ознаками розрізняти категорії. В даному випадку ознаками виступатимуть векторизовані тексти дописів, а категоріями – їхнє емоційне забарвлення.

Існує декілька алгоритмів для вирішення задачі класифікації, серед яких найпопулярнішими є: наївний Баєсів класифікатор, дерева рішень, логістична регресія, метод опорних векторів та найближчих сусідів. Для задачі визначення емоційного забарвлення дописів було прийнято рішення використати наївний Баєсів класифікатор. Перевагою даного алгоритму є те, що кожна ознака розглядається незалежно від інших ознак певного класу. Це значна перевага у випадку семантичного аналізу текстових даних, адже одне слово може зустрічатись як в негативних, так і в позитивних текстах.

Якщо значення однієї ознаки не впливає на значення інших ознак, то такі ознаки називаються незалежними. Але не можна припускати що всі ознаки з набору є повністю незалежними. Якщо розглядати текст, то більшість слів пов'язані одне з одним певним сенсом, і якщо розглядати їх окремо, то весь сенс може втратитись. Тому, теорема Баєса дозволяє передбачити клас на підставі набору ознак, використовуючи ймовірність.

Спрощене рівняння для класифікації за Баєсом виглядає так:

$$P(\text{Клас}A|\text{Ознака}1, \text{Ознака}2) = \frac{P(\text{Ознака}1|\text{Клас}A)*P(\text{Ознака}2|\text{Клас}A)*P(\text{Клас}A)}{P(\text{Ознака}1)*P(\text{Ознака}2)} \quad (2.1)$$

де P – ймовірність, $\text{Клас}A$ – певний клас A , $\text{Ознака}1$, $\text{Ознака}2$ – певні ознаки, які належать класу A .

За формулою (2.1) можна знайти ймовірність класифікації заданих даних до певного класу, на підставі ознак, які є в цих даних.

Баєсів класифікатор потребує навчання на заздалегідь розміченому наборі даних. Для класифікації нерозмічених даних потрібно обчислити ймовірність входження даних у кожен наявний клас та обрати найбільшу ймовірність. Попри свою простоту, даний метод може бути досить точним для визначення емоційного забарвлення тексту, якщо використовувати його в поєднанні з іншими алгоритмами (векторизація з використанням TF-IDF).

Приклад роботи Баєсового класифікатора можна побачити на рисунку 2.4.

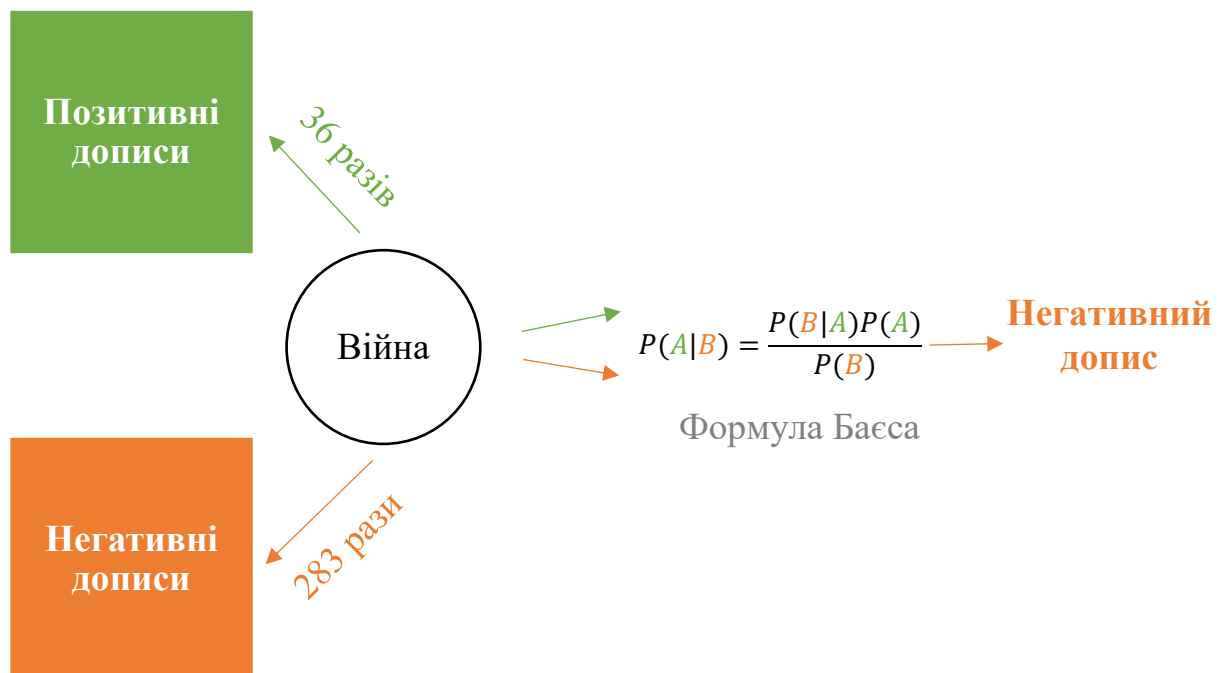


Рисунок 2.4 – Застосування Баєсового класифікатора для визначення емоційного забарвлення дописів

Як видно з рисунка 2.4, при використанні Баєсового класифікатора необхідно обчислити скільки разів задане слово входить до негативних та

позитивних дописів. Далі обчислюється ймовірність за формулою Баеса та видається результат.

Отже, для розробки методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів вирішено використовувати алгоритм TF-IDF для формування векторів семантично важливих слів та наївний Баєсів класифікатор для класифікації дописів за категоріями емоцій.

2.3 Формування навчальних множин дописів для аналізу емоційного забарвлення

Обраний метод класифікації, а саме, наївний Баєсів класифікатор, потребує навчання з вчителем для проведення подальшої класифікації дописів за емоційним забарвленням. Навчання з учителем означає необхідність створення навчальних множин, тобто, наборів ознак із заздалегідь визначеними категоріями.

Метою даної роботи є й прикладене застосування методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів, тому навчальними множинами будуть дописи користувачів із заздалегідь визначеним емоційним забарвленням.

Тематикою для створення соціально орієнтованого вебресурсу було обрано обговорення книг. Таким чином, створена інформаційна система буде мати вигляд блогу в якому користувачі матимуть змогу зареєструватись та публікувати дописи про книги на своїх особистих сторінках. Такі дописи дуже добре підійдуть для навчання класифікатора для задачі розпізнавання емоцій у тексті. Адже, зазвичай, відгуки та враження від прочитаного, за своєю природою, мають або позитивне, або негативне емоційне забарвлення. Також, такі дописи наповненні багатьма словами, які мають чітке емоційне навантаження, що допомагає краще визначити емоційне забарвлення тексту.

Зазвичай, для подібних задач використовуються готові датасети, які можна знайти в інтернеті. Такі датасети можуть бути як безкоштовні, так і

платні. Пошук готового датасету не дав бажаних результатів, тому було прийнято рішення створити набір необхідних даних власноруч. Для цього було використано чат-бот зі штучним інтелектом ChatGPT [26].

Для того щоб створити датасет з використанням ChatGPT необхідно написати чат-боту чіткий запит із вимогами до даних (Рисунок 2.5). Програма добре виконує поставленні перед нею завдання та генерує тексти позитивного та негативного емоційного забарвлення.

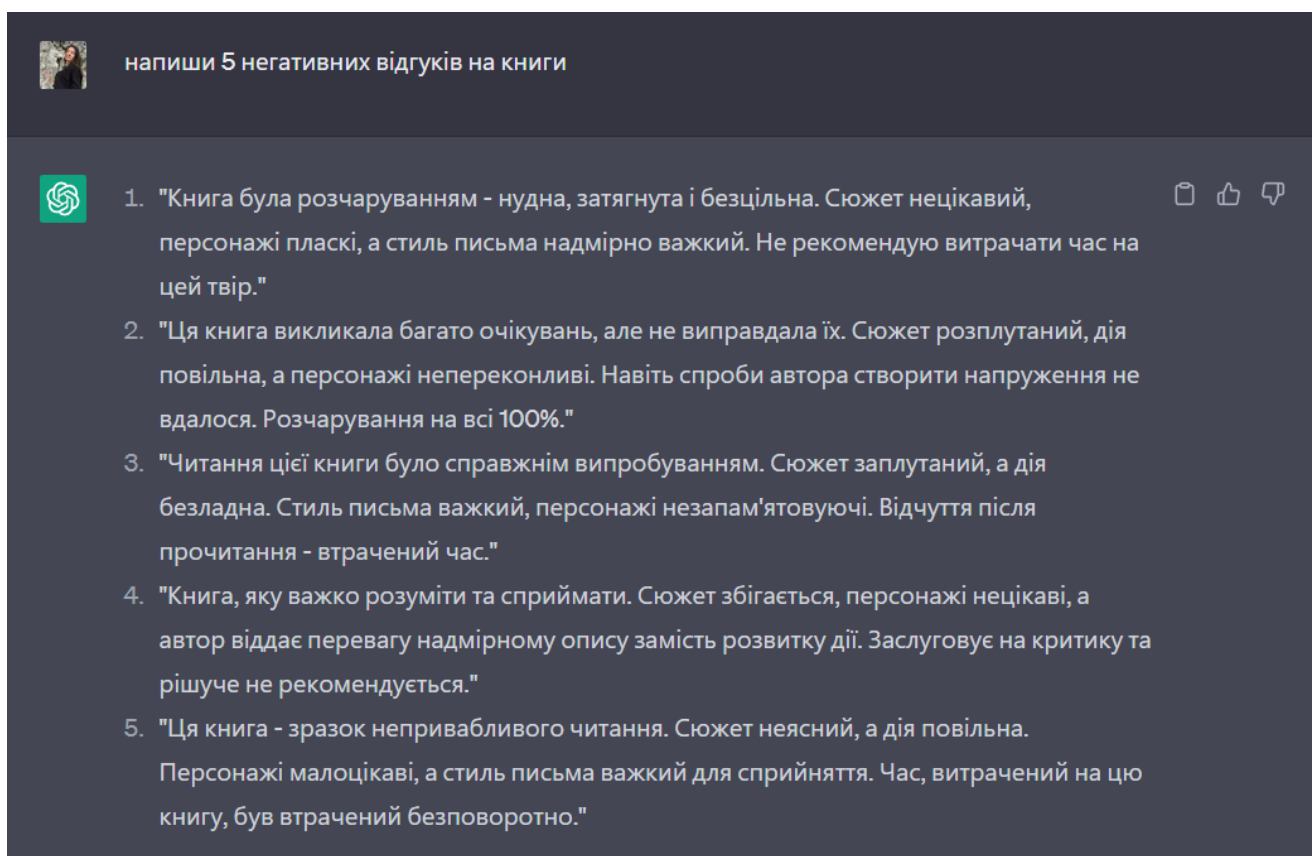


Рисунок 2.5 – Приклад створення негативних відгуків через ChatGPT [26]

Проте, інколи ChatGPT допускає помилки в написанні слів. Найпоширенішими помилками є неправильне відмінювання слів. Тому, кожен відгук, написаний ChatGPT, необхідно перечитувати та редагувати за потреби. З точки зору економії часу, це значно швидше ніж шукати реальні відгуки у соціально орієнтованих вебресурсах. Адже, в такому випадку, кожен реальний відгук потрібно самостійно класифікувати до певного емоційного класу, що не завжди легко та можливо.

Якість класифікації дописів відповідно до емоційного забарвлення залежить від кількості навчальних зразків на яких навчатиметься класифікатор. Для ефективного навчання класифікатора було створено 500 навчальних зразків (450 з яких для навчання класифікатора, а інші 50 – для тестування його роботи) (Рисунок 2.6). Як навчальні, так і тренувальні дані, порівну розділенні на дві категорії – позитивні та негативні відгуки.

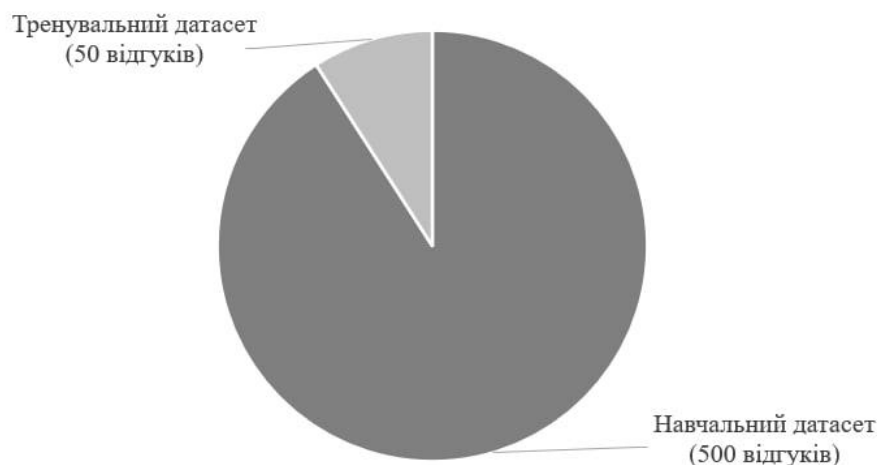


Рисунок 2.6 – Зразки робочих вхідних даних інформаційної системи

Отже, можна зробити висновок, що формування зразків робочих вхідних даних це дуже важливий етап перед початком програмної реалізації інформаційної системи. При самостійному створенні датасетів варто взяти до уваги відповідність створених даних усім вимогам. Інакше, в майбутньому, це може призвести до виникнення помилок та неправильної роботи програми.

2.4 Проектування структури соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів

Соціально орієнтований вебресурс з автоматизованим визначенням емоційного забарвлення дописів є програмною реалізацією, яка використовує розроблений метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом й забезпечує функції

аутентифікації користувачів, роботи користувача з персональними даними, додавання власних дописів, перегляду особистих сторінок інших користувачів, перегляду журналу обраного користувача з усіма дописами, визначення емоційного забарвлення обраного допису та візуалізації емоційно значущих слів.

Структура соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів складається з бази даних і двох підсистем (Рисунок 2.7):

- прикладна підсистема (соціально орієнтований вебресурс);
- підсистема визначення емоційного забарвлення дописів (використовує розроблений метод).

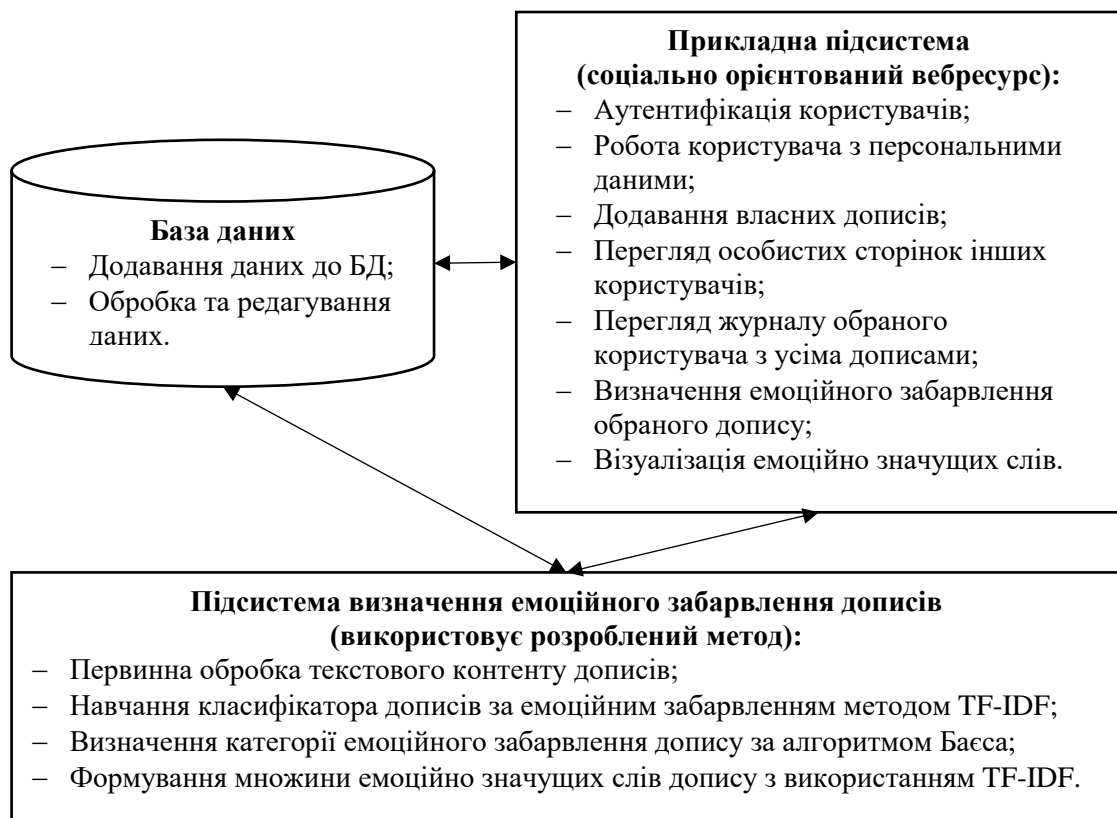


Рисунок 2.7 – Схема структури соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів

База даних потрібна для забезпечення можливості керувати даними, необхідними для функціонування соціально орієнтованого вебресурсу. Для цього потрібно реалізувати функціонал для додавання даних у БД, отримання

даних з БД, їх обробки та редагування. Створення БД є важливим етапом, адже будь-яка інформаційна система в процесі свого існування працює з даними, які потребують місця для зберігання.

Оскільки, прикладна реалізація методу передбачає створення соціально орієнтованого вебресурсу, який використовуватиме розроблений метод то важливо відокремити процеси пов'язані із роботою методу та роботою користувача із програмою розділивши весь функціонал на дві підсистеми. Це дасть змогу відокремити різні процеси, які відбуваються під час використання програмного продукту.

Функціонал, пов'язаний із вирішенням задачі автоматизованого визначення емоційного забарвлення дописів, зосереджено в підсистемі визначення емоційного забарвлення дописів. Дана підсистема використовує розроблений в п. 2.1 метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом.

Перш за все, підсистема для визначення емоційного забарвлення дописів аналізуватиме текстовий контент множин навчальних дописів та дописів, розміщених користувачами у соціально орієнтованому вебресурсі. Під аналізом текстового контенту дописів мається на увазі його первинна обробка, тобто видалення розділових знаків та зміна регістру. Після цього, відбуватиметься навчання класифікатора з використанням методу TF-IDF, який формуватиме вектори семантично важливих слів за категоріями емоцій. Визначення категорії емоційного забарвлення допису відбувається через використання алгоритму Баєса. Коли емоційне забарвлення допису буде визначено, система додаватиме отримані дані у БД. Пізніше ці дані будуть використані для відображення емоційного забарвлення дописів у зручній та зрозумілій для користувача формі.

Оскільки, метою прикладної реалізації розробленого методу є створення соціально орієнтованого вебресурсу то потрібно спроектувати необхідний функціонал для користувачів. Саме тому було прийнято рішення об'єднати весь функціонал пов'язаний із взаємодією користувача із програмою в одну прикладу підсистему.

Перш за все, для користувачів необхідно реалізувати систему аутентифікації, тобто, можливість реєстрації на вебресурсі та збереження особистих даних для подальшого повторного входу у систему. Також, це зумовлює реалізацію функціоналу, який дозволить користувачеві працювати з особистими даними, тобто додавати їх, редагувати або, взагалі, видаляти.

Використання соціально орієнтованих вебресурсів передбачає можливість ведення особистої сторінки для зареєстрованих користувачів. На особистій сторінці користувачі можуть ділитись своїми фотографіями, думками та іншою інформацією. Також, варто відмітити, що ведення особистої сторінки не можливе без реалізації функціоналу, який дозволяє додавання дописів.

Оскільки, метою існування соціально орієнтованих вебресурсів є обмін інформацією між користувачами через публікацію дописів, то важливо, передбачити можливість перегляду сторінок інших користувачів. Такий підхід дозволяє користувачам знаходити однодумців або просто цікавих людей. Перегляд сторінок інших користувачів дає змогу переглянути їхні дописи та надихнутись новими ідеями або дізнатись щось нове.

Згідно з тим, що створений вебресурс використовує розроблений метод, то прикладна підсистема містить в собі функціонал, який відповідає за візуалізацію результатів роботи методу. Мається на увазі, що кожен допис у вебресурсі має визначене емоційне забарвлення, яке відображається для користувачів у зрозумілій формі (кольорова рамка, напис). Також, крім визначеного емоційного забарвлення користувачі можуть переглянути перелік ключових слів, які мають емоційне забарвлення.

Отже, прикладна підсистема в першу чергу відповідає за взаємодію користувачів із інформаційною системою. Також, ця підсистема взаємодіє із БД, адже усі відображені користувачеві дані зберігаються у базі даних. Також, ця підсистема взаємодіє з підсистемою для визначення емоційного забарвлення дописів, оскільки, кожен доданий користувачем допис одразу після публікації проходить процес визначення його емоційного забарвлення.

Основною сторінкою для користувачів соціально орієнтованих вебресурсів є їхня персональна сторінка (Рисунок 2.8). На цій сторінці користувач може переглянути свою особисту інформацію та відредагувати її за потреби. Також, саме на цій сторінці реалізовано інтерфейс, який дозволяє користувачеві додавати дописи. Звичайно ж, на персональній сторінці користувачам доступні написані ними дописи. Це дає змогу переглядати додаткову інформацію про дописи, просто натиснувши на бажаний допис.

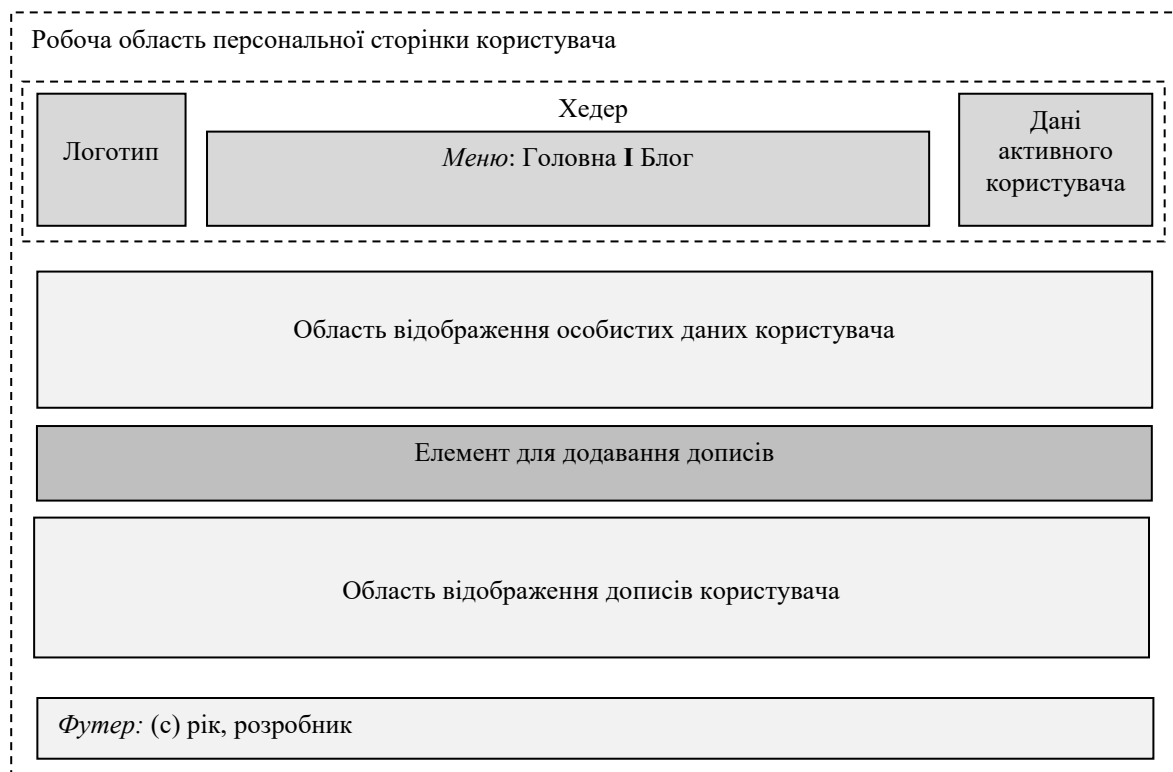


Рисунок 2.8 – Схема інтерфейсу персональної сторінки зареєстрованого користувача

Прикладне застосування методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом можна переглянути на сторінці обраного допису (Рисунок 2.9). Для зручності та кращого розуміння користувач бачитиме до якої категорії емоцій відноситься обраний допис, а також, зможе переглянути слова допису, які мають емоційне забарвлення.

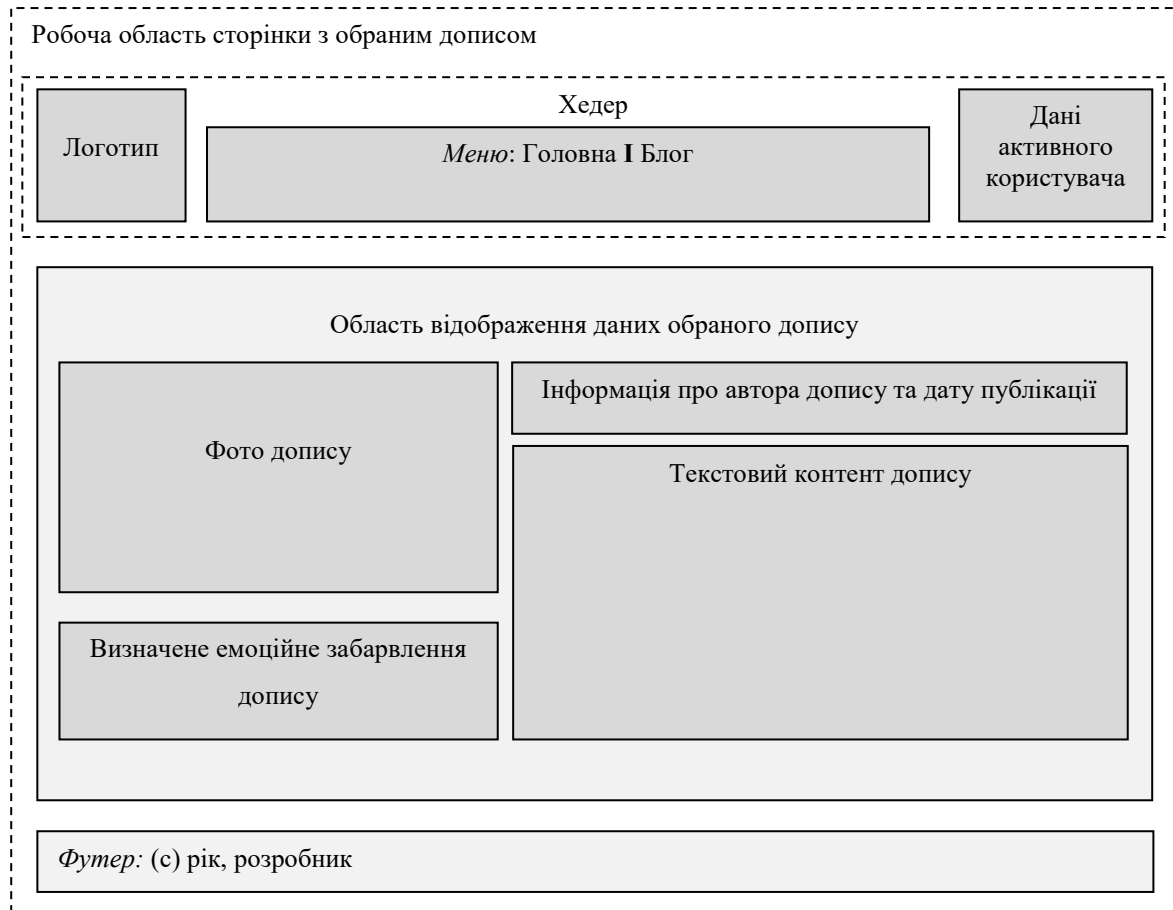


Рисунок 2.9 – Схема інтерфейсу сторінки обраного допису

На етапі проєктування схеми структури інформаційної системи важливо чітко розділити весь функціонал на окремі підсистеми. Це допоможе уникнути плутанини та дозволить краще зрозуміти процеси, які відбуваються всередині інформаційної системи. Також, це дає змогу побачити, як окремі компоненти програми пов'язані між собою.

Отже, спроектовано соціально орієнтований вебресурс з автоматизованим визначенням емоційного забарвлення дописів, який використовує розроблений метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом й забезпечує функції аутентифікації користувачів, роботи користувача з персональними даними, додавання власних дописів, перегляду особистих сторінок інших користувачів, перегляду журналу обраного користувача з усіма дописами, визначення емоційного забарвлення обраного допису та візуалізації емоційно значущих слів.

2.5 Даталогічна модель бази даних соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів

Проектування даталогічної моделі бази даних є чи не найважливішим етапом в розробці програмних продуктів. Адже, саме БД зберігає усі необхідні дані, які використовуються під час роботи програми. Правильно спроектована БД дозволяє з легкістю зберігати, отримувати та змінювати дані, які в ній зберігаються. Дотримання правил розробки баз даних зменшить ризик втрати даних чи їх некоректного використання.

Даталогічна модель бази даних соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів можна розділити на дві частини. Перша частина зберігає дані для навчання класифікатора, а друга – дані, необхідні для функціонування соціально орієнтованого вебресурсу. В загальному БД складається з восьми таблиць: Users, User type, Sex, Comments, Likes, Posts, Emotions, Posts for learning (Рисунок 2.10).

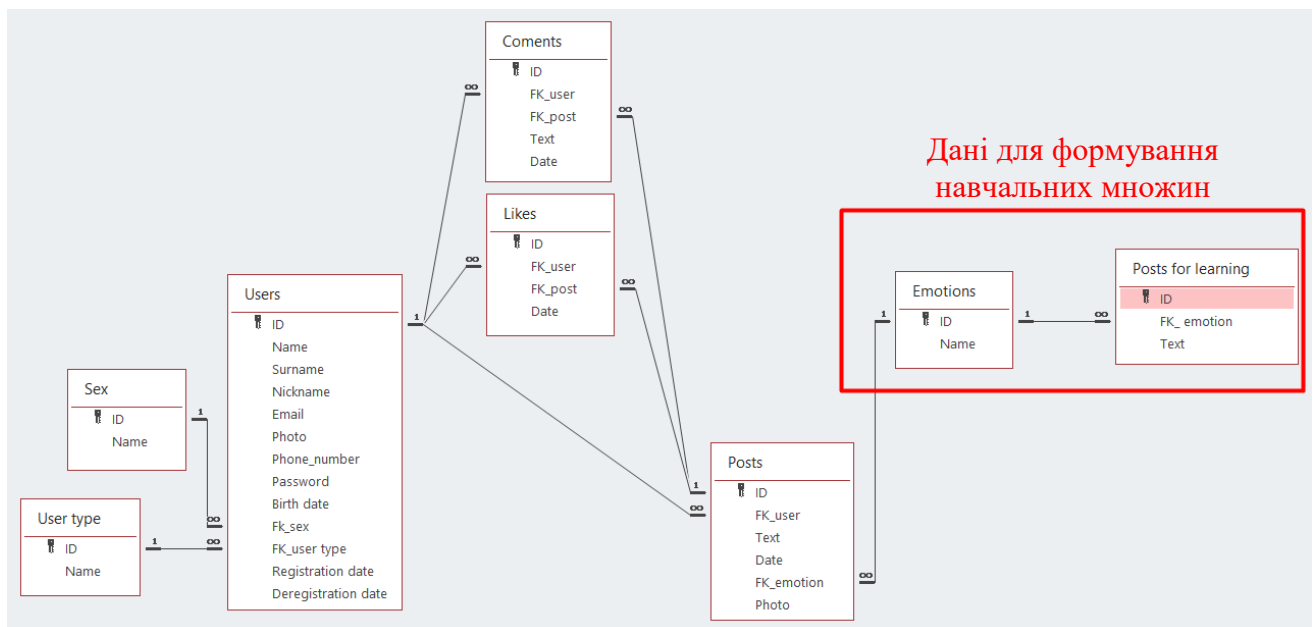


Рисунок 2.10 – Даталогічна модель БД соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів

Головною таблицею у схемі БД є таблиця «Posts for learning». У ній зберігається інформація про дописи, які складатимуть вибірку навчальних даних для класифікатора. Оскільки, навчальні дані повинні мати заздалегідь визначене емоційне забарвлення кожного допису, то ця таблиця також зберігатиме дані про емоційні забарвлення дописів. Отже, таблиця «Posts» має такі атрибути: ID, FK_emotion, Text (Таблиця 2.1).

Таблиця 2.1 – Атрибути таблиці «Posts»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ.
2.	FK_emotion	int	Вторинний ключ. Використовується для позначення емоційного забарвлення допису.
3.	Text	text	Текстовий вміст допису.

Зважаючи на те, що основною метою створення соціально орієнтованого вебресурсу є автоматизоване визначення емоційного забарвлення дописів, доцільним є створення окремої таблиці, яка зберігатиме існуючі емоційні забарвлення. Для досягнення цієї мети, буде створено таблицю «Emotions» з такими атрибутами: ID, Name (Таблиця 2.2).

Таблиця 2.2 – Атрибути таблиці «Emotions»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ.
2.	Name	text	Назва емоційного забарвлення

Оскільки, створена інформаційна система матиме вигляд соціально орієнтованого вебресурсу, важливою функцією якого є система аутентифікації користувачів, є необхідність у створенні таблиці для збереження даних користувачів. Тому, було спроектовано структуру таблиці «Users» з такими атрибутами: ID, Name, Surname, Nickname, Email, Avatar, Password, FK_sex, FK_user_type, Registration_date, Deregistration_date (Таблиця 2.3).

Таблиця 2.3 – Атрибути таблиці «Users»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ.
2.	Name	text	Особисте ім'я користувача
3.	Surname	text	Особисте прізвище користувача
4.	Nickname	text	Унікальний нікнейм користувача
5.	Email	text	Поштова скринька користувача
6.	Avatar	text	Фото користувача (аватар)
7.	Password	text	Унікальний пароль придуманий користувачем
8.	FK_sex	int	Вторинний ключ. Використовується для позначення статі користувача.
9.	FK_user_type	int	Вторинний ключ. Використовується для позначення типу користувача.
10.	Registration_date	date	Дата реєстрації користувача на вебресурсі.
11.	Deregistration_date	date	Дата видалення профілю користувача на вебресурсі.

Ще однією важливою таблицею з БД є таблиця «Posts». Ця таблиця призначена для зберігання інформації про дописи користувачів. Адже, написання дописів користувачами та їхня публікація – це одна із задач по створенню соціально орієнтованого вебресурсу. Таким чином, таблиця «Posts» має такий набір атрибутів: ID, FK_user, Date, Text, FK_emotion, Photo (Таблиця 2.4).

Таблиця 2.4 – Атрибути таблиці «Posts»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ.
2.	FK_user	int	Вторинний ключ. Використовується для позначення автора допису.
3.	Date	date	Дата публікації допису.
4.	Text	text	Текстовий контент допису
5.	FK_emotion	int	Вторинний ключ. Використовується для позначення емоційного забарвлення допису.
6.	Photo	text	Фото допису

Часто, в соціально орієнтованих вебресурсах, для користувачів передбачена можливість взаємодії із дописами інших користувачів. Це здійснюється через написання коментарів або реакцій на дописи, які називають «лайками». Тому, для забезпечення цього функціоналу спроектовано дві додаткові таблиці «Likes» та «Comments» для лайків та коментарів відповідно.

Під час проектування даталогічної моделі БД виникає краще розуміння, які сутності необхідні для створення програмного продукту, та які існують зв'язки між ними. Детальний опис кожної сутності дає змогу зрозуміти, які процеси відбуваються при обробці даних. Також, це зменшує ризик виникнення непередбачуваних проблем та помилок на етапі наповнення БД даними.

Отже, було спроектовано даталогічну модель бази даних соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів, яка складається з восьми таблиць, що зберігають усю необхідну інформацію про користувачів та написані ними дописи. Крім цього, у БД зберігаються навчальні дописи, необхідні для роботи методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом.

2.6 Вибір засобів розробки соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів

Для створення будь-якого програмного продукту важливо правильно обрати набір необхідних засобів розробки. Це допоможе пришвидшити реалізацію програмного застосунку та уникнути небажаних проблем на різних етапах розробки програми.

Оскільки, поставлене завдання передбачає створення веб сервісу, то для його реалізації було прийнято рішення використати фреймворк Laravel для мови програмування PHP. Обраний фреймворк є одним із найкращих засобів для розробки динамічних та інтерактивних вебсайтів [27]. Щодо мови

програмування PHP, можна сказати, що це одна з популярних мов у сфері веброзробки. Дана мова об'єктно орієнтована та досить проста у вивченні та розумінні.

Зважаючи на те, що розробка вебресурсу потребує збереження великої кількості даних, необхідно обрати засіб для створення та збереження БД. Одним із лідерів серед безплатних варіантів збереження та обробки даних є СКБД MySQL. Обрана система гнучка та досить проста у використанні та налаштуванні під власні потреби. MySQL дозволяє створювати бази даних із необмеженою кількістю взаємопов'язаних сутностей. Також, обрану СКБД легко можна під'єднати до проєкту, створеного на платформі PHP. СКБД MySQL має досить високу продуктивність, що дозволяє швидко виконувати необхідні операції над даними, збереженими в БД [28].

Проаналізувавши обрані засоби розробки можна зробити висновок, що поєднання платформи PHP із СКБД MySQL надає усі необхідні інструменти для реалізації соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів.

2.7 Висновки до розділу 2

Було розроблено метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів, який призначений для визначення категорії допису за його емоційним забарвленням та переліку емоційно значущих слів допису. Виконання розробленого методу поділяється на два етапи. На першому етапі методу формуються вектори семантично важливих слів за категоріями емоцій на основі завчасно підготованих тестових зразків з використанням методу TF-IDF. На другому етапі проводиться визначення емоційного забарвлення допису використовуючи алгоритм Баєса. Крім цього, для кращого розуміння визначеної методом категорії емоційного забарвлення, формується перелік емоційно значущих слів.

Для навчання класифікатора було необхідно заздалегідь сформувати множини навчальних даних, які включатимуть в себе позитивні та негативні відгуки. Формування навчальних даних відбувається з використанням чат-боту зі штучним інтелектом ChatGPT, який може генерувати текст відповідно до заданого запиту.

Соціально орієнтований вебресурс з автоматизованим визначенням емоційного забарвлення дописів є прикладною програмною реалізацією методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом. Спроектований вебресурс забезпечуватиме функції аутентифікації користувачів, роботи користувача з персональними даними, додавання власних дописів, перегляду особистих сторінок інших користувачів, перегляду журналу обраного користувача з усіма дописами, визначення емоційного забарвлення обраного допису та візуалізації емоційно значущих слів.

Наявність великої кількості даних, потрібних для роботи системи, призводить до виникнення необхідності у створенні БД, для збереження цих даних. Тому, в ході проектування було спроектовано даталогічну модель бази даних соціально орієнтованого вебресурсу, яка складається з восьми таблиць, що зберігають усю необхідну інформацію про користувачів та написані ними дописи.

Для реалізації БД відповідно до спроектованої даталогічної моделі бази даних було прийнято рішення використати СКБД MySQL. У поєднанні із PHP фреймворком Laravel, обраний набір засобів розробки, дає змогу використовувати велику кількість інструментів, необхідних для розробки вебсервісу.

Розділ 3 Програмна реалізація соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів

3.1 Структура модулів соціально орієнтованого вебресурсу

При реалізації вебресурсу використовується фреймворк Laravel, який передбачає створення програмних продуктів використовуючи шаблон MVC. Використання шаблону MVC, під час створення програмних продуктів, дозволяє відокремити інтерфейс користувача від програмної логіки застосунку. MVC складається з трьох компонентів, кожен з яких відповідає за певні процеси, які відбуваються в інформаційній системі. Models або моделі відповідають за отримання даних із БД. Views або представлення – це файли, які зберігають структуру сторінок вебсайту, тобто – це інтерфейс користувача. Controllers або контролери – це свого роду посередники між моделями та представленнями, які отримують дані із моделей, виконують над ними необхідні операції та передають їх в представлення для відображення користувачеві [29].

Відповідно до шаблону проєктування MVC, та спроектованої в п. 2.5 даталогічної моделі БД в проєкті створено такі моделі:

- Dataset – модель, яка отримує вибірку навчальних датасетів, необхідних для формування навчальних множин;
- Emotion – модель, яка отримує інформацію про емоційне забарвлення дописів із навчальних множин та дописів опублікованих користувачами;
- Post – модель для отримання даних про дописи користувачів;
- User – модель з допомогою якої можна отримати дані про користувачів, зареєстрованих на вебресурсі;
- UserType – модель для отримання даних про типи користувачів.

Для реалізації інтерфейсу користувача, згідно із шаблоном MVC, необхідно створити окремі файли з розміткою сторінок. Розмітка сторінок створюється з використанням звичних HTML тегів, в поєднанні із методами РНР, які дають змогу відображати користувачам дані. Отже, для реалізації зручного інтерфейсу в проєкті було створено такі файли представлень:

- AuthLogin – інтерфейс сторінки для авторизації зареєстрованих на вебресурсі користувачів;
- AuthRegister – інтерфейс сторінки для реєстрації нових користувачів;
- LayoutsMain – загальна розмітка сторінок, яка включає шапку вебсайту та підвал, які є однаковими на усіх інших сторінках;
- PostsIndex – інтерфейс для стрічки дописів;
- PostsShow – розмітка для сторінки з відображенням обраного користувачем допису;
- Home – особиста сторінка авторизованого користувача, яка включає його особисту інформацію та опубліковані дописи;
- Index – головна сторінка вебресурсу з якої починається взаємодія користувача із вебсайтом.

Реалізація основного функціоналу програми зберігається у файлах контролерів. Тому, для взаємодії із даними вебресурсу було створено такі контролери:

- HomeController – контролер для роботи із головною сторінкою вебресурсу;
- PostsController – контролер, який відповідає за збереження даних про опубліковані користувачами дописи до БД, а також, відображення цих дописів;
- RegisterController – контролер для керування процесом реєстрації нових користувачів на вебресурсі;
- LoginController – контролер, який відповідає за авторизацію зареєстрованих користувачів на вебресурсі.

Отже, відповідно до шаблону MVC користувач взаємодіє через інтерфейс користувача відправляючи запити контролерам. До цих запитів відноситься автентифікація, перегляд стрічки дописів та особистої сторінки, а також, додавання нових дописів та перегляд обраного допису. Після того, як користувач створить запит, який його цікавить, контролер передає запит на отримання даних в відповідну модель. Лише модель може взаємодіяти з БД, тому після отримання потрібних даних, модель передає їх контролеру. В даному випадку даними, які

можна отримати через модель може бути інформація про користувачів та дописи. В контролері відбувається маніпуляція з отриманими даними, наприклад, сортування дописів за датою. Далі оброблені дані передаються в представлення та відображаються в інтерфейсі користувача. Описану структуру модулів системи та взаємодію користувача із системою зображено на рисунку 3.1.

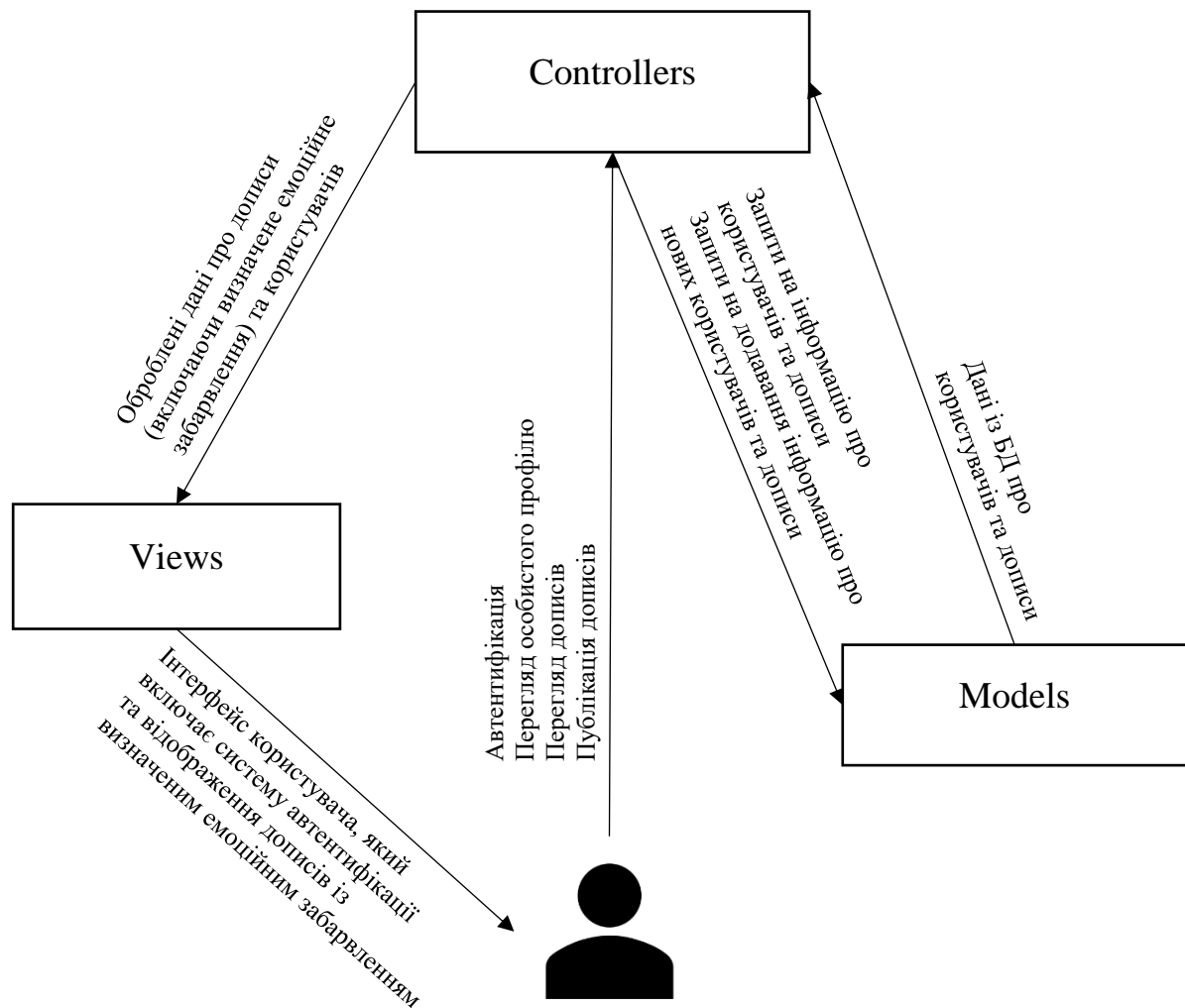


Рисунок 3.1 – Структура модулів соціально орієнтованого вебресурсу

Якщо інформаційна система потребує реалізацію додаткових методів, які потрібно використовувати в різних частинах програми, варто використовувати допоміжні інструменти – помічники (Helpers). Helpers – це окремі файли з програмним кодом, в яких реалізований додатковий функціонал. Під час

реалізації методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом, було прийнято рішення винести програмний код методу в окремий файл помічник. Таким чином, це дасть змогу викликати метод визначення емоційного забарвлення тексту у будь-якій точці програми.

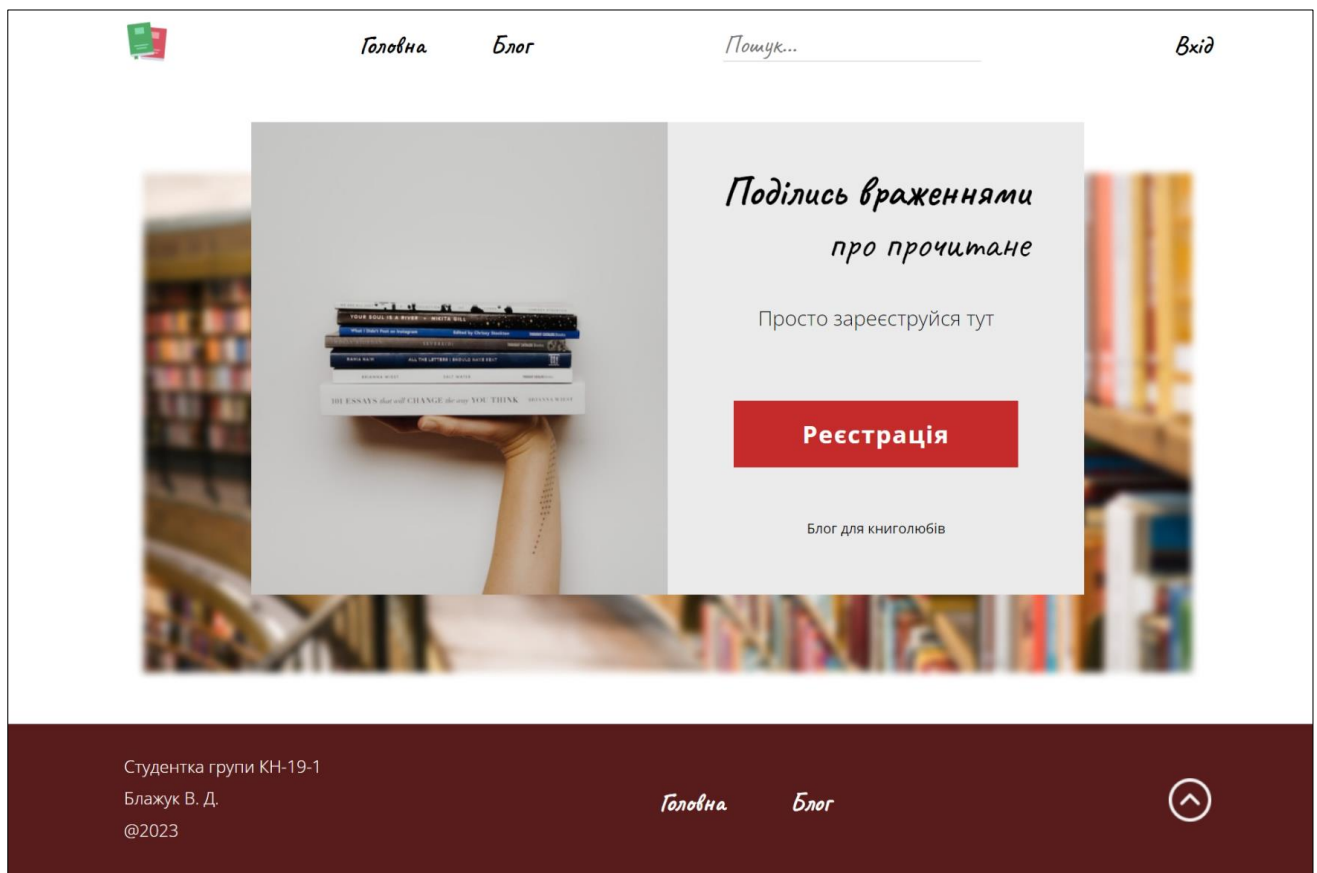
Реалізована структура модулів соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів відповідно до шаблону MVC, дозволяє розділити різні компоненти програми на окремі файли. Це значно пришвидшує розробку та підтримку програмного продукту. Тому, згідно з шаблоном MVC функціонал створеної інформаційної системи було розділено на п'ять моделей, сім представлень, чотири контролери та один файл помічник із реалізацією методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом.

3.2 Особливості реалізації соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів

Знайомство користувачів із будь-яким програмним продуктом завжди починається із головної сторінки на якій присутня уся основна інформація про ресурс. Оскільки, темою соціально орієнтованого вебресурсу було обрано обговорення книг, то програмний продукт матиме вигляд блогу для книголюбів (Рисунок 3.2). Вдало реалізована головна сторінка дає змогу користувачам з перших секунд зрозуміти з чим вони матимуть справу. Саме тому, на головній сторінці прийнято розміщувати загальну ввідну інформацію, яка буде корисною для нових користувачів вебресурсу.

Оскільки, завдання кваліфікаційної роботи бакалавра полягає в реалізації соціально орієнтованого вебресурсу, тому важливим етапом розробки є створення системи автентифікації. Для цього у фреймворку Laravel використовується бібліотека UI, яка автоматично створює набір потрібних контролерів та представлень. Потрібно лише відредагувати ці файли під власні

потреби. В результаті редагування файлу з розміткою сторінки для реєстрації ця сторінка матиме вигляд зображений на рисунку 3.3.



Рисунк 3.2 – Головна сторінка блогу для книголюбів

Для реєстрації нового користувача, перш за все, потрібно перевірити чи користувач не залишив пустих полів, адже при внесенні даних до БД важливо мати повний перелік даних. Якщо користувач випадково залишить поле пустим воно підсвітиться червоним кольором. Так буде продовжуватись доти, доки користувач не заповнить усі поля.

Коли користувач заповнить усі поля, необхідно перевірити введені користувачем дані, перш ніж додавати їх до БД. Для цього створено функцію `validator()`, яка перевіряє відповідність даних вимогам. При виникненні помилок функція повертає текст помилки у відповідне поле. Таким чином, користувач бачитиме де саме він допустив помилку і чому.

Рисунок 3.3 – Вигляд форми для реєстрації нових користувачів

Якщо дані введено вірно, функція `create()` створює в базі даних новий запис з інформацією про користувача. Оскільки, на даному етапі можуть виникнути проблеми пов'язані із доступом до БД то для інформування користувача про їхнє виникнення створено окрему сторінку, яка відкривається у випадку помилки при додаванні даних до бази. Описаний алгоритм реалізовано в контролері `RegisterController` та проілюстровано за допомогою блок-схеми (Рисунок 3.4).

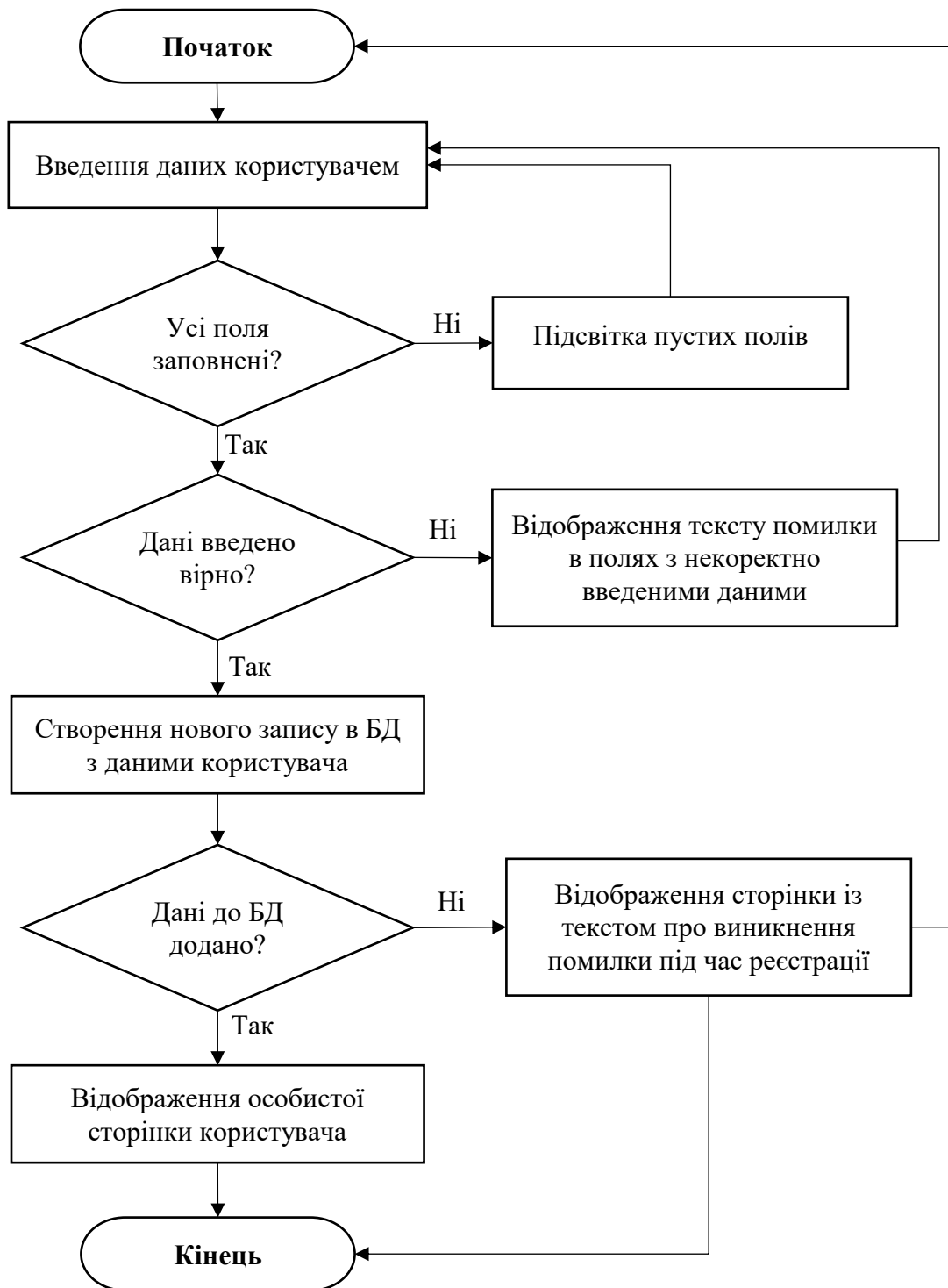


Рисунок 3.4 – Схема виконання алгоритму реєстрації нового користувача

Після того як користувач зареєструється в системі йому буде доступний його особистий профіль (Рисунок 3.5). Тут користувачу доступна персональна інформація, введена під час реєстрації.

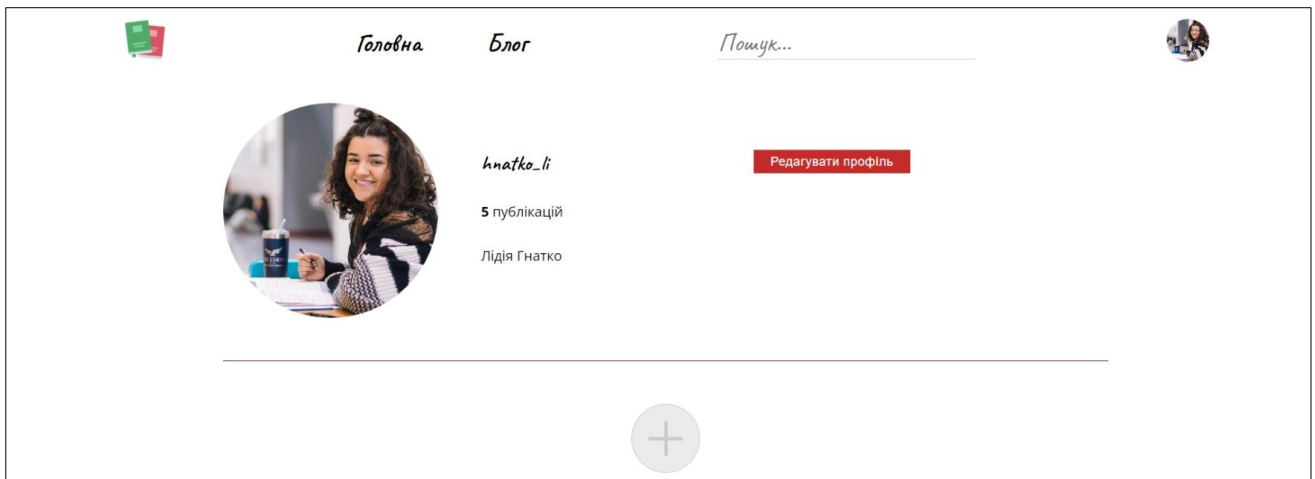


Рисунок 3.5 – Особиста сторінка зареєстрованого користувача

Основна мета існування соціально орієнтованих вебресурсів – це наповнення вебресурсу контентом від користувачів. Тому, при розробці соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів, важливо реалізувати можливість публікування дописів для авторизованих користувачів.

Для початку, користувачу потрібно ввести текст допису у відповідне поле, зберігаючи усі абзаци та форматування. Це необхідно для подальшого коректного відображення тексту допису на сторінці. Далі, текст проходить попередню перевірку, яка перевіряє чи розмір тексту не занадто великий та чи зазначене поле не пусте. Якщо умови перевірки не виконуються, користувачу відображається текст помилки у полі для введення тексту.

Перед створенням нового запису у БД відбувається аналіз тексту з використанням реалізованого методу визначення емоційного забарвлення дописів. В результаті цієї операції визначається категорія до якої належить текст, відповідно до його емоційного забарвлення. Також, формується список емоційно значущих слів допису. Уся зазначена інформація зберігається у БД разом із текстовим контентом допису та інформацією про його автора. Функціонал, який дозволяє користувачам публікувати нові дописи, реалізовано в контролері PostsController та проілюстровано за допомогою блок-схеми (Рисунок 3.6).

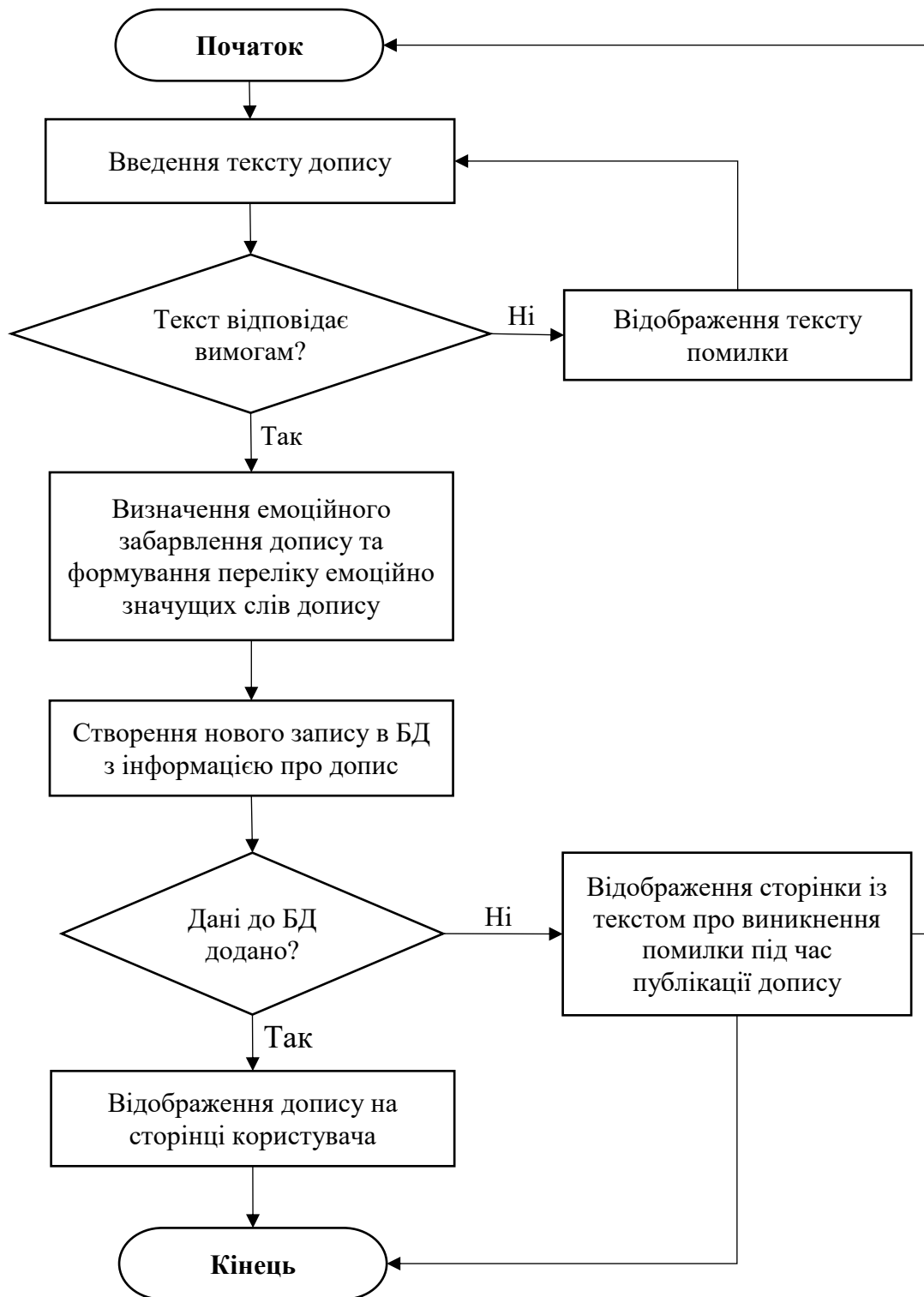


Рисунок 3.6 – Схема виконання алгоритму публікації допису

Якщо при публікації допису не виникло проблем та помилок то новий допис буде автоматично відображено на особистій сторінці користувача (Рисунок 3.7). Визначене емоційне забарвлення дописів відображається через кольорові межі допису. Зеленим кольором виділяються позитивні дописи, а червоним – негативні. Виділення емоційного забарвлення дописів кольором

відбувається шляхом перевірки даних про кожен допис, які зберігаються в БД. Таким чином, користувач має змогу побачити прикладне застосування розробленого методу визначення емоційного забарвлення дописів на прикладі власної сторінки.

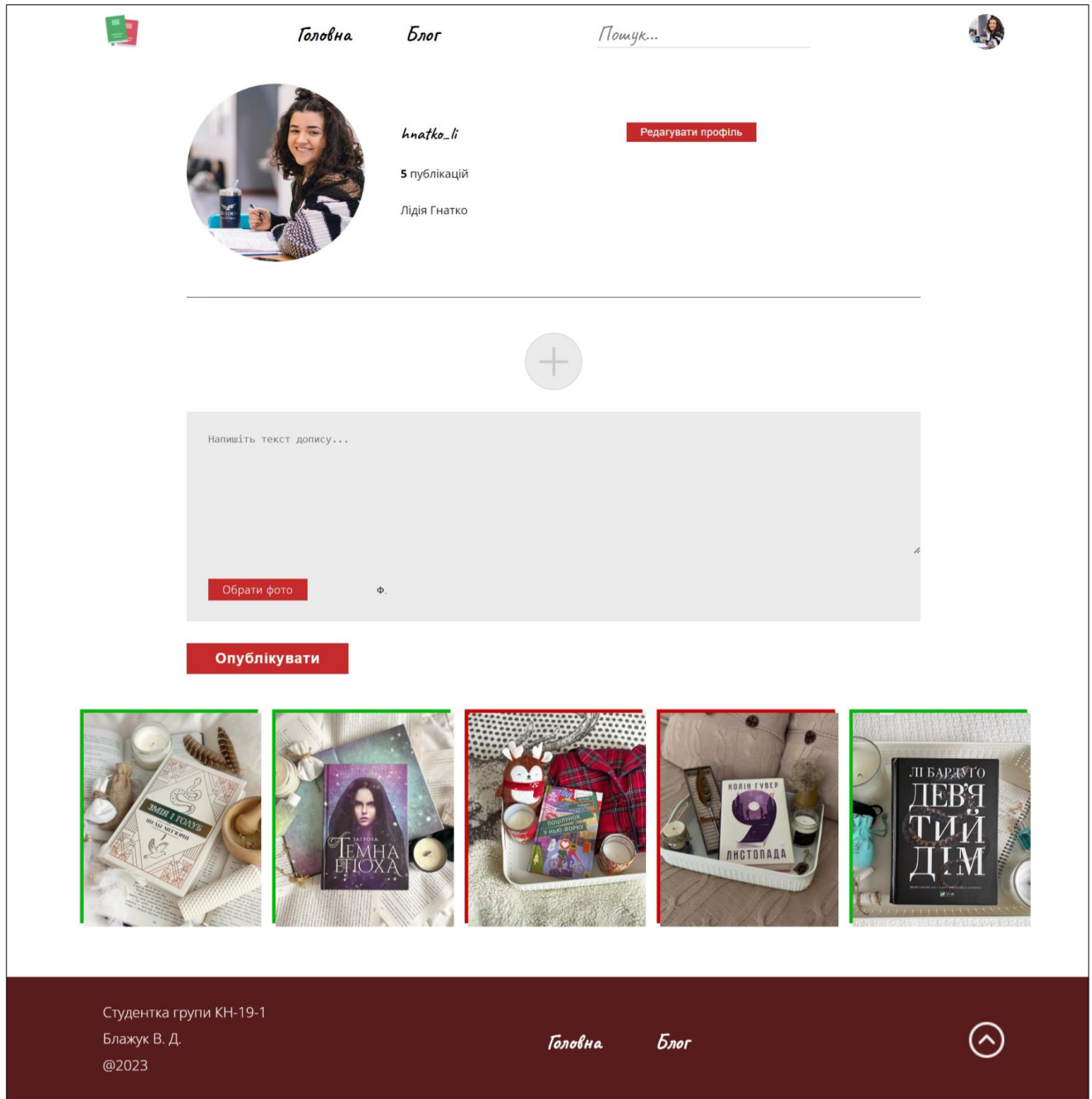


Рисунок 3.7 – Особиста сторінка користувача з опублікованими дописами

Визначення емоційного забарвлення дописів відбувається з використанням PHP бібліотеки PHP-ML [30]. Дана бібліотека дозволяє

використовувати різні алгоритми для вирішення задач класифікації та навчання нейронних мереж.


Відповідно до розробленого у 2 розділі методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом, спершу потрібно видалити розділові знаки із тексту та перетворити весь текст у символи нижнього регістру. Це реалізовано через цикл, який перебирає весь текст.

Після цього, оброблений текст передається в клас `TfidfTransformer` (один із класів бібліотеки `RNP-ML`). В даному класі рахується емоційне навантаження кожного слова, відповідно до кількості його появ в дописах кожної категорії. Тобто, якщо певне слово частіше з'являється в позитивних дописах, то воно матиме більшу позитивну вагу. Аналогічно зі словами, які частіше зустрічаються в негативних дописах. Щодо слів, які зустрічаються як в негативних, так і в позитивних дописах приблизно однакову кількість разів, то такі слова матимуть нейтральну вагу. Саме так відсіюватимуться усі загально вживані слова та слова, які не мають емоційного забарвлення. Таким чином, використання класу `TfidfTransformer` дозволяє сформувати вектори семантично важливих слів за категоріями.


Для визначення емоційного забарвлення дописів використовується інший клас бібліотеки `RNP-ML` – клас `NaiveBayes`, в якому містяться методи для роботи із класифікатором Баєса. Оскільки, класифікатор не потребує навчання, адже замість простого підрахунку кількості появи слів у дописах були сформовані вектори з визначеним емоційним забарвленням слів, то для визначення емоційного забарвлення дописів використовується метод `predict()`, який передбачає емоційне забарвлення допису.

Результат виконання методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом зображено на рисунку 3.8. Емоційне забарвлення допису визначається з використанням наївного Баєсового класифікатора. Також, для кращого розуміння чому допис віднесено до того, чи іншого класу емоційного забарвлення, виділяються слова з

відповідним емоційним навантаженням. Визначення слів, які потрібно виділити відбувається з використанням раніше сформованих векторів семантично важливих слів. Тобто, слова які мають позитивну вагу виділяються зеленим кольором, а слова з негативною вагою – червоним.



Емоційне забарвлення допису
Позитивне

 **hnatko_li** 2023-04-25 16:46:08

Я **дуже** **дуже** **люблю** цю книгу. І не лише тому, що головну героїню звати так само, як мене, а тому що вона дійсно була **дуже** **цікава**. Книга буквально з перших сторінок затягує у свої переплетіння.

Персонажі
Луїза та Рід це просто вогонь. Вони **настільки** **прикольні** разом. Їхня хімія була **дуже** **класною**, навіть попри обставини їхньої зустрічі, було чимало **милих** моментів.
Лу це дійсно одна з **улюблених** героїнь, тому що настільки було **весело** спостерігати за тим, як Рід реагує на її фразочки, пісеньки, підколи. Вона **весела**, відкрита, відважна, **емоційна** та з характером, в той час як Рід стриманий, набожний, не надто емоційний, серйозний, цей контраст між ними був **чудовий**. Вони доповнювали один одного. Ну і звісно розвиток їхніх стосунків від ненависті до кохання, що просто не могло не радувати.

Сюжет
Він дійсно **захоплює**. Сюжетні повороти просто **шикарні** та не дають відірватись від книги. Коли здається, що нас більше вже не здивувати, авторці все ж вдається це знову зробити. Книга закінчилась **дуже** **круто**, **емоційно**, **епічно**, але **круто**.

Отже якщо вагаєтесь, чи братись за цю книгу, я без вагань вам скажу, читати її. Вона дійсно **хороша** та **варта уваги**. Тепер з **нетерпінням** очікуєм на наступні частини цієї трилогії.




Рисунок 3.8 – Результат визначення емоційного забарвлення допису

Використання готових бібліотек значно пришвидшує та полегшує розробку програмних продуктів. Бібліотека RHP-ML надає великий набір класів, які можна використовувати при вирішенні задач, пов'язаних із аналізом цифрових текстів. Отже, в результаті використання бібліотеки RHP-ML було реалізовано прикладне застосування методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом, що дозволяє за текстовим контентом допису визначити його емоційне забарвлення та виділити ключові слова, які мають емоційне навантаження.

3.3 Тестування функціональних можливостей соціально орієнтованого вебресурсу

Зазвичай, для того щоб мати змогу публікувати дописи в соціально орієнтованих вебресурсах існує система аутентифікації. Це необхідно для того, щоб кожен користувач мав свою особисту сторінку на якій будуть відображатись усі написані ним дописи. Також, це необхідно для збереження інформації про авторів дописів.

Отже, для того щоб отримати доступ до можливості публікувати дописи, користувачу необхідно пройти процес реєстрації. Для того щоб зареєструватись користувач повинен натиснути на кнопку «Реєстрація», яка знаходиться на головній сторінці вебресурсу. Далі, користувач повинен ввести персональні дані у відповідні поля та натиснути на кнопку «Зареєструватись» (Рисунок 3.9).

Головна Блог Пошук... Вхід

Реєстрація

Вікторія

Блажук

ttori

ttori@gmail.com

.....

.....

Вибір файлу photo.jif

Зареєструватись

Студентка групи КН-19-1
Блажук В. Д.
@2023

Головна Блог

Рисунок 3.9 – Процес реєстрації користувача на вебресурсі

Для перевірки процесу реєстрації створено тест-кейс (Таблиця 3.1). Тест-кейс перевіряє правильність введеної користувачем інформації. Це дає змогу відстежити чи вірно працює система реєстрації, та знайти помилки, якщо такі є.

Таблиця 3.1 – Тест-кейс АТ0001

Тест-кейс ID: АТ0001	Пріоритет: 1	Створено: 13.05.2023, В. Блажук
Назва: Перевірка процесу реєстрації користувача		
Вхідні дані: Прізвище = «Блажук», Ім'я = «Вікторія», Нікнейм = «ttori», Email = «ttori2002@gmail.com», Пароль = «ttori2002», Аватар = «photo.jfif»		
Кроки	Очікуваний результат	
<p><i>Передумова:</i> користувач повинен бути не зареєстрованим у системі</p> <ol style="list-style-type: none"> 1. Відкрити вебресурс 2. Натиснути кнопку «Реєстрація» 3. Ввести необхідні дані в поля 4. Натиснути кнопку «Зареєструватись» 5. Порівняти фактичний результат з очікуваним 	<p>Сторінка з формою для реєстрації закрилась, та відкрилась персональна сторінка користувача із особистими даними.</p> <p>В БД у таблиці Users з'явився новий запис з інформацією про зареєстрованого користувача.</p>	
Результат виконання тест-кейсу: пройдено успішно		

Якщо при реєстрації не виникло помилок, що означає, що користувач ввів вірні дані, в БД з'явиться новий запис (Рисунок 3.10). Цей запис зберігатиме інформацію про щойно зареєстрованого користувача. Також, ця інформація буде використовуватись при повторному вході користувача у систему, тобто при авторизації.

	id	name	surname	nickname	email	password	avatar	user_type_id	registration_date	deregistration_date
<input type="checkbox"/>	1	Гнатко	Лідія	hnatko_li	hnatko_li@gmail.com	\$2y\$10\$хgp.s	img/users/C	1	2023-04-12 15:04:58	NULL
<input type="checkbox"/>	2	Вікторія	Блажук	ttori	ttori2002@gmail.com	\$2y\$10\$OA3i	img/users/F	1	2023-05-13 12:58:36	NULL

Рисунок 3.10 – Новий запис в БД з інформацією про зареєстрованого користувача

Після того, як користувач зареєструвався на вебресурсі, йому стане доступна його персональна сторінка з особистою інформацією. У нових користувачів особистий профіль немає жодних дописів, тому користувачі

можуть опублікувати їх будь-коли. Для цього потрібно написати текст для допису у відповідному полі на сторінці та обрати фото допису (Рисунок 3.11).

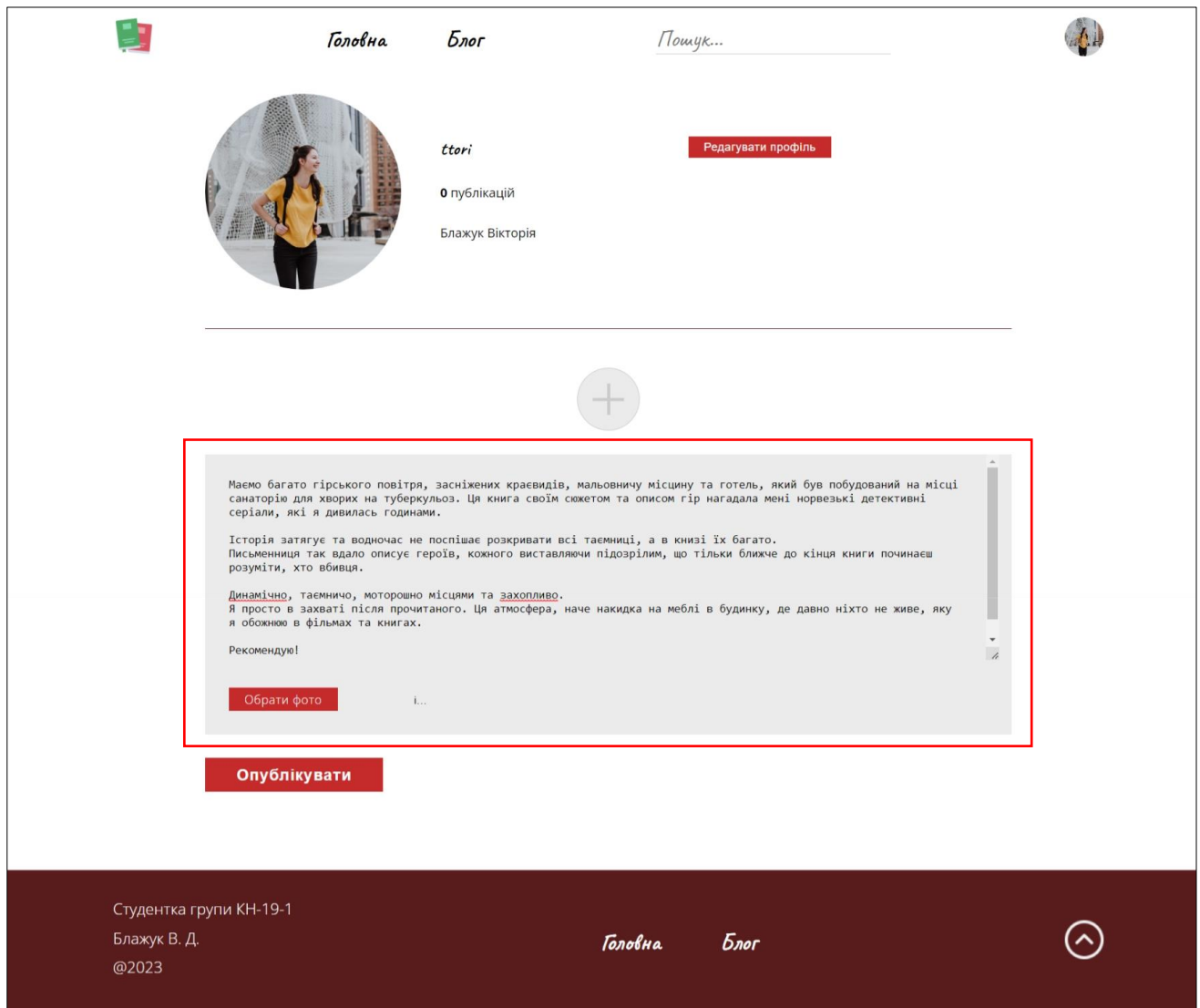


Рисунок 3.11 – Процес публікації допису

Коли текст допису написаний, користувач повинен натиснути на кнопку «Опублікувати». Одразу ж після цього визначається емоційне забарвлення тексту за допомогою розробленого методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом. Далі, уся інформація про допис, включаючи визначене емоційне забарвлення, додається у базу даних. Перевірку описаного процесу проведено через тест-кейс (Таблиця 3.2).

Таблиця 3.2 – Тест-кейс АТ0002

Тест-кейс ID: АТ0002	Пріоритет: 1	Створено: 14.05.2023, В. Блажук
Назва: Перевірка процесу публікації дописів		
Вхідні дані: Текст допису, фото		
Кроки	Очікуваний результат	
<p><i>Передумова:</i> користувач повинен бути авторизованим у системі</p> <ol style="list-style-type: none"> 1. Відкрити вебресурс 2. Ввійти в особистий профіль користувача 3. Відкрити особисту сторінку 4. Написати текст допису у відповідне поле та додати фото 5. Натиснути на кнопку «Опублікувати» 6. Порівняти фактичний результат з очікуваним 	<p>Особиста сторінка користувача перезавантажилась. На сторінці відображається новий допис з визначеним емоційним забарвленням (кольорова рамка навколо фото).</p> <p>Очікуване визначення емоційного забарвлення: позитивне.</p> <p>Фактичне визначення емоційного забарвлення: позитивне.</p> <p>В БД в таблиці «Post» з'явився новий запис з інформацією про допис.</p>	
Результат виконання тест-кейсу: пройдено успішно		

Після успішної публікації допису, на сторінці користувача відобразатиметься новий допис. Оскільки, визначення емоційного забарвлення відбувається автоматизовано, то допис матиме кольорову рамку, відповідно до емоційного забарвлення (Рисунок 3.12). У даному випадку, як і очікувалось, рамка допису зеленого кольору, адже текст допису має чітке позитивне забарвлення. Такий підхід дає змогу користувачеві бачити емоційні забарвлення усіх дописів не переглядаючи сторінку конкретного допису.

Крім цього, у БД можна побачити новий запис з інформацією про щойно опублікований допис (Рисунок 3.13). Ця інформація включає в себе: текст допису, шлях до фото, автора допису, дату публікації та визначене емоційне забарвлення допису.

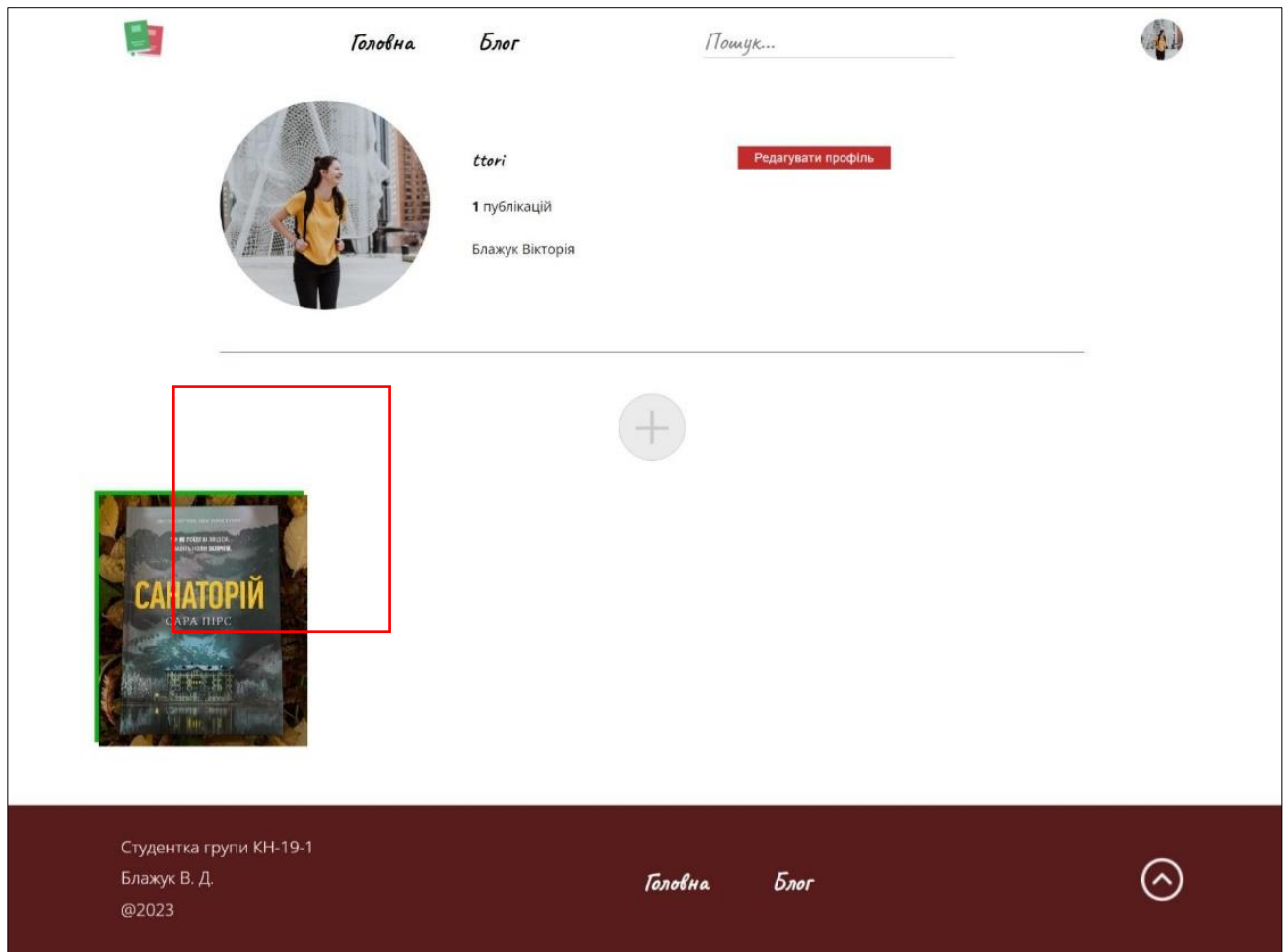


Рисунок 3.12 – Відображення нового допису на персональній сторінці користувача

	id	text	photo	user_id	emotion_id	date	created_at	updated_at
<input type="checkbox"/>	13	Я дуже дуже люблю цю книгу. І не лише тому, що гол...	imglbooks/Oo	1	1	2023-04-25 16:46:08	2023-04-25 13:46:08	2023-04-25 13:46:08
<input type="checkbox"/>	14	Знаєте, я проковтнула цю історію буквально за дві ...	imglbooks/UZ	1	1	2023-04-25 17:04:32	2023-04-25 14:04:32	2023-04-25 14:04:32
<input type="checkbox"/>	15	Що ж... в мене дуже змішані почуття до цієї книги аб...	imglbooks/Mq	1	2	2023-04-25 17:46:50	2023-04-25 14:46:50	2023-04-25 14:46:50
<input type="checkbox"/>	16	Думаю варто обґрунтувати свою оцінку. Книга напевн...	imglbooks/AR	1	2	2023-04-25 17:50:55	2023-04-25 14:50:55	2023-04-25 14:50:55
<input type="checkbox"/>	17	Хоча і читала книгу досить давно, я все-одно розпо...	imglbooks/lJT	1	1	2023-04-25 23:27:09	2023-04-25 20:27:09	2023-04-25 20:27:09
<input type="checkbox"/>	19	Маємо багато гірського повітря, засніжених краєвид...	imglbooks/AC	2	1	2023-05-14 14:31:51	2023-05-14 11:31:51	2023-05-14 11:31:51

Рисунок 3.13 – Відображення інформації про новий допис у БД

Щоб прочитати текст допису користувачу потрібно натиснути на фото допису. Після цього відкриється нова сторінка, на якій буде відображено фото допису, текст та визначене емоційне забарвлення (Рисунок 3.14).

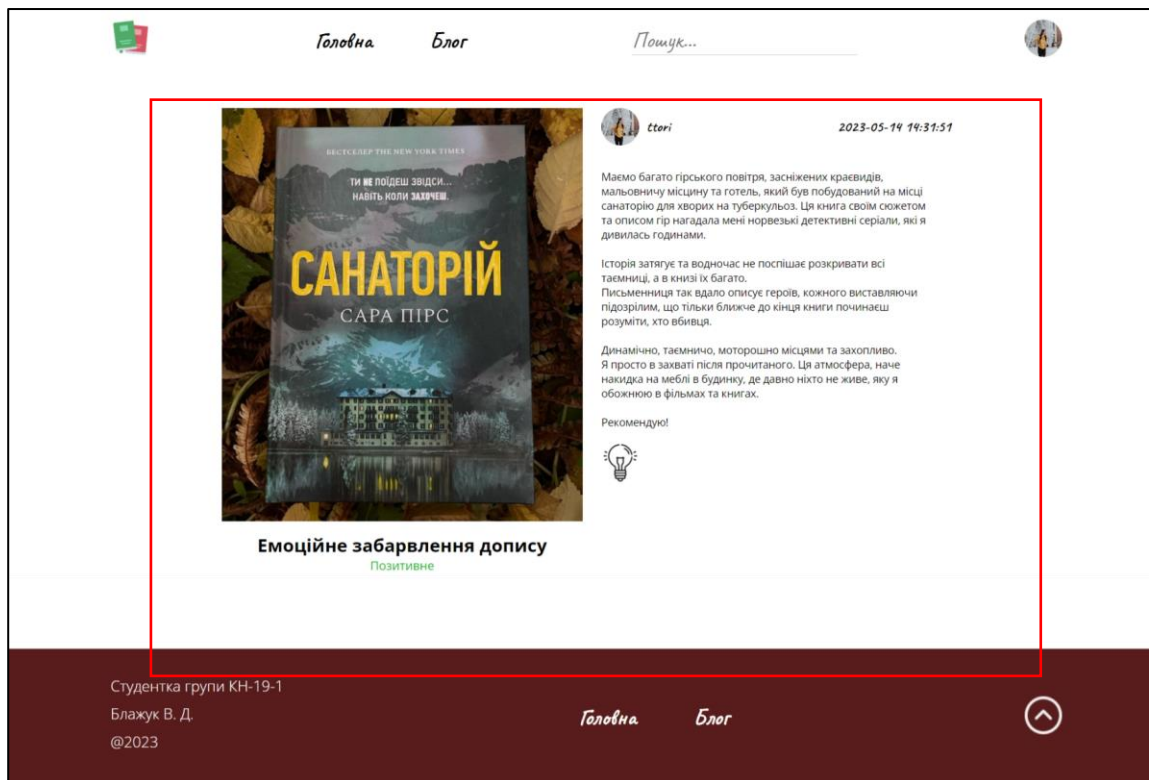


Рисунок 3.14 – Відображення сторінки з обраним дописом

Для кращого розуміння, чому система класифікувала обраний допис як позитивний, користувач може скористатись функцією підсвітки емоційно забарвлених слів. Для цього потрібно натиснути на зображення лампочки під текстом допису. Це виділить зеленим кольором позитивні слова у тексті, які були визначені як позитивні, ще на етапі формування векторів семантично важливих слів (Рисунок 3.15). Щодо зображення лампочки, то воно також зміниться – лампочка засвітиться жовтим кольором. Такий прийом застосовується для того, щоб показати користувачеві, що певний функціонал активований. Повторне натиснення на лампочку «вимкне» її, а виділення зі слів зникне.

Для перевірки алгоритму виділення емоційно забарвлених слів створено відповідний тест-кейс (Таблиця 3.3). У тест-кейсі порівнюється очікуваний список виділених слів із списком отриманих виділених слів. В результаті порівняння можна зробити висновок, що практично всі очікувані слова були виділені системою. Також, система більше зосереджена на виділенні загальних слів, ніж на абстрактних описах.



 **ttori** 2023-05-14 14:31:51

Маємо багато гірського повітря, засніжених краєвидів, **мальовничу** місцину та готель, який був побудований на місці санаторію для хворих на туберкульоз. Ця книга своїм сюжетом та описом гір нагадала мені норвезькі детективні серіали, які я дивилась годинами.

Історія **затягує** та водночас не поспішає розкривати всі таємниці, а в книзі їх багато. Письменниця так **вдало** описує героїв, кожного виставляючи підозрілим, що тільки ближче до кінця книги починаєш розуміти, хто вбивця.

Динамічно, таємничо, моторошно місцями та **захопливо**. Я просто **в захваті** після прочитаного. Ця атмосфера, наче накидка на меблі в будинку, де давно ніхто не живе, яку я обожаю в фільмах та книгах.

Рекомендую!



Емоційне забарвлення допису
Позитивне

Рисунок 3.15 – Результат виділення позитивних слів у тексті допису

Таблиця 3.3 – Тест-кейс АТ0003

Тест-кейс ID: АТ0003	Пріоритет: 2	Створено: 14.05.2023, В. Блажук
Назва: Перевірка алгоритму виділення емоційно забарвлених слів із тексту		
Вхідні дані: Текст допису, очікуваний список виділених слів, отриманий список виділених слів		
Кроки	Очікуваний результат	
<p><i>Передумова:</i> допис повинен бути опублікований</p> <ol style="list-style-type: none"> Відкрити вебресурс Обрати допис та натиснути на нього Натиснути на зображення лампочки на сторінці з дописом Порівняти фактичний результат з очікуваним 	<p>Після натиснення на зображення лампочки, зображення змінилось (лампочка засвітилась).</p> <p>Очікуваний список виділених слів: багато гірського повітря, засніжених краєвидів, мальовничу місцину, історія затягує, вдало, динамічно, таємничо, захопливо, просто в захваті, обожною, рекомендую.</p> <p>Отриманий список виділених слів: мальовничу, затягує, вдало, динамічно, таємничо, захопливо, захваті, обожною, рекомендую.</p>	
Результат виконання тест-кейсу: пройдено успішно		

Таким чином, користувач може коректно додавати множину дописів та переглядати їхнє емоційне забарвлення. Проведене тестування розробленої інформаційної системи доводить, що реалізований функціонал працює вірно. Користувачі мають змогу реєструватись без виникнення проблем та помилок. Публікація дописів, також, не викликає непорозумінь, адже система працює вірно. Щодо автоматизованого визначення емоційного забарвлення дописів, можна сказати, що у розробленому вебресурсі реалізована така можливість і, що цей функціонал працює правильно.

3.4 Дослідження ефективності методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом

Дослідження ефективності розробленого методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом дасть змогу визначити наскільки успішно працює цей метод. Також, це дозволить виявити потенційні проблеми у роботі методу або його слабкі місця, що дасть можливість зробити метод більш точним та ефективнішим.

Для дослідження ефективності методу визначення емоційного забарвлення дописів було відібрано 50 тестових дописів (25 негативних та 25 позитивних). Для проведення дослідження необхідно було порівняти оцінку експерта та результати роботи методу, щодо визначення емоційного забарвлення допису. За оцінку експерта було прийнято рішення взяти відповідь чат-боту ChatGPT. Для отримання оцінки емоційного забарвлення допису необхідно написати чат-боту запит, який включатиме текстовий контент допису та вимогу для визначення його емоційного забарвлення.

Для демонстрації процесу проведення дослідження ефективності розробленого методу було взято негативний відгук на книгу. Спочатку, обраний відгук аналізувався через ChatGPT. Як видно з проведеного аналізу чат-бот правильно класифікував відгук, як негативний (Рисунок 3.16). Крім цього,

штучний інтелект надає користувачеві експертну оцінку, яка пояснює причини отриманого результату.

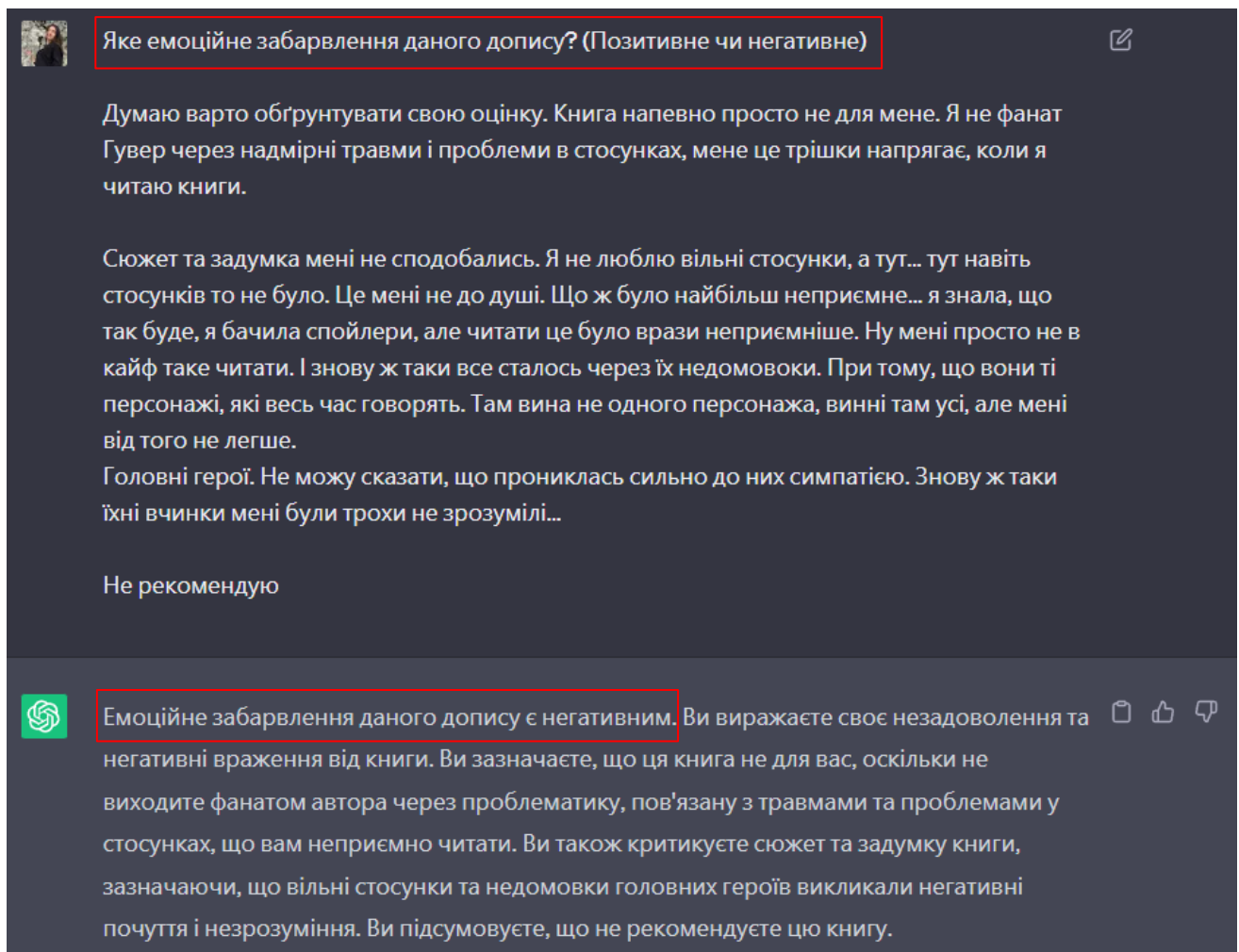



Рисунок 3.16 – Визначення емоційного забарвлення допису через ChatGPT

Далі, цей же відгук було проаналізовано через розроблений метод визначення емоційного забарвлення дописів. В даному випадку, метод, також, визначає допис як негативний (Рисунок 3.17). Для підтвердження отриманої методом оцінки емоційного забарвлення допису, в тексті виділяються слова, які мають негативне емоційне забарвлення.



hnatko_li

2023-04-25 17:50:55

Думаю варто обґрунтувати свою оцінку. Книга напевно просто не для мене. Я не фанат Гувер через надмірні травми і проблеми в стосунках, мене це трішки напругає, коли я читаю книги.

Сюжет та задумка мені не сподобались. Я не люблю вільні стосунки, а тут... тут навіть стосунків то не було. Це мені не до душі. Що ж було найбільш неприємне... я знала, що так буде, я бачила спойлери, але читати це було вразі неприємніше. Ну мені просто не в кайф таке читати. І знову ж таки все сталося через їх недомовки. При тому, що вони ті персонажі, які весь час говорять. Там вина не одного персонажа, винні там усі, але мені від того не легше.

Головні герої. Не можу сказати, що прониклась сильно до них симпатією. Знову ж таки їхні вчинки мені були трохи не зрозумілі...

Не рекомендую

Емоційне забарвлення допису

Негативне

Рисунок 3.17 – Результат визначення емоційного забарвлення допису з використанням розробленого методу

Для решти дописів із множини тестових дописів було проведено аналогічне порівняння. Такий підхід дає змогу швидко порівняти роботу розробленого методу із результатами роботи добре навченої нейронної мережі, яку використовує чат-бот ChatGPT.

Проведене дослідження ефективності розробленого методу визначення емоційного забарвлення дописів, шляхом порівняння експертної оцінки сформованої через ChatGPT та результату роботи методу, дало змогу отримати дані по ефективності роботи методу. Таким чином, було встановлено, що емоційне забарвлення 89,91% дописів із тестової множини було визначено правильно. Інші 10,09% – це похибка у роботі методу визначення емоційного забарвлення дописів (Рисунок 3.18).

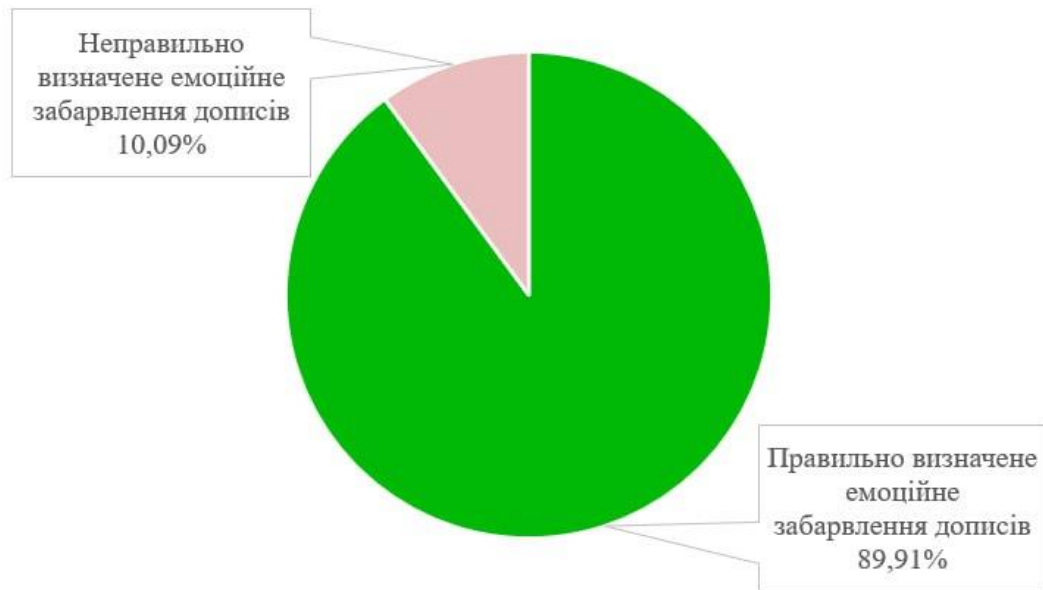


Рисунок 3.18 – Загальна діаграма ефективності роботи методу

В ході порівняння результатів роботи методу та отриманої експертної оцінки було встановлено, що 46,31% із 50% позитивних дописів були правильно визначені методом. У випадку негативних дописів метод допускає більше помилок, адже правильно визначеними були 44,60% із 50% негативних дописів (Рисунок 3.19).

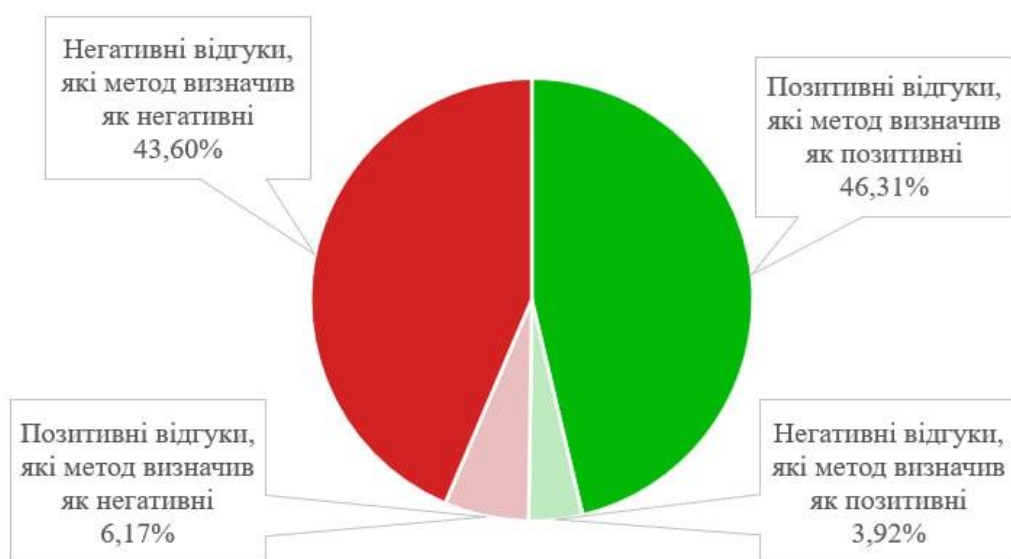


Рисунок 3.19 – Детальна діаграма ефективності роботи методу

Отже, розроблений метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом трохи краще визначає емоційне забарвлення позитивних дописів. Однією з причин такого результату може бути краще підготовлена множина позитивних навчальних дописів. Отже, для підвищення ефективності роботи методу, варто підготувати більше зразків дописів із негативним забарвленням.

Проведене дослідження ефективності роботи розробленого методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом дало змогу встановити, що ефективність роботи методу становить 89,91%, адже саме такий відсоток емоційного забарвлення визначеного методом співпав із експертною оцінкою отриманою засобом ChatGPT.

3.5 Висновки до розділу 3

Використання фреймворків та бібліотек при розробці програмних продуктів значно пришвидшує та покращує процес програмування. Використання шаблону проєктування MVC дало змогу розділити різну логіку програми на окремі взаємопов'язані компоненти. В майбутньому, це допомагає для кращого розуміння програмного коду, що в свою чергу потрібне для ефективної підтримки та покращення продукту.

Для розробки прикладного застосування методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом було обрано предметну область блогу для написання відгуків на книги. Вебресурс було реалізовано з використанням фреймворку Laravel для мови програмування PHP. Фреймворк надає велику кількість бібліотек та готових програмних рішень, серед яких система аутентифікації користувачів. Для реалізації розробленого методу визначення емоційного забарвлення дописів було використано PHP бібліотеку PHP-ML, яка надає інструменти для роботи із класифікаторами та іншими статистичними методами, серед яких TF-IDF.

Розроблена інформаційна система, яка включає реєстрацію та авторизацію користувачів, публікацію дописів, визначення їх емоційного забарвлення та візуалізацію емоційно забарвлених слів працює вірно. Завдяки системі автентифікації, користувачі можуть створювати свої особисті сторінки та публікувати дописи про враження від прочитаних книг. Процес реєстрації користувачів відбувається правильно та без помилок, що дозволяє зберігати дані про користувачів у БД та ідентифікувати авторів дописів. Процес публікації дописів також працює правильно. Після публікації допису система автоматично визначає його емоційне забарвлення. Користувач може переглянути результати роботи методу через функцію виділення емоційно забарвлених слів. Це допомагає краще зрозуміти роботу програми, щодо класифікації текстів за емоціями.

Отже, результатом програмної реалізації є вебресурс у вигляді блогу для написання відгуків на книги з автоматизованим визначенням емоційного забарвлення дописів. Вебресурс включає весь необхідний функціонал для користувачів, зокрема: реєстрацію, публікацію дописів та перегляд емоційного забарвлення дописів. Розроблена інформаційна система працює правильно, а у разі виникнення помилок повідомляє про це користувача. Інтерфейс користувача допомагає легко орієнтуватись в програмі та використовувати необхідний функціонал. Проведене дослідження ефективності роботи розробленого методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом дало змогу встановити, що ефективність роботи методу становить 89,91%, саме такий відсоток емоційного забарвлення визначеного методом співпав із експертною оцінкою. Таким чином, метод вирішує поставлену задачу та дає змогу проаналізувати допис і коректно визначити його емоційне забарвлення й сформувати множину ключових слів, які найбільше впливають на емоційне забарвлення допису.

Висновки

Кваліфікаційна робота бакалавра присвячена розробці методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом та відповідної інформаційної системи у вигляді соціально орієнтованого вебресурсу, що дозволяє за текстовим контентом допису визначати його емоційне забарвлення.

Під час виконання кваліфікаційної роботи бакалавра, перш за все, було проаналізовано предметну область, яка стосується обробки природної мови з використанням інформаційних технологій. В процесі дослідження предметної області було встановлено, що завдання, яке полягає у визначенні емоційного забарвлення цифрових текстів дуже важливе та актуальне, адже кожна текстова інформація може нести в собі певний сенс, який супроводжується конкретним емоційним забарвленням. Оскільки, ця інформація може нести вплив на інших людей чи на певні групи суспільства, то варто відстежувати таку інформацію.

Метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів призначений для одержання вихідних даних у вигляді категорії допису за його емоційним забарвленням та переліку емоційно значущих слів тестового допису за вхідними даними у вигляді позитивної й негативної навчальних множини дописів і тестового допису для аналізу емоційного забарвлення, для чого застосовуються формування векторів семантично важливих слів за категоріями емоцій та визначення емоційного забарвлення допису з використанням Баєсового класифікатора та методу TF-IDF.

Метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом полягає в вирішенні задачі класифікації цифрових текстів на дві категорії: позитивні та негативні дописи. Для вирішення цього завдання використовується бібліотека PHP-ML, яка пропонує інструменти для роботи із класифікаторами. Крім цього, розробка вебресурсу здійснюється з використанням PHP фреймворку Laravel. А для

збереження даних про користувачів вебресурсу, дописи які вони публікують та іншої додаткової інформації використовується СКБД MySQL.

Отже, в результаті поєднання обраних засобів розробки було реалізовано соціально орієнтований вебресурс з автоматизованим визначенням емоційного забарвлення дописів, який використовує розроблений метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом й забезпечує такий функціонал:

- аутентифікація користувачів;
- робота користувача з персональними даними;
- додавання власних дописів;
- перегляд особистих сторінок інших користувачів;
- перегляд журналу обраного користувача з усіма дописами;
- первинна обробка текстового контенту дописів;
- навчання класифікатора дописів за емоційним забарвленням методом TF-IDF;
- визначення категорії емоційного забарвлення допису за алгоритмом Баєса;
- формування множини емоційно значущих слів допису з використанням TF-IDF;
- візуалізація емоційно значущих слів.

Створена інформаційна система є дослідницькою, проте може бути використана для аналізу дописів у існуючих соціальних мережах з метою створення певної статистики або відстеження громадської думки. Програмний продукт можна розширити та вдосконалити. Перш за все, це стосується розширення бази для навчання класифікатора, адже чим більшою буде кількість навчальних зразків, тим кращими будуть результати класифікації. Також, оскільки розроблена інформаційна система – це соціально орієнтований вебресурс, її можна значно розширити додавши прикладні можливості створювати чати між користувачами та об'єднуватись у групи зі спільними інтересами.

Дослідження ефективності роботи розробленого методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом, визначена як відсоток від кількості дописів, що були коректно визначені з використанням розробленого методу згідно з експертною оцінкою отриманою через ChatGPT, показує, що ефективність роботи методу становить 89,91%.

Проаналізувавши отриманий результат можна зробити висновок, що мета кваліфікаційної роботи бакалавра досягнута. Тестування прикладного застосування методу визначення емоційного забарвлення дописів показує, що метод вирішує поставлену задачу та дає змогу проаналізувати допис і коректно визначити його емоційне забарвлення та сформувати множину ключових слів, які найбільше впливають на позитивне та негативне емоційне забарвлення допису.

Перелік посилань

1. Idea Digital Agency. Семантика. URL:
<https://ideadigital.agency/blog/semantika>
2. ThoughtCo. An Introduction to Semantics. URL:
<https://www.thoughtco.com/semantics-linguistics-1692080>
3. Тег. Що таке семантика слова і тексту: аналіз і приклади. URL:
<http://teg.com.ua/shho-take-semantika-slova-i-tekstu-analiz-i-prikladi/>
4. Вікіпедія. Природна мова. URL:
https://uk.wikipedia.org/wiki/Природна_мова
5. Онищенко К. Г., Данієль Я., Каменєв Р. Аналіз методів обробки природної мови. Інформаційні системи та технології : матеріали 9-ї Міжнародної науково-технічної конференції, 2020. Харків. С 186–190.
6. IBM. What is natural language processing? URL:
<https://www.ibm.com/topics/natural-language-processing>
7. Вікіпедія. Аналіз тональності тексту. URL:
https://uk.wikipedia.org/wiki/Аналіз_тональності_тексту
8. Освіта. Емоції: види та значення, прояви, механізм виникнення. URL:
<https://osvita.ua/vnz/reports/biolog/26132/>
9. Карпенко Є. В. Сучасні вектори дослідження емоційного інтелекту особистості. Психологія і особистість, 2019. С 60–70
10. Вікіпедія. Ключове слово. URL:
https://uk.wikipedia.org/wiki/Ключове_слово
11. Жиркова А. П, Ігнатенко О. П. Аналіз методів машинного навчання в задачі класифікації документів. Моделі та методи машинного навчання. Київ, 2020. С 81-87
12. Оленич І., Притула М., Сінькевич О, Хамар О. Система автоматизованого визначення тональності тексту. Електроніка та інформаційні технології (випуск 15), 2021. С 16–23
13. Вікіпедія. Веб 2.0. URL: https://uk.wikipedia.org/wiki/Веб_2.0

14. FutureNow. Що таке соціальні мережі?(види, класифікація, безпека)
URL: <https://futurenow.com.ua/shho-take-sotsialni-merezhi-vydy-klasyfikatsiya-bezpeka/>
15. На урок. Поняття форуму, спілкування в чатах, етикет інтерактивного спілкування. URL: <https://naurok.com.ua/ponyattya-forumu-spilkuvannya-v-chatah-etiket-interaktivnogo-spilkuvannya-156091.html>
16. Вікіпедія. Відеохостинг. URL: <https://uk.wikipedia.org/wiki/Відеохостинг>
17. Smartik. Які є електронні пошти? Топ-5 безкоштовних поштових сервісів. URL: <https://smartik.kiev.ua/iaki-ie-elektronni-poshty/>
18. Vseosvita. Що таке месенджери, які у них можливості – топ популярних месенджерів. URL: <https://vseosvita.ua/library/so-take-mesendzeri-aki-u-nih-mozlivosti-top-popularnih-mesendzeriv-281047.html>
19. Пода Т.А. Феномен соціальних мереж та їхній вплив на суспільно-політичні процеси у глобалізованому світі. Вісник Національного авіаційного університету. Збірник наукових праць, 2019. С 141-145
20. Opinosis analytics. What is Term Frequency? URL: <https://www.opinosis-analytics.com/knowledge-base/term-frequency-explained/>
21. Вікіпедія. TF-IDF. URL: <https://uk.wikipedia.org/wiki/TF-IDF>
22. Cran. Textrank for summarizing text. URL: <https://cran.r-project.org/web/packages/textrank/vignettes/textrank.html>
23. Batool Armouty, Sara Tedmori. Automated Keyword Extraction using Support Vector Machine. Jordan International Joint Conference on Electrical Engineering and Information Technology, Amman, 2019. С 109-125
24. Istio. Семантичний аналіз тексту istio. URL: <https://istio.com/uk/text/analyz>
25. MonkeyLearn. Sentiment Analyzer. URL: <https://monkeylearn.com/sentiment-analysis-online/>
26. OpenAI. ChatGPT. URL: <https://openai.com/blog/chatgpt>
27. Laravel. URL: <https://laravel.com/>

28. Hostinger tutorials. Що таке MySQL: пояснення MySQL для початківців. URL: <https://www.hostinger/rukovodstva/shto-takoje-mysql>

29. Освіта. Що таке MVC? URL: <https://uk.education-wiki.com/3886982-what-is-mvc>

30. PHP-ML. URL: <https://php-ml.readthedocs.io/en/latest/>

ДОДАТКИ

Додаток А

Схема методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів



Додаток Б

Схема структури соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів



Додаток В

Основні програмні коди

Лістинг Dataset.php

```
<?php
namespace App\Models;
use Illuminate\Database\Eloquent\Factories\HasFactory;
use Illuminate\Database\Eloquent\Model;
class Dataset extends Model
{
    use HasFactory;
    public function emotion(){
        return $this->belongsTo(Emotion::class);
    }
}
```

Лістинг Emotion.php

```
<?php
namespace App\Models;
use Illuminate\Database\Eloquent\Factories\HasFactory;
use Illuminate\Database\Eloquent\Model;
class Emotion extends Model
{
    use HasFactory;
    public function datasets(){
        return $this->hasMany(Dataset::class, foreignKey:'emotion_id', localKey:'id');
    }
    public function posts(){
        return $this->hasMany(Post::class, foreignKey:'emotion_id', localKey:'id');
    }
}
```

Лістинг Post.php

```
<?php
namespace App\Models;
use Illuminate\Database\Eloquent\Factories\HasFactory;
use Illuminate\Database\Eloquent\Model;
class Post extends Model
{
    use HasFactory;
    protected $fillable = ['user_id', 'emotion_id', 'text', 'photo'];
    public function user(){
        return $this->belongsTo(User::class);
    }
    public function emotion(){
        return $this->belongsTo(Emotion::class);
    }
}
```

Лістинг User.php

```
<?php
namespace App\Models;
// use Illuminate\Contracts\Auth\MustVerifyEmail;
use Illuminate\Database\Eloquent\Factories\HasFactory;
use Illuminate\Foundation\Auth\User as Authenticatable;
use Illuminate\Notifications\Notifiable;
use Laravel\Sanctum\HasApiTokens;
class User extends Authenticatable
{
    use HasApiTokens, HasFactory, Notifiable;
    /**
     * The attributes that are mass assignable.
     *
     * @var array<int, string>
     */
    protected $fillable = ['surname', 'name', 'nickname', 'login', 'email', 'password', 'avatar'];
    /**
     * The attributes that should be hidden for serialization.
     *
     * @var array<int, string>
     */
}
```

```

    */
    protected $hidden = [
        'password',
        'remember_token',
    ];
    public function user_type(){
        return $this->belongsTo(UserType::class);
    }
    public function posts(){
        return $this->hasMany(Post::class, foreignKey:'user_id', localKey:'id');
    }
}

```

Лістинг UserType.php

```

<?php
namespace App\Models;
use Illuminate\Database\Eloquent\Factories\HasFactory;
use Illuminate\Database\Eloquent\Model;
class UserType extends Model
{
    use HasFactory;
    public function users(){
        return $this->hasMany(User::class, foreignKey:'user_type_id', localKey:'id');
    }
}

```

Лістинг Classification.php

```

<?php
use App\Models\Dataset;
use Phpml\FeatureExtraction\TokenCountVectorizer;
use Phpml\FeatureExtraction\TfIdfTransformer;
use Phpml\Tokenization\WhitespaceTokenizer;
use Phpml\Classification\NaiveBayes;
function bayes($input)
{
    $data = Dataset::all();
    $samples = [];
    $labels = [];
    $chars = ['.', '?', ',', '!', '-'];
    $text = [$input];
    for ($i = 0; $i < count($data); $i++) {
        $samples[$i] = strtolower(str_replace($chars, '', $data[$i]->text));
        $labels[$i] = $data[$i]->emotion->id;
    }
    $vectorizer = new TokenCountVectorizer(new WhitespaceTokenizer());
    $vectorizer->fit($samples);
    $vectorizer->transform($samples);
    $vectorizer->fit($text);
    $vectorizer->transform($text);
    $tfidfTransformer = new TfIdfTransformer();
    $tfidfTransformer->fit($samples);
    $tfidfTransformer->transform($samples);
    $classifier = new NaiveBayes();
    $classifier->train($samples, $labels);
    $emotion = $classifier->predict($text);
    return ($emotion[0]);
}

```

Лістинг HomeController.php

```

<?php
namespace App\Http\Controllers;
use Illuminate\Http\Request;
class HomeController extends Controller
{
    /**
     * Create a new controller instance.
     * @return void
     */
    public function __construct()
    {
        $this->middleware('auth');
    }
    /**
     * Show the application dashboard.
     *
     * @return \Illuminate\Contracts\Support\Renderable
     */
    public function index()
    {

```

```

    return view('home');
}
}

```

Лістинг PostsController.php

```

<?php
namespace App\Http\Controllers;
use App\Models\Post;
use Illuminate\Http\Request;
use Illuminate\Support\Facades\Auth;
class PostsController extends Controller
{
    public function store(Request $request){
        $data=$request->validate([
            'text'=>'string',
        ]);
        $path = '/img/users/default.png';
        if (isset($request['photo'])) {
            $path = $request['photo']->store('img\books', 'public');
        }
        $emotion=bayes($data['text']);
        $post=Post::create([
            'user_id'=>Auth::user()->id,
            'emotion_id'=>$emotion,
            'text'=>$data['text'],
            'photo'=>$path
        ]);
        if($post){
            return redirect()->route('home');
        }
        else{
            return redirect()->route('home')->withErrors([
                'formError'=>'Виникла помилка при публікації допису'
            ]);
        }
    }
    public function show(Post $post){
        return view('posts.show', compact('post'));
    }
    public function index(){
        $posts=Post::all();
        return view('posts.index', compact('posts'));
    }
}

```

Лістинг main.blade.php

```

<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    @vite(['resources/css/app.css', 'resources/js/app.js'])
    <link rel="preconnect" href="https://fonts.googleapis.com">
    <link rel="preconnect" href="https://fonts.gstatic.com" crossorigin>
    <link href="https://fonts.googleapis.com/css2?family=Caveat:wght@400;700&family=Open+Sans:wght@300;400;700&display=swap"
rel="stylesheet">
    <link rel="icon" href="/img/logo.png">
    <title>Diplom</title>
</head>
<body>
    <header>
        <div class="navigation">
            <div class="logo">
                <a href="/">
                    
                </a>
            </div>
            <div class="menu">
                <h2><a href="/" class="menu_item">Головна</a></h2>
                <h2><a href="{{route('posts.index')}}" class="menu_item">Блог</a></h2>
            </div>
            <div class="search">
                <input type="text" placeholder="Пошук..." class="search_input">
            </div>
            @guest
                <h2><a href="{{route('login')}}" class="menu_item">Вхід</a></h2>
            @endguest
            @auth
                <a href="{{route('home')}}" class="header__button">
                    
                </a>
            @endauth

```

```

    </div>
</header>

@yield('content')
<footer>
    <div class="footer_content">
        <div class="about">
            <p class="about_item">Студентка групи КН-19-1</p>
            <p class="about_item">Блажук В. Д.</p>
            <p class="about_item">@2023</p>
        </div>
        <div class="menu">
            <h2><a href="#" class="menu_item_footer">Головна</a></h2>
            <h2><a href="#" class="menu_item_footer">Блог</a></h2>
        </div>
        <div class="up">
            
        </div>
    </div>
</footer>
</body>
</html>

```

Лістинг posts.index.blade.php

```

@extends('layouts.main')
@section('content')
<div class="blog_wrapper">
    <h1>Нові дописи</h1>
    <div class="user_posts_section_wrapper margin_top_25">
        @foreach($posts as $post)
            @if($post->emotion->id==1)
                <div class="posts_card_wrapper">
                    <div class="post_info_user margin_bottom_15 margin_left_10">
                        
                        <h4 class="caveat margin_left_10">{{$post->user->nickname}}</h4>
                    </div>
                    <div class="post_card grin">
                        <a href="{{route('posts.show', $post->id)}}">
                            
                        </a>
                    </div>
                </div>
            </div>
            @endif
            @if($post->emotion->id==2)
                <div class="posts_card_wrapper">
                    <div class="post_info_user margin_bottom_15 margin_left_10">
                        
                        <h4 class="caveat margin_left_10">{{$post->user->nickname}}</h4>
                    </div>
                    <div class="post_card red">
                        <a href="{{route('posts.show', $post->id)}}" id="post">
                            
                        </a>
                    </div>
                </div>
            </div>
            @endif
        @endforeach
    </div>
</div>
@endsection

```

Лістинг posts.show.blade.php

```

@extends('layouts.main')
@section('content')
<div class="post_show_section_wrapper">
    <div class="post_show_section">
        <div class="post_show_first_section">
            
            <div class="post_emotion_info">
                @if($post->emotion->id==1)
                    <h3>Емоційне забарвлення допису</h3>
                    <p class="green_text">{{$post->emotion->name}}</p>
                @endif
                @if($post->emotion->id==2)
                    <h3>Емоційне забарвлення допису</h3>
                    <p class="red_text">{{$post->emotion->name}}</p>
                @endif
            </div>
        </div>
        <div class="post_content">
            <div class="post_info">

```

```

        <div class="post_info_user">
            
            <h4 class="caveat margin_left_10">{{ $post->user->nickname}}</h4>
        </div>
        <h4 class="caveat">{{ $post->date}}</h4>
    </div>
    <p id="PostText" class="post_text">{{nl2br($post->text)}}</p>
    <button id="button" class="img_btn"></button>
</div>
</div>
</div>
@endsection

```

Лістинг home.blade.php

```

@extends('layouts.main')
@section('content')
<div class="home_wrapper">
    <div class="about_user_wrapper">
        
        <div class="user_info_wrapper">
            <div class="user_info">
                <h3 class="caveat">{{Auth::user()->nickname}}</h3>
                <form id="logout-form" action="{{route('logout')}}" method="POST">
                    @csrf
                    <input type="submit" value="Редагувати профіль" class="button_2">
                </form>
            </div>
            <div class="user_info">
                <p><b>{{Auth::user()->posts->count()}}</b> публікацій</p>
                <p class="white"><b>{{Auth::user()->posts->count()}}</b> підписників</p>
                <p class="white"><b>{{Auth::user()->posts->count()}}</b> підписок</p>
            </div>
            <div class="user_info">
                <p>{{Auth::user()->surname.' '.Auth::user()->name}}</p>
            </div>
        </div>
    </div>
    <div class="post_editor_section">
        
        <form action="{{route('post.store')}}" method="post" enctype="multipart/form-data">
            @csrf
            <div class="textarea_wrapper">
                <textarea name="text" id="" cols="30" rows="10" placeholder="Напишіть текст допису..."
class="textarea"></textarea>
                @error('text')
                <div class="danger">{{ $message}}</div>
                @enderror
                <input type="file" class="file_button margin_left_30" name="photo">
            </div>
            <button class="button_smaller">Опублікувати</button>
        </form>
    </div>
    <div class="user_posts_section_wrapper">
        @foreach(Auth::user()->posts as $post)
        @if($post->emotion->id==1)
        <div class="post_card grin">
            <a href="{{route('posts.show', $post->id)}}">
                
            </a>
        </div>
        @endif
        @if($post->emotion->id==2)
        <div class="post_card red">
            <a href="{{route('posts.show', $post->id)}}">
                
            </a>
        </div>
        @endif
        @endforeach
    </div>
</div>
@endsection

```

Лістинг index.blade.php

```

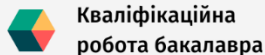
@extends('layouts.main')
@section('content')
<div class="main_page_content">
    
    <div class="intro">
        
        <div class="offer">

```

```
<div class="offer_main_text">
  <h1>Поділись враженнями</h1>
  <h1 class="light">про прочитане</h1>
</div>
<div class="offer_add_text">
  <h3 class="light">Просто зареєструйся тут</h3>
</div>
<div class="main_page_button">
  <a href="{{route('register')}}" class="button">Реєстрація</a>
</div>
<div class="second_add_text">
  <p>Блог для книголюбів</p>
</div>
</div>
</div>
</div>
@endsection
```

Додаток Г

Презентаційний матеріал



Метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом

Виконала: студентка 4 курсу, групи КН-19-1 Блажук Вікторія Дмитрівна

Керівник: к.т.н, доцент кафедри КН Мазурець Олександр Вікторович



Розвиток та активне використання соціально орієнтованих вебресурсів призводять до щоденного збільшення кількості контенту в мережі. Кожен написаний та опублікований допис несе в собі певне емоційне забарвлення, і як результат, може впливати на читачів. Тому, аналіз дописів допомагає відстежувати думки та настрої соціуму.

Проблема визначення емоційного забарвлення дописів стає особливо актуальною під час значних змін у суспільстві. Саме тому, дослідження даної проблеми є актуальним в наш час.





Мета і задачі роботи

МЕТОД ВИЗНАЧЕННЯ ЕМОЦІЙНОГО ЗАБАРВЛЕННЯ ДОПИСІВ СОЦІАЛЬНО ОРІЄНТОВАНИХ ВЕБРЕСУРСІВ ЗА СЕМАНТИЧНИМ АНАЛІЗОМ

МЕТОЮ кваліфікаційної роботи бакалавра є розробка й програмна реалізація методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом

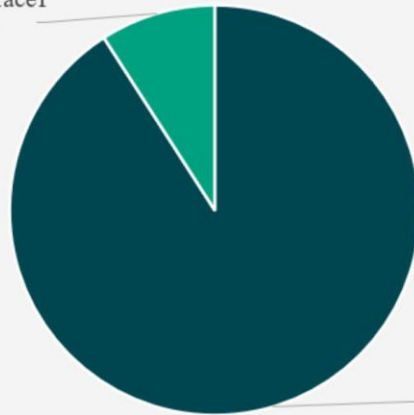
ЗАВДАННЯ:

- Аналіз предметної області визначення емоційного забарвлення цифрових текстів за семантичним аналізом;
- Аналіз методів пошуку ключових слів у тексті;
- Розробка методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом;
- Проектування структури прикладного соціально орієнтованого вебресурсу та БД;
- Програмна реалізація соціально орієнтованого вебресурсу;
- Дослідження ефективності розробленого методу



Формування навчальних множин дописів для аналізу емоційного забарвлення

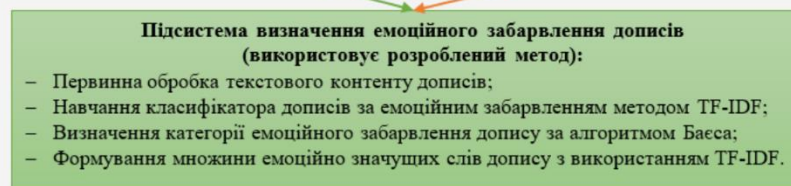
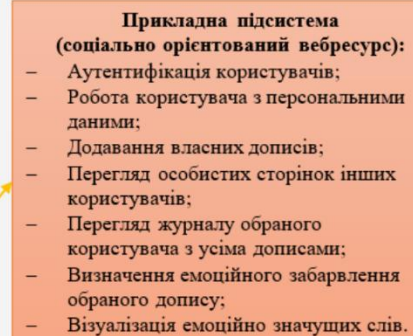
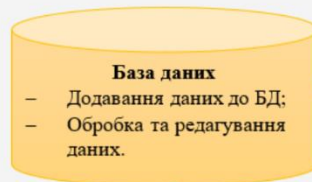
Тренувальний датасет
(50 відгуків)



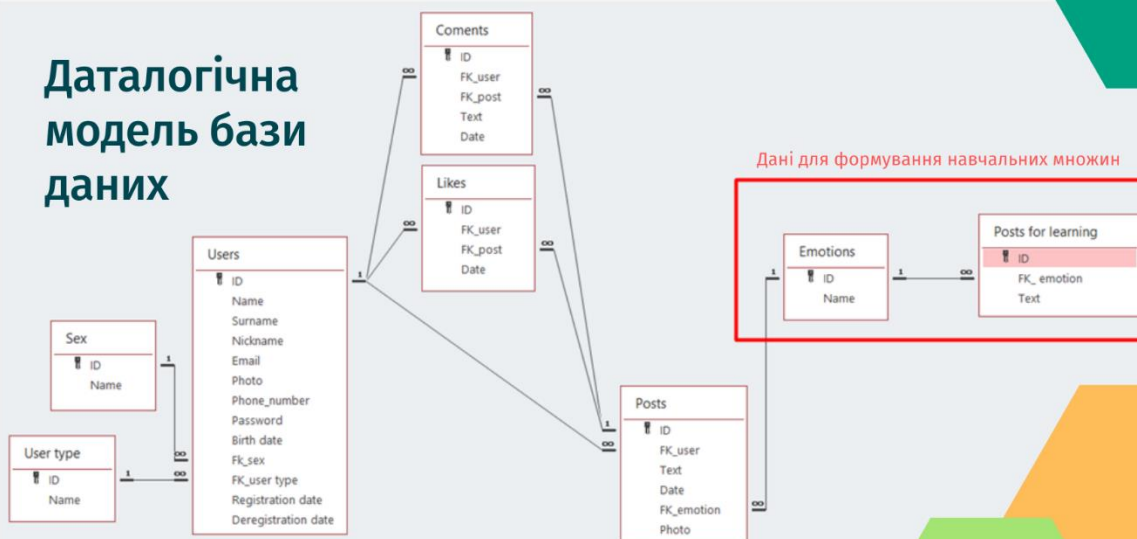
Навчальний датасет
(500 відгуків)

Датасети створено з використанням ChatGPT

Схема структури соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів

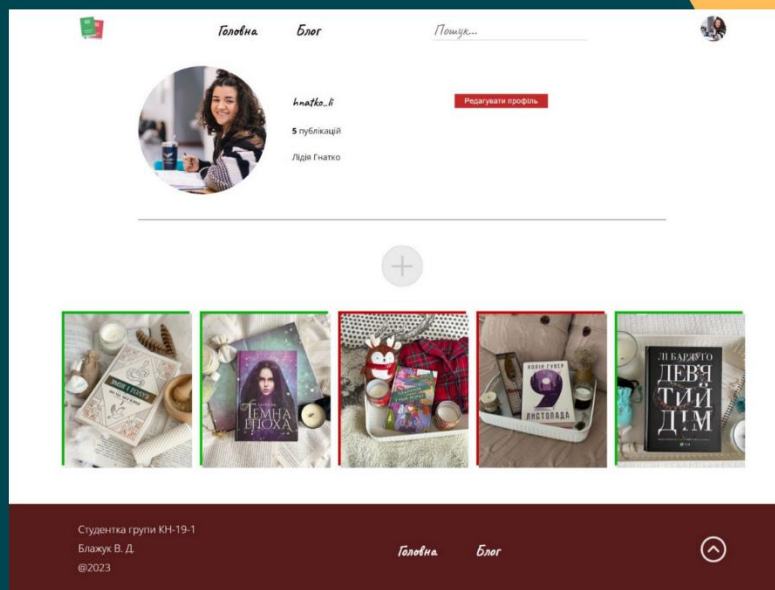


Даталогічна модель бази даних



Програмна реалізація

Особиста сторінка авторизованого користувача із переліком опублікованих дописів



Дослідження ефективності методу



Програмна реалізація

Сторінка з обраним дописом

Результат роботи методу на прикладі визначення позитивного емоційного забарвлення допису



Емоційне забарвлення допису

Позитивне



hnatko_li

2023-04-25 16:46:08

Я **дуже** **дуже** люблю цю книгу. І не лише тому, що головну героїню звати так само, як мене, а тому що вона дійсно була **дуже** **цікава**. Книга буквально з перших сторінок затягує у свої переплетіння.

Персонажі

Луза та Рід це просто вогонь. Вони **настільки** **прикольні** разом. Їхня хімія була **дуже** **класною**, навіть попри обставини їхньої зустрічі, було чимало **милих** моментів.

Лу це дійсно одна з **улюблених** героїнь, тому що **настільки** було **весело** спостерігати за тим, як Рід реагує на її фразочки, пісеньки, підколи. Вона **весела**, відкрита, **відважна**, **емоційна** та з характером, в той час як Рід стриманий, **набожний**, не надто емоційний, **серйозний**, цей контраст між ними був **чудовий**. Вони доповнювали один одного. Ну і звісно розвиток їхніх стосунків від ненависті до кохання, що просто не могло не радувати.

Сюжет

Він дійсно **захоплює**. Сюжетні повороти просто **шикарні** та не дають відірватись від книги. Коли здається, що нас більше вже не здивувати, автори все ж вдаються це знову зробити. Книга закінчилась **дуже** **круто**, **емоційно**, **епічно**, але **круто**.

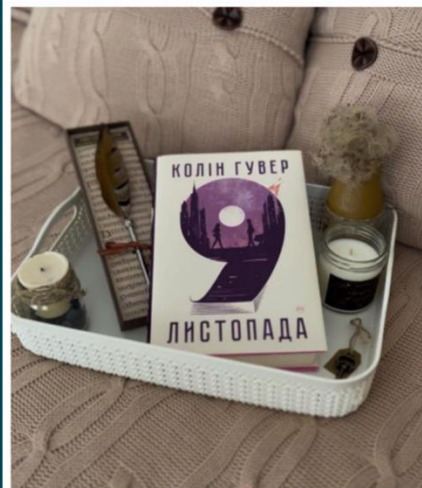
Отож якщо вагаєтесь, чи братись за цю книгу, я без вагань вам скажу, читати її. Вона дійсно хороша та варта уваги. Тепер з **нетерпінням** очікуєм на наступні частини цієї трилогії.



Програмна реалізація

Сторінка з обраним дописом

Результат роботи методу на прикладі визначення негативного емоційного забарвлення допису



Емоційне забарвлення допису

Негативне



hnatko_li

2023-04-25 17:50:55

Думаю варто обґрунтувати свою оцінку. Книга напевно просто не для мене. Я не фанат Гувєр через надмірні травми і проблеми в стосунках, мене це трішки напружає, коли я читаю книги.

Сюжет та задумка мені не сподобались. Я не люблю вільні стосунки, а тут... тут навіть стосунків то не було. Це мені не до душі. Що ж було найбільш неприємне... я знала, що так буде, я бачила спойлери, але читати це було вразі неприємніше. Ну мені просто не в кайф таке читати. І знову ж таки все сталося через їх недомоовки. При тому, що вони ті персонажі, які весь час говорять. Там вина не одного персонажа, винні там усі, але мені від того не легше.

Головні герої. Не можу сказати, що прониклась сильно до них симпатією. Знову ж таки їхні вчинки мені були трохи не зрозумілі...

Не рекомендую



Висновки по виконанню кваліфікаційної роботи бакалавра

Проведено аналіз предметної області

Спроектовано та описано метод визначення емоційного забарвлення дописів за семантичним аналізом

Реалізовано прикладне застосування методу у вигляді соціально орієнтованого вебресурсу

Досліджено ефективність роботи розробленого методу

Ім'я користувача:
Кафедра КН

ID перевірки:
1015353902

Дата перевірки:
31.05.2023 22:10:38 EEST

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
31.05.2023 22:13:12 EEST

ID користувача:
100005671

Назва документа: КН-19-1 Блажук

Кількість сторінок: 71 Кількість слів: 12609 Кількість символів: 99393 Розмір файлу: 2.73 MB ID файлу: 1015021615

Виявлено модифікації тексту (можуть впливати на відсоток схожості)

3.32% Схожість

Найбільша схожість: 1.53% з джерелом з Бібліотеки (ID файлу: 1014963757)

2.82% Джерела з Інтернету 163

Сторінка 73

2.38% Джерела з Бібліотеки 107

Сторінка 74

0% Цитат

Не знайдено жодних цитат

Не знайдено жодних посилань

0% Вилучень

Немає вилучених джерел

Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Замінені символи 5

Підозріле форматування 12 сторінок

Anti-Plagiarism v-15.257**Максимальне співпадіння з одним документом 2.0%**

Словники перевірки: en_US, ru_RU, ua_UA. Помилки в документах: 7%

ID: 114411 Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА Додано в БД: 2023-05-31 Автора: В.Д. Блажук Керівники: О.В. Мазурець Консультанти: Опоненти:	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	83815	1239	3202 (4%)	44 (4%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

РІШЕННЯ ЕКСПЕРНОЇ КОМІСІЇ КАФЕДРИ КОМП'ЮТЕРНИХ НАУК
ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

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

Назва: Метод визначення емоційного забарвлення дописів соціально орієнтованих
вб ресурсів за семантичним аналізом

Автор: студентка групи КН-19-1 Блажук Вікторія Дмитрівна

Спеціальність: 122 – Комп'ютерні науки

Освітня програма: освітньо-професійна

Науковий керівник: к.т.н., доцент кафедри КН Мазурець О.В.

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.	<i>відповідає</i>
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 Блажук В.Д., не є плагіатом, оскільки: запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; серед запозичень знаходяться загальновідомі терміни, скорочення та матеріали оглядових джерел.

Обсяг запозичень, визначений системами виявлення збігів/ідентичності/схожості, складає:

- за системою Anti-Plagiarism: 2%;

- за системою Unichек: 3,32 %.

Керівник роботи

Олександр МАЗУРЕЦЬ

Гарант ОП

Олександр МАЗУРЕЦЬ

Завідувач кафедри КН

Олександр БАРМАК



ВІДГУК НАУКОВОГО КЕРІВНИКА
на кваліфікаційну роботу бакалавра

студентки гр. КН-19-1 Блажук Вікторії Дмитрівни

за темою Метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом

1. Актуальність теми

Актуальність теми обґрунтована, оскільки активне використання соціально орієнтованих вебресурсів призводять до щоденного збільшення кількості контенту у мережі. Створення вебресурсу для аналізу емоційного забарвлення дописів допомагає відстежувати думки та настрої соціуму, особливо під час значних змін у суспільстві.

2. Відповідність роботи предметній області Стандарту спеціальності 122 Комп'ютерні науки

Тема кваліфікаційної роботи відповідає предметній області спеціальності 122 Комп'ютерні науки та вимогам до кваліфікаційної роботи бакалавра, оскільки метою роботи є розробка методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом. При вирішенні поставленої задачі використано математичні моделі, методи та алгоритми розв'язання теоретичних і прикладних задач, що виникають при розробці інформаційних систем.

3. Професійні та особистісні якості бакалавра

Студентка Блажук В. Д. під час роботи над кваліфікаційною роботою бакалавра продемонструвала високий рівень знань та умінь за спеціальністю "Комп'ютерні науки", проявила себе відповідальною та сумлінною студенткою. Студентка опанувала всі необхідні професійні навички за напрямком «Комп'ютерні науки».

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Одержані в роботі результати є наслідком особистої діяльності студентки, яка самостійно виконувала всі поставлені задачі.

5. Ступінь оволодіння методами дослідження

При реалізації кваліфікаційної роботи бакалавра студентка Блажук В. Д. показала достатній рівень компетентностей та володіння необхідними інструментами та обладнанням, методами, методиками та технологіями предметної області комп'ютерних наук.

6. Повнота та якість розкриття теми роботи

Усі поставлені вимоги до роботи виконані в повному обсязі, проведено аналіз актуальності обраної теми, також реалізоване відповідне прикладне застосування розробленого методу.

7. Логічність, послідовність, аргументованість, літературна грамотність викладення матеріалу

Викладення матеріалу логічне, послідовне та аргументоване. Мова і стиль викладення кваліфікаційної роботи бакалавра відповідають стандартам, що забезпечує доступність сприймання матеріалу і відповідає вимогам до сучасних наукових робіт.

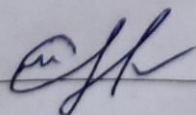
8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин

Реалізоване прикладне застосування розробленого методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом може бути застосоване для визначення емоційного забарвлення дописів (позитивного та негативного) та формування переліку емоційно забарвлених слів допису.

9. Висновок про можливість допуску кваліфікаційної роботи бакалавра 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**».

Керівник



к.т.н., доцент кафедри КН Олександр МАЗУРЕЦЬ



РЕЦЕНЗІЯ

на кваліфікаційну роботу бакалавра

студентки гр. КН-19-1 Блажук Вікторії Дмитрівни

за темою: Метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом

1. Актуальність обраної теми

Розвиток та активне використання соціально орієнтованих вебресурсів призводять до щоденного збільшення кількості контенту в мережі. Кожен написаний та опублікований допис несе в собі певне емоційне забарвлення, і як результат, може впливати на читачів. Тому, аналіз дописів допомагає відстежувати думки та настрої соціуму, особливо під час значних змін у суспільстві.

2. Повнота розкриття мети та завдань роботи

За виконання кваліфікаційної роботи бакалавра мету та завдання роботи було розкрито у повній мірі, зокрема було проведено аналіз предметної області, розроблено та описано кроки методу та реалізовано його прикладне застосування у вигляді соціально орієнтованого вебресурсу.

3. Зміст кожного розділу роботи

Записка кваліфікаційної роботи бакалавра містить три розділи. У першому розділі проведено аналіз предметної області, методів та алгоритмів для вирішення поставлених завдань. У другому розділі розроблено метод визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом та описано його кроки. У третьому розділі описано процес програмної реалізації методу у вигляді соціально орієнтованого вебресурсу з автоматизованим визначенням емоційного забарвлення дописів та досліджено його ефективність.

4. Оцінка розробленої інформаційної системи, її практична цінність

Для розробки прикладного застосування методу визначення емоційного забарвлення дописів соціально орієнтованих вебресурсів за семантичним аналізом було обрано предметну область блогу для написання відгуків на книги. Реалізований вебресурс, який використовує розроблений метод, може бути застосований для визначення емоційного забарвлення дописів (позитивного та негативного) та формування переліку

емоційно забарвлених слів допису.

5. Якість оформлення кваліфікаційної роботи бакалавра

Записка оформлена якісно, згідно з вимогами та правилами. Матеріал викладено логічно та послідовно, з грамотним формулюванням речень.

6. Недоліки кваліфікаційної роботи бакалавра

Рекомендовано розширити навчальну множину дописів й дослідити ефективність методу за його використання з іншими мовами, зокрема англійською.

7. Загальний висновок (допускається чи не допускається до захисту), та оцінка на яку заслуговує кваліфікаційна робота.

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

Рецензент

Дж

Л. Корецька, к.т.н., доцент
кадр. АРІТ та Р