

Хмельницький національний університет
Факультет інформаційних технологій
Кафедра комп'ютерних наук

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

на тему Метод підбору подібних за настроєм коротких художніх текстів
засобами обробки природної мови

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

Виконав: _____ студент групи КН-20-2 _____ Артем БАБІЙ _____
Група виконавця Підпис Ім'я, ПРІЗВИЩЕ

Керівник: _____ к.т.н., доц. каф. КН _____ Олександр ПАСІЧНИК _____
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

Нормоконтроль: _____ к.т.н., доц. каф. КН _____ Руслан БАГРІЙ _____
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

До захисту допускаю:

Зав. кафедри КН, д.т.н., професор _____ Олександр БАРМАК _____
Підпис Ім'я, ПРІЗВИЩЕ

20 06 2024 р.

Хмельницький 2024

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

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних наук



(підпис)

д.т.н., професор Олександр БАРМАК

«16» Метого 2024 року

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА

1. Тема кваліфікаційної роботи бакалавра: «Метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови»
2. Завдання видано студенту Артему БАБІЮ
(ім'я, прізвище)
3. Керівник роботи доцент кафедри КН Олександр ПАСІЧНИК
(посада, ім'я, прізвище)
4. Затверджено наказом університету від «15» Метого 2024 р. № 8
5. Дата видачі завдання студенту: «16» Метого 2024 р.
6. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Мета роботи – спрощення роботи з літературними сервісами за рахунок автоматизованого підбору подібних за настроєм текстів. Для досягнення мети необхідно провести аналіз предметної області семантичного аналізу текстів за емоційним настроєм. Виконати аналіз існуючих програмних рішень та наукових публікацій для вирішення подібних задач. Розробити метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови. Виконати підготовку вхідних даних для навчання. Програмно реалізувати розроблений метод у вигляді вебресурс для обміну літературними творами. Використовуючи реалізований вебресурс здійснити дослідження ефективності роботи розробленого методу.

7. Календарний план виконання кваліфікаційної роботи бакалавра:

№	Назва етапів (розділів) кваліфікаційної роботи бакалавра	Термін виконання	Примітка
1	Вибір напряму дослідження та узгодження тематики кваліфікаційної роботи бакалавра з керівником, складання календарного графіка виконання роботи	січень 2024	Виконав
2	Ознайомлення з предметною областю, формулювання мети та задач дослідження, визначення об'єкта та предмета дослідження	лютий 2024	Виконав
3	Проектування та розробка загальної архітектури програмного забезпечення, інтерфейсу користувача, вибір засобів реалізації програмного забезпечення	березень 2024	Виконав
4	Створення та тестування програмного забезпечення	квітень 2024	Виконав
5	Написання пояснювальної записки, урахування зауважень керівника, оформлення згідно вимог	травень 2024	Виконав
6	Розробка презентаційних матеріалів та попередній захист кваліфікаційної роботи	травень 2024	Виконав
7	Отримання відгуку керівника, рецензії, перевірка на плагіат, нормоконтроль	червень 2024	Виконав
8	Підготовка до захисту та захист кваліфікаційної роботи бакалавра	червень 2024	Виконав

Виконавець: студент групи КН-20-1

Група виконавця


Підпис

Артем БАБІЙ

Ім'я, ПРІЗВИЩЕ

Керівник: к.т.н., доц. каф. КН

Науковий ступінь, посада


Підпис

Олександр ПАСІЧНИК

Ім'я, ПРІЗВИЩЕ

Анотація

Тема кваліфікаційної роботи бакалавра: «Метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови»

Виконавець кваліфікаційної роботи бакалавра: студент групи КН-20-2
Артем БАБІЙ

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

Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
84	43	18	27	4

Метою кваліфікаційної роботи бакалавра є спрощення роботи з літературними сервісами за рахунок автоматизованого підбору подібних за настроєм текстів. Для досягнення поставленої мети було розроблено метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови з його програмною реалізацією у вигляді вебресурсу з обміну літературними творами.

Реалізована інформаційна система дає змогу користувачам реєструватись на вебресурсі та публікувати власні літературні твори, для яких автоматично визначаються оцінки прояву емоційних настроїв. Використовуючи реалізований метод, користувачі можуть проводити пошук подібних за настроєм творів до обраного твору серед числа опублікованих на вебсервісі. Крім цього, адміністратор вебсервісу має доступ до управління методом підбору подібних за настроєм текстів шляхом внесення змін в множину навчальних текстів.

Ключові слова: емоційний настрій тексту, обробка природної мови, вебресурс, літературні твори.

Виконавець: студент групи КН-20-2
Група виконавця


Підпис

Артем БАБІЙ
Ім'я, ПРІЗВИЩЕ

Зміст

Перелік скорочень	4
Вступ.....	5
Розділ 1 Характеристика предметної області семантичного аналізу текстів за настроєм: аналіз моделей, методів та реалізацій	7
1.1 Аналіз інформаційних моделей.....	7
1.2 Огляд теоретичних підходів до розв’язку задач порівняння текстів за настроєм	11
1.3 Аналіз існуючих програмних засобів та наукових рішень щодо порівняння художніх текстів за настроєм.....	15
1.4 Мета, задачі та вимоги до реалізації інформаційної системи	18
Розділ 2 Проектування вебресурсу з підбору подібних за настроєм художніх текстів	20
2.1 Кроки методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови.....	20
2.2 Аналіз та автоматизація обробки потоків даних вебресурсу з обміну літературними творами.....	25
2.3 Функціональна структура вебресурсу з обміну літературними творами	28
2.4 Проектна архітектура вебресурсу з обміну літературними творами та взаємозв’язок компонентів.....	31
2.5 Проектування даталогічної моделі бази даних вебресурсу з обміну літературними творами.....	34
2.6 Підготовка словників для подання категорій настрою художніх текстів....	41
2.7 Особливості використання спеціалізованих програмних компонентів	43
2.8 Визначення шляхів дослідження та засобів створення програмного забезпечення	43
2.9 Висновки до розділу 2	45
Розділ 3 Програмна реалізація вебресурсу з підбору подібних за настроєм художніх текстів	47

3.1 Вибір засобів розробки вебресурсу з підбору подібних за настроєм художніх текстів.....	47
3.2 Структура та функціональне призначення програмних складових вебресурсу з підбору подібних за настроєм художніх текстів.....	48
3.3 Особливості реалізації програмних складових вебресурсу з підбору подібних за настроєм художніх текстів.....	53
3.4 Тестування вебресурсу та вимоги до розгортання	57
3.6 Аналіз функціональності вебресурсу з підбору подібних за настроєм художніх текстів.....	65
3.7 Результати досліджень ефективності методу підбору подібних за настроєм коротких художніх текстів.....	73
3.8 Висновки до розділу 3	78
Загальні висновки.....	80
Перелік посилань.....	82
Додатки	

Перелік скорочень

Скорочення, термін, позначення	Пояснення
ІС	Інформаційна система
КРБ	Кваліфікаційна робота бакалавра
КН	Комп'ютерні науки
ПЗ	Пояснювальна записка
СКБД	Система керування базами даних
ХНУ	Хмельницький національний університет
NLP	Natural Language Processing
SVM	Support Vector Machine
RNN	Recurrent Neural Network
CNN	Convolutional Neural Networks
TF	Term Frequency
TF-IDF	Term Frequency – Inverse Document Frequency
ІІ	Штучний інтелект
GPT	Generative Pre-trained Transformer
БД	База даних
ID	Identity Document
FK	Foreign Key
PHP	Hypertext Preprocessor
ML	Machine Learning
MVC	Model View Controller
SQL	Structured query language

Вступ

Мета кваліфікаційної роботи бакалавра – спрощення роботи з літературними сервісами за рахунок автоматизованого підбору подібних за настроєм текстів. Для досягнення поставленої мети необхідно розробити метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови та програмно реалізувати у вигляді вебресурсу з обміну літературними творами.

Актуальність. В умовах збільшення інформації та великої кількості літературних творів існує попит на автоматизовані системи пошуку подібних творів з емоційним настроєм. Це може допомогти як читачам, для пошуку творів для читання, подібних до їхніх улюблених творів. Так і письменникам для пошуку натхнення та ідей, враховуючи настрої творів над якими вони працюють. Розробка методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови допоможе покращити досвід користувача під час пошуку подібних творів та релевантних рекомендацій.

Об'єкт дослідження – процес підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови.

Предмет дослідження – методи машинного навчання для роботи з текстовою інформацією.

Мета кваліфікаційної роботи бакалавра – спрощення роботи з літературними сервісами за рахунок автоматизованого підбору подібних за настроєм текстів.

Завдання кваліфікаційної роботи бакалавра. Провести аналіз предметної області визначення семантичного аналізу текстів за настроєм. Виконати огляд теоретичних підходів щодо порівняння текстів за настроєм. Виконати аналіз існуючих програмних рішень та наукових публікацій для вирішення подібних задач. Розробити метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови. Виконати підготовку вхідних даних для навчання. Спроекувати структуру інформаційної

системи та базу даних для збереження необхідної інформації. Виконати вибір засобів розробки для програмної реалізації методу. Програмно реалізувати розроблений метод у вигляді вебресурс для обміну літературними творами. Використовуючи реалізований вебресурс здійснити тестування інформаційної системи та провести дослідження ефективності роботи розробленого методу.

Розділ 1 Характеристика предметної області семантичного аналізу текстів за настроєм: аналіз моделей, методів та реалізацій

1.1 Аналіз інформаційних моделей

З розвитком багатьох сфер людського життя, письмо стало невід'ємним інструментом для збереження і передачі інформації та знань. Використовуючи письмо люди змогли вести облік та покращувати торгівлю, писати закони для врегулювання суспільства, укладати договори між державами, документувати важливі історичні події або просто записувати свої думки. Розвиток писемності впливав і на розвиток культури, адже, це давало змогу зберігати та поширювати культурну спадщину не лише через великі відстані а й через роки. Таким чином, тексти стали одним із засобів комунікації та передачі інформації і знань між людьми.

З наукової точки зору, текст – це послідовність пов'язаних між собою знаків, які передають певний сенс та несуть якусь інформацію [1]. Текст складається із послідовності речень, пов'язаних між собою змістом і побудованих відповідно до правил мовної системи в якій цей текст написаний. В свою чергу, речення складаються із послідовності слів. Тому, саме окремі слова є тими цеглинками з яких будується текст.

Отже, основними елементами з яких будується текст є [2]:

- слова, як базова одиниця мовлення, які мають власне значення та поєднуються з іншими словами утворюючи речення;
- речення, як групи слів, поєднаних між собою певною думкою відповідно до граматичних правил;
- абзаци, як групи речень, поєднаних між собою ідеєю задля організації структури тексту;
- розділи, які дозволяють розбивати великий текст на частини, пов'язані однією темою або ідеєю.

Аналізуючи елементи з яких складаються тексти, стає зрозуміло, що текст – це щось більше ніж просто послідовність знаків. Кожен текст наділений рядом ознак, які його характеризують. До основних ознак тексту можна віднести [3]:

- послідовність, адже текст складається з конкретної послідовності слів, які формують речення, які в свою чергу формують абзаци та розділи;
- зв'язність, яка характеризується розвитком теми тексту;
- цілісність, яка забезпечує повний виклад змісту тексту;
- інформативність, адже тексти призначені для збереження та передачі певної інформації;
- організованість, яка характеризується поділом тексту на абзаци та розділи для полегшення сприйняття його змісту;
- завершеність, яка характеризується тим, наскільки завершеною є основна думка чи ідея тексту.

Крім цього, тексти можна класифікувати на різні типи в залежності від їхніх жанрів, стилів, змісту, особливостей побудови та призначення. Наприклад, в українській мовній системі за стилем тексти діляться на 5 типів: розмовний, офіційно-діловий, науковий, публіцистичний та художній [4]. Стиль тексту залежить від засобів виразності, за допомогою яких, автори тексту передають основний його зміст. Також, різні стилі тексту призначені для різних аудиторій, тому одна й та сама тема може бути висвітлена зовсім по різному, в залежності від обраного стилю. Наприклад, інформація пов'язана із стихійними явищами у науковому стилі буде висвітлювати причини виникнення цих явищ з точки зору науки, в той час як в художньому стилі може міститись детальний опис відчуттів людини, яка зіткнулась із природним лихом.

Художній стиль текстів найбільше відрізняється від усіх інших тим, що він тісно пов'язаний із творчістю та уявою. Головна мета художніх текстів – викликати в читача певні емоції та естетичне задоволення від прочитаного. Отже, художній стиль текстів – це набір мовних засобів для передачі думок, ідей та образів, здатних викликати у людини почуття, емоції та переживання [5].

Тому, можна сказати, що основна мета художніх текстів – це вплив на емоційний стан читача з метою виклику певних почуттів.

Художні тексти можна розділити на три групи, відповідно до існуючих типів настрою: оптимістичні, нейтральні та песимістичні.

Оптимістичний тип настрою пов'язаний із станом емоційного піднесення людини, який характеризується вірою в позитивний перебіг подій. Зазвичай, оптимістичні люди наділені більшою людяністю та вірою в добро, що спонукає їх творити більше добрих вчинків [6]. Оптимістичні художні тексти, викликають у читача позитивні емоції пов'язані із радістю, задоволенням, захопленням, вірою в добро. Часто, такі тексти наповнюють читача енергією та мотивують його до хороших вчинків, які можуть принести користь людині та її оточенню. Автор тексту може передавати оптимізм через яскраві описи природи, хороші події та внутрішній стан героїв. Наприклад, опис сонячного літнього ранку, який викликає радість у головного героя та мотивує його прожити цей день якнайкраще, неодмінно викличе подібні емоції й у читача.

Нейтральний тип настрою характеризується відсутністю переживання яскраво виражених позитивних чи негативних емоцій. Нейтральний настрій може переважати тоді, коли у житті людини не відбуваються якісь яскраві події [7]. Нейтральний настрій – це стан душевного спокою, гармонії та стабільності. Це сіра зона, в якій емоційні ваги людини знаходяться в рівновазі. Нейтральні художні тексти не викликають у читача якихось конкретних емоцій. Вони можуть використовуватись для наведення певної інформації з додаванням наукових фактів для опису об'єктивної реальності без впливу на емоційний стан читача. Наприклад, у фантастичній історії про космічні подорожі інформація про історію виникнення планети може бути подана в нейтральній формі та не викликати у читача емоцій, а лише інформувати про побудову всесвіту.

Песимістичний тип настрою, зазвичай, характеризується пригніченим емоційним станом. Песимісти схильні ставитись до життя негативно, більше сумувати та злитись. Також, песимісти більше зосереджені на несприятливих ситуаціях та схильні шукати мінуси навіть там де їх немає [8]. Песимістичні

художні тексти здатні викликати у читача негативні емоції такі як сум, злість, роздратування чи огида. Автор може викликати песимістичний настрій у читача описуючи невдачі героїв та створюючи складні життєві ситуації з якими важко впоратись. Також, описи похмурих, темних, занедбаних, брудних локацій можуть викликати в читача асоціації з чимось поганим та песимістичним.

Настрій в художніх текстах може проявлятися за допомогою різних засобів та прийомів, які використовують автори задля впливу на емоційний стан читача. Найпростіший метод впливу на емоції читача – це наповнення тексту емоційно забарвленими словами. Наприклад, слова і фрази, які виражають оптимізм можуть бути наступними: «добре», «чудово», «світле майбутнє», «віра в краще». Протилежні за сенсом слова та фрази будуть викликати песимістичні емоції. Крім цього, автори часто використовують добре відомі образи та метафори для передачі настрою тексту. Таким чином, опис похмурого дощового дня, який зруйнував чийсь плани може посилити відчуття песимізму в читача [9].

Отже, тексти – це особливий засіб збереження інформації та комунікації між людьми, який має складну структуру та особливості. В залежності від призначення, тексти класифікуються на різні стилі, що дозволяє розглядати різні варіанти передачі однієї і тієї ж інформації. Особливістю художніх текстів є їхня спрямованість на вплив та маніпуляцію емоційним настроєм читача. Таким чином, автори текстів можуть виражати оптимістичний, песимістичний чи нейтральний настрій через відповідні слова, фрази, образи та метафори, та викликати аналогічний емоційний стан у читачів. Тому, аналізуючи структуру та особливості текстів, можна зробити висновок, що текст – це не просто набір знаків чи слів, це складна форма комунікації, яка здатна впливати на сприйняття та емоційні відчуття людей.

1.2 Огляд теоретичних підходів до розв'язку задач порівняння текстів за настроєм

На сучасному етапі розвитку писемності, тексти стали чимось більшим ніж просто засіб збереження та передачі інформації між людьми, адже тепер тексти – це один із засобів комунікації між людиною та комп'ютером. Проте, для розуміння тексту, створеного людиною, інформаційній системі потрібно провести ряд складних операцій над текстом. Такий процес називається обробкою природної мови або NLP.

NLP (Natural Language Processing) – це один із напрямків у галузі комп'ютерних наук, пов'язаний із розробкою та застосуванням методів та систем автоматизованого аналізу природної мови [10]. Природна мова – це засіб спілкування, виключно між людьми, для передачі інформації, ідей, думок та почуттів. Отже, основна мета NLP – це зрозуміти природну мову.

NLP включає в себе досить багато різноманітних завдань, серед яких основними є:

- розпізнавання мовлення та тексту;
- генерація мовлення та тексту;
- машинний переклад;
- аналіз семантики та синтаксису тексту для розуміння його значення;
- визначення та класифікація іменованих сутностей таких як імена, дати, організації, власні назви;
- аналіз тональності тексту;
- сумаризація тексту, тобто стисле узагальнення змісту тексту.

Класифікація текстів за настроєм засобами NLP полягає у визначенні емоційного забарвлення текстів, такого як оптимістичне, нейтральне та песимістичне. Вирішення даної проблеми може бути корисним при аналізі відгуків користувачів соціальних мереж, новин, статей та іншої текстової інформації.

Для класифікації текстів за настроєм можуть застосовуватись моделі машинного навчання та класифікатори, які заздалегідь навчені на спеціально підготовлених даних. До таких моделей можна віднести наївний Баєсів класифікатор або метод опорних векторів.

Основою наївного Баєсового класифікатора є Теорема Баєса, яка базується на припущенні про незалежність між ознаками даних. Тобто, наявність однієї ознаки в наборі даних ніяк не впливає на наявність інших ознак. Перевагою використання цього класифікатора є його простота в реалізації та висока ефективність у випадку малої кількості даних. Проте, припущення про незалежність ознак може погіршити результати класифікації у випадку текстових даних, адже слова в тексті взаємопов'язані [11].

Метод опорних векторів або SVM (Support Vector Machine) є потужним алгоритмом для вирішення завдань класифікації, зокрема класифікації текстів за настроєм. Цей алгоритм відображає дані в багатовимірному просторі ознак так щоб дані можна було категоризувати, навіть якщо ці дані лінійно не подільні іншими методами. Після того, як виявляється роздільник між категоріями, дані перетворюються таким чином, щоб роздільник можна було зобразити як гіперплощину. Таким чином, за характеристиками нових даних можна передбачити до якої категорії вони належать [12].

Крім методів машинного навчання, для вирішення задачі класифікації текстів за настроєм, також використовуються методи глибокого навчання. До методів глибокого навчання належать: рекурентні нейронні мережі, згорткові нейронні мережі та нейронні мереже трансформери.

Рекурентні нейронні мережі або RNN (Recurrent Neural Network) застосовуються для обробки лінійних даних до яких належать текстові дані. Основною особливістю RNN є те, що вхідні дані попереднього кроку подаються як вхідні дані для наступного кроку (Рисунок 1.1). Звідси й назва «рекурентні», що також означає «повторювані». Такий підхід ефективно себе показує у випадку, коли потрібно передбачити наступне слово у реченні, адже для цього потрібно знати попередні слова, а отже, є потреба в їхньому запам'ятовуванні.

RNN вирішує цю проблему за рахунок прихованого шару, який запам'ятовує інформацію про послідовність даних [13].

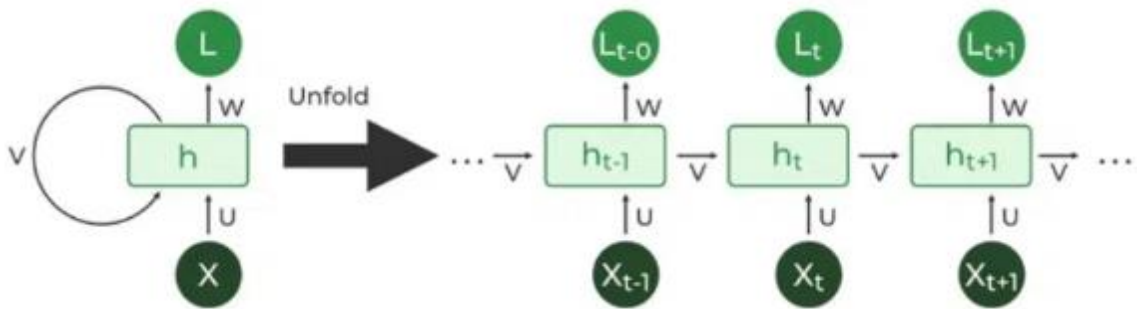


Рисунок 1.1 – Архітектура рекурентної нейронної мережі [13]

Згорткові нейронні мережі або CNN (Convolutional Neural Networks) часто використовуються для розпізнавання тексту на зображеннях. Згорткова нейронна мережа складається з кількох шарів: вхідний шар, згортковий шар, шар об'єднання, шар щільності та вихідний шар (Рисунок 1.2). Згортковий шар застосовує фільтри до вхідного зображення, щоб виділити з нього особливості. Шар об'єднання зменшує дискретизацію зображення, щоб зменшити кількість обчислень та навантаження на обчислювальну систему. Після всіх обчислень, шар щільності робить остаточний прогноз. Згорткова нейронна мережа знаходить оптимальні фільтри для згорткового шару через зворотнє поширення помилок та градієнтний спуск [14].

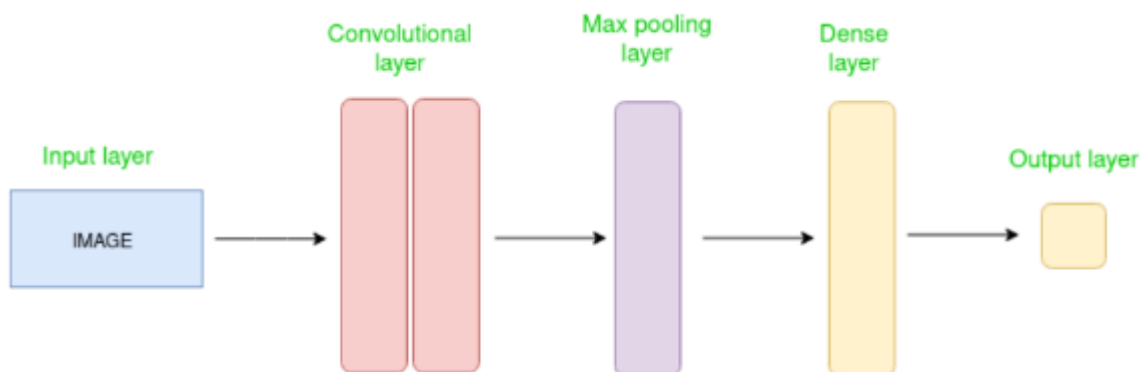


Рисунок 1.2 – Архітектура згорткової нейронної мережі [14]

Ще одним поширеним підходом для класифікації текстів за настроєм є використання словників тональності. Словники тональності – це набори слів із визначеним класом емоційного забарвлення. Перевагою використання таких словників є те, що слова в них мають визначену оцінку емоційності, що спрощує процес самостійного визначення емоційності слів [15].

Оскільки, комп'ютери не здатні розуміти звичайний текст, то перед використанням будь якого із підходів необхідно провести векторизацію тексту. Векторизація тексту – це процес перетворення тексту в числовий формат, де кожне слово представлене числом. Зазвичай, для кожного слова обчислюється числова оцінка значимості слова в тексті. Для вирішення цієї задачі можна використати статистичний метод TF-IDF (Term Frequency - Inverse Document Frequency). Метод TF-IDF складається з двох частин: TF – частотної оцінки та IDF – зворотної документарної частоти [16].

Частотна оцінка обчислює як часто кожне унікальне слово зустрічається в тексті. Чим більше входжень одного і того ж слова у тексті, тим вищою буде його вага. Формула для обчислення TF має наступний вигляд:

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

де TF_i – унікальне слово із тексту, n_i – кількість зустрічань цього слова в тексті, $\sum_k n_{ik}$ – загальна кількість слів в тексті.

В текстах є багато слів, які не несуть жодного смислового значення й використовуються лише для зв'язності оповіді. Тому, виникає потреба в пошуку та відсіюванні таких слів. Для цього необхідно використовувати альтернативні тексти, тобто тексти відмінні за тематикою від тих, для яких обчислювалась оцінка TF, адже загальноживані слова зустрічатимуться в усіх текстах, незалежно від тематики. Отже, через обчислення IDF можна зменшити значимість загальноживаних слів за такою формулою:

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

де IDF_i – унікальне слово, для якого обчислюється інверсна документарна частота, D – загальна кількість текстів у вибірці, d_i – кількість текстів у яких присутнє дане слово.

Таким чином, обчислюючи оцінку значимості слів для текстів можна швидко векторизувати текст для подальшого аналізу та класифікації за настроєм. В ході аналізу існуючих методів для вирішення задачі визначення емоційного настрою текстів прийнято рішення використати словники тональності, адже це зручний та надійний інструмент, який дозволить отримати більш точні результати.

1.3 Аналіз існуючих програмних засобів та наукових рішень щодо порівняння художніх текстів за настроєм

Щоб провести порівняння та пошук подібних за настроєм коротких художніх текстів, спочатку необхідно вирішити завдання визначення емоційного настрою у цих текстах. Вирішуючи задачу визначення емоційного настрою у текстах методами обробки природної мови, автори наукових робіт підходять з точки зору вирішення задачі класифікації текстів.

Автор наукової роботи на тему «Дослідження методів аналізу емоційного забарвлення тексту українською мовою» Рябишев О. В., у своїй роботі описав найбільш ефективні методи аналізу емоційного забарвлення текстів для української мови [17]. В ході дослідження було розглянуто підхід, заснований на словниках настроїв, підхід, заснований на правилах та підхід з використанням машинного навчання. Для кожного з підходів було визначено переваги та недоліки. Перевагами підходу з використанням словників є те, що даний метод не потребує навчання, проте знайти хороший словник може бути проблемою, а створювати його самостійно – це дуже довгий і кропіткий процес, який потребує поглиблених лінгвістичних знань. Підхід заснований на правилах може добре працювати в межах певного домену, але для високої точності роботи необхідна велика кількість правил, які необхідно прописувати вручну. Використання

машинного навчання може надати досить хороші результати класифікації, але цей підхід вимагає попереднього навчання класифікатора на якісному наборі даних, адже від цього залежить подальша якість класифікації.

Автори Біленький В. С., Кобзев В. Г., Матвеев Д. І. у своїй роботі присвяченій темі «Методи оцінювання тонального забарвлення тексту» розглянули методи, що базуються на методах найближчих сусідів, опорних векторів та наївному баєсівському класифікаторі [18]. В ході дослідження було встановлено, що для поліпшення процесу автоматизованого визначення тональності тексту варто використовувати поєднання декількох методів одночасно, адже це дає значно кращі результати класифікації. Також, для покращення роботи класифікаторів варто забезпечити автоматичну перевірку та виправлення орфографічних помилок, які можуть ускладнювати отримання якісних результатів. Щодо методів, заснованих на словниках тональності потрібно вдосконалювати та розширювати словникову базу. Теж саме стосується навчальних датасетів для методів машинного навчання.

У вільному доступі можна знайти готові сервіси, здатні класифікувати тексти за настроєм. Одним із таких сервісів є семантичний аналізатор тексту від «MonkeyLearn» [19]. Даний сервіс аналізує текст та надає користувачеві оцінку позитивності чи негативності тексту у відсотках. На рисунку 1.3 зображено спробу визначити настрій позитивного тексту.

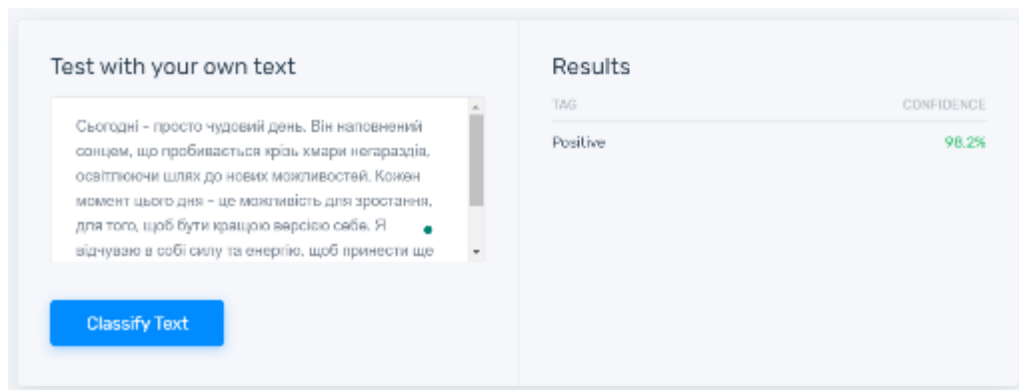


Рисунок 1.3 – Визначення настрою тексту через сервіс «MonkeyLearn» [19]

Сервіс для аналізу текстів від «MonkeyLearn» дуже простий та зручний у використанні, що є його основною перевагою. Недоліком даного сервісу можна

назвати те, що невідомо за якими алгоритмами та методами відбувається аналіз текстів. Проте, сервіс показує хороші результати при визначенні емоційного настрою текстів.

Іншим зручним та досить популярним варіантом для визначення настрою тексту є використання ChatGPT [20]. ChatGPT здатен аналізувати контекст тексту та надавати розгорнуту відповідь на запит користувача (Рисунок 1.4). З однієї сторони це досить зручно, адже можна отримати додатковий коментар по тексту від ШІ, а з іншої – це надмірна та зайва інформація.

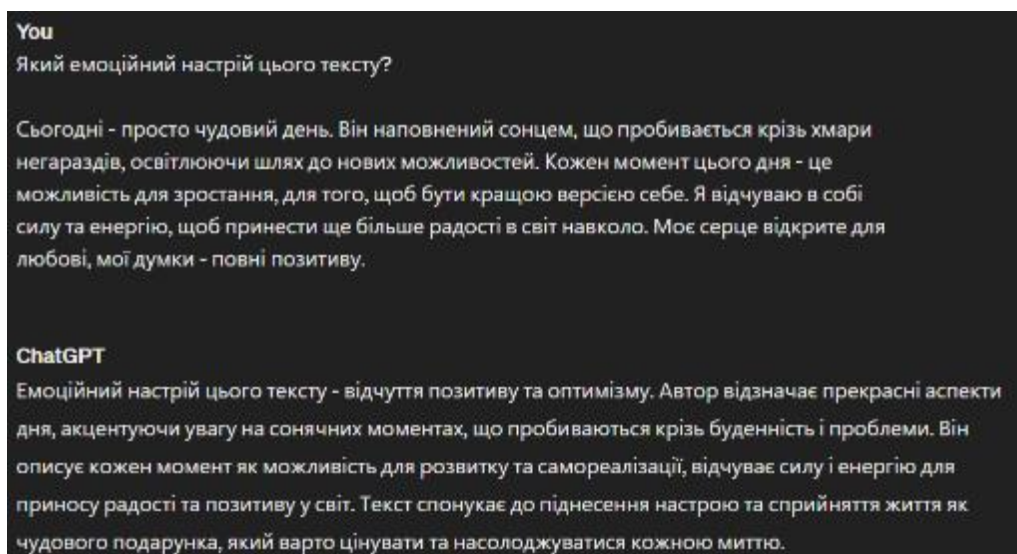


Рисунок 1.4 – Визначення настрою тексту з використанням ChatGPT [20]

У своїй роботі ChatGPT використовує модель GPT (Generative Pre-trained Transformer), основою якої є архітектура трансформерної нейронної мережі. Трансформерні нейронні мережі – це, відносно, нові підходи до вирішення задач обробки природної мови. Модель GPT тренувана на величезному об’ємі даних, тому здатна добре розуміти та генерувати текст, схожий на природну мову.

Отже, дослідження методів для вирішення зада обробки природної мови є актуальним завданням, адже вдосконалення існуючих методів допоможе отримати кращі результати в даній області. Аналіз існуючих наукових рішень щодо вирішення задачі, яка полягає в визначенні емоційного настрою текстів зводиться до того, що необхідно вирішити задачу класифікації текстів. Варто

відмітити, що використання словників тональності досить ефективний спосіб визначення настрою тексту у випадку, якщо словник якісно сформований. Тому, для визначення настрою коротких художніх текстів з метою подальшого пошуку подібних за настроєм текстів, прийнято рішення використати словники тональності.

Крім цього, створення сервісу для аналізу текстів у вигляді вебзастосунку досить вдале рішення, адже це надає користувачам вільний доступ до програми без потреби завантажувати додаткове програмне забезпечення.

1.4 Мета, задачі та вимоги до реалізації інформаційної системи

Метою кваліфікаційної роботи бакалавра є спрощення роботи з літературними сервісами за рахунок автоматизованого підбору подібних за настроєм текстів. Для досягнення мети слід розв'язати наступні задачі:

- виконати дослідження предметної області семантичного аналізу текстів за настроєм;
- у рамках дослідження предметної області виконати огляд теоретичних підходів щодо порівняння текстів за настроєм;
- виконати аналіз наукових надбань й існуючих програмних рішень у області порівняння художніх текстів за настроєм;
- розробити метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови;
- сформувати вектори значущих слів для подання типів настрою художніх текстів;
- виконати підготовку навчальних художніх текстів для обрахунку значень оцінок важливості слів словників для типів настрою текстів;
- на основі розробленого методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови виконати проектування структури інформаційної системи;

- спроектувати даталогічну модель бази даних до вебресурсу з обміну літературними творами;
- здійснити вибір засобів розробки для вебресурсу з обміну літературними творами;
- здійснити програмну реалізацію вебресурсу з обміну літературними творами;
- провести тестування розробленого вебресурсу;
- здійснити дослідження ефективності розробленого методу з використанням створеного вебресурсу з обміну літературними творами.

Створений вебресурс з обміну літературними творами має використовувати розроблений метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови й забезпечувати виконання таких функцій:

- реєстрація та авторизація користувачів;
- робота користувачів з персональною сторінкою;
- публікація власних літературних творів;
- перегляд оцінок прояву емоційних настроїв опублікованих творів;
- пошук подібних творів за емоційним настроєм;
- написання коментарів до опублікованих творів;
- можливість редагувати множину навчальних текстів адміністратором вебресурсу;
- перегляд адміністратором вебресурсу проміжних даних роботи методу у вигляді векторів значущих слів для типів настрою;

Результатом досягнення поставленої мети є вебресурс з обміну літературними творами, який використовує розроблений метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови та надає користувачам можливість перегляду оцінок прояву емоційного настрою для опублікованих творів та пошук подібних творів за емоційним настроєм.

Розділ 2 Проектування вебресурсу з підбору подібних за настроєм художніх текстів

2.1 Кроки методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови

Метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови складається із трьох послідовних етапів (Рисунок 2.1).

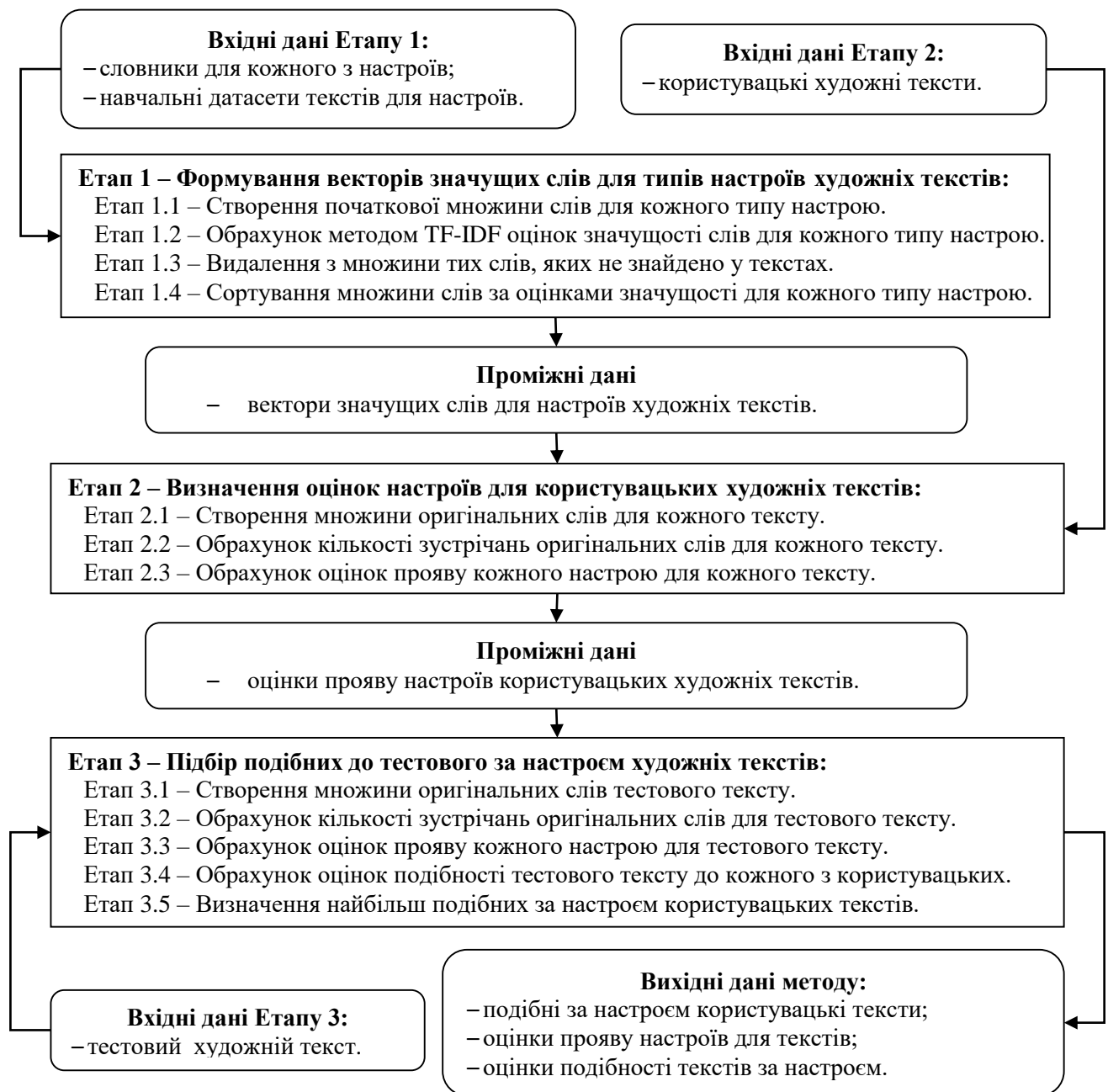


Рисунок 2.1 – Схема методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови

На першому етапі відбувається формування векторів значущих слів для кожного типу настрою. Вхідними даними для цього етапу є: словники тональності для оптимістичного, нейтрального та песимістичного настроїв та навчальні датасети текстів для кожного з вказаних типів настрою. Таким чином, для роботи першого етапу потрібно використати три словники, тобто три переліки слів з визначеним емоційним забарвленням та три набори коротких художніх текстів із визначеним емоційним настроєм (Рисунок 2.2).



Рисунок 2.2 – Схема роботи етапу формування векторів значущих слів для типів настроїв художніх текстів

На першому кроці даного етапу відбувається створення початкової множини слів для кожного із трьох наборів текстів. Цей крок включає в себе

попередню обробку тексту, яка необхідна для покращення результатів роботи методу. Попередня обробка тексту включає в себе видалення із тексту розділових знаків, символів, цифр та стоп-слів. Також, на даному етапі усі слова варто привести до базової форми з метою уникнення повторів.

Далі, використовуючи метод TF-IDF обчислюється оцінка значущості для кожного слова із множин. Оскільки, TF-IDF потребує використання альтернативних текстів, то у випадку обрахунку оцінок значущості для оптимістичних текстів, нейтральні та песимістичні тексти будуть альтернативними (аналогічно для нейтральних та песимістичних).

Вихідними даними першого етапу є вектори значущих слів для кожного типу настрою, які формуються шляхом перетину множини слів із словників тональності та множини слів з оцінками значущості. Тобто, якщо слово із множини слів оцінок значущості присутнє у словнику тональності – воно потрапляє у вектор значущих слів для даного типу настрою.

На другому етапі відбувається визначення оцінок настроїв для користувацьких художніх текстів, тобто текстів з числа яких пізніше буде проводитись пошук подібних за настроєм до тестового тексту. Вхідними даними для цього етапу є вектори значущих слів для кожного типу настрою, отримані на попередньому етапі, та набір користувацьких текстів (Рисунок 2.3).

Спочатку, для кожного тексту з набору формується множина оригінальних слів. Далі, підраховується кількість зустрічань кожного оригінального слова в тексті. Таким чином, для кожного тексту формується перелік слів та число входжень кожного слова у текст.

На наступному кроці другого етапу відбувається обрахунок оцінок прояву кожного типу настрою для кожного тексту. Тобто, якщо слово є в множині оригінальних слів користувацького художнього тексту і в множині векторів значущих слів для певного типу настрою, то оцінку прояву певного типу настрою E можна обчислити за формулою (2.1).

$$E = \sum_{i=1}^n N_i * W_i \quad (2.1)$$

де N_i – кількість зустрічань слова із множини оригінальних слів користувацьких художніх текстів, W_i – вага слова у векторі значущих слів для конкретного типу настрою, n – загальна кількість слів у множині оригінальних слів користувацьких художніх текстів.

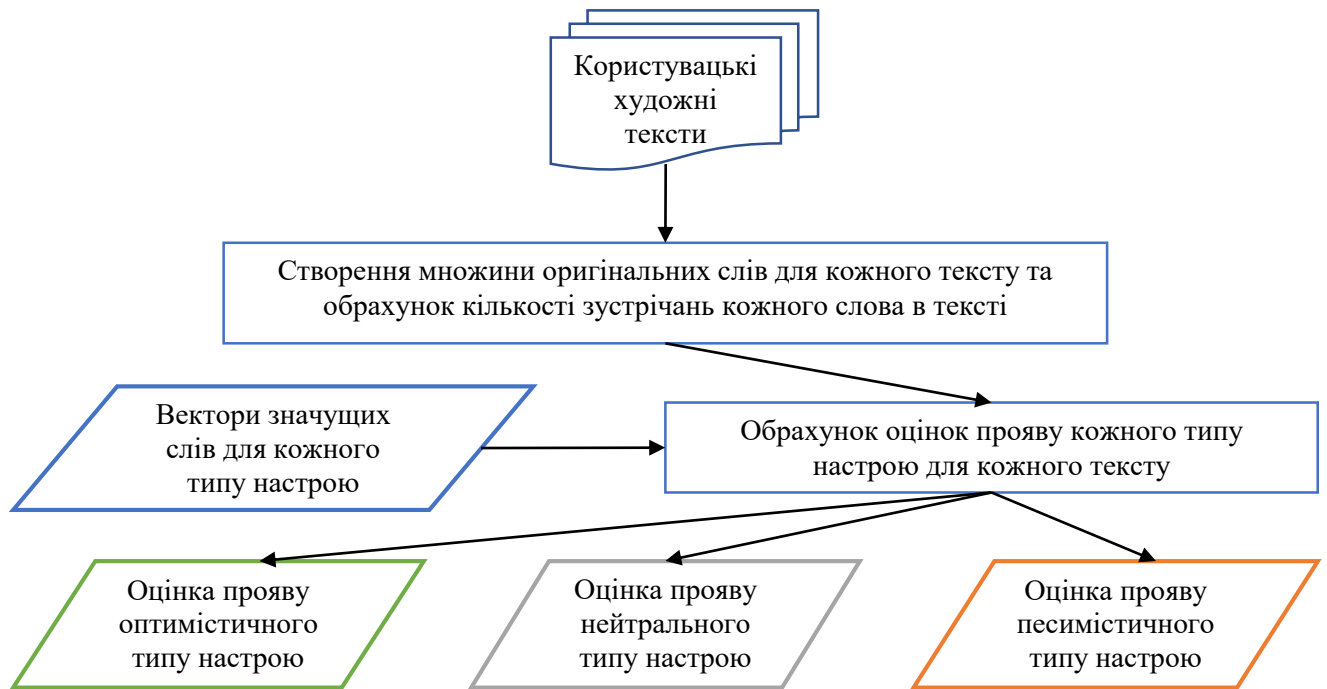


Рисунок 2.3 – Схема роботи етапу визначення оцінок настроїв для користувацьких художніх текстів

Таким чином, для кожного користувацького художнього тексту буде отримано значення трьох оцінок прояву настрою: оптимістичної, нейтральної та песимістичної оцінки прояву настрою в тексті. Саме ці дані будуть вихідними даними другого етапу методу на якому відбувається визначення оцінок настроїв для користувацьких художніх текстів.

На завершальному третьому етапі відбувається підбір подібних до тестового за настроєм художніх текстів. Вхідними даними для цього етапу є тестовий текст та множина оцінок прояву настроїв для користувацьких художніх текстів, отримана на другому етапі (Рисунок 2.4).



Рисунок 2.4 – Схема роботи етапу підбору подібних до тестового за настроєм художніх текстів

Спочатку проводить аналіз тестового тексту за тим же алгоритмом, що й для користувацьких текстів: створюється множина оригінальних слів тексту та обраховується кількість зустрічань кожного слова в тексті. Далі, обраховується оцінка прояву кожного типу настрою для тестового тексту за формулою (2.1). Тепер, відбувається обрахунок оцінок подібності тестового тексту із кожним користувацьким текстом. Отриманий перелік оцінок сортується від найбільшої до найменшої. Отже, перелік подібних за настроєм користувацьких текстів з оцінками прояву настроїв для цих текстів та оцінками подібності текстів із тестовим текстом є вихідними даними роботи методу.

Отже, було розроблено метод підбору подібних за настроєм коротких художніх текстів засобами природної мови, який призначений для одержання результатів у вигляді переліку подібних за настроєм користувацьких текстів з оцінками прояву настроїв для цих текстів та оцінками подібності текстів із

тестовим текстом, з використанням словників та навчальних текстів для кожного типу настрою.

2.2 Аналіз та автоматизація обробки потоків даних вебресурсу з обміну літературними творами

Оскільки, програмна реалізація методу підбору подібних за настроєм коротких художніх текстів засобами природної мови відбувається у формі вебресурсу з обміну літературними творами, то важливим кроком є визначення користувацьких ролей. Це необхідно для того, щоб розділити функціонал програми таким чином, щоб для кожної користувацької ролі був свій набір доступних функцій, що призводить до виникнення різних сценаріїв використання застосунку.

У вебресурсі з обміну літературними творами передбачено три типи користувацьких ролей: адміністратор, гість та зареєстрований користувач. Відповідно, кожна з вказаних користувацьких ролей має свій доступний функціонал.

Роль адміністратора у вебресурсі, перш за все, необхідна для управління даними. Адміністратор може переглядати, змінювати, видаляти та додавати дані до бази даних вебресурсу. Часто, основною метою існування адміністраторів вебресурсів є слідкування за порядком у спілкуванні між користувачами. Наприклад, адміністратор може обмежити можливість написання коментарів для користувача, який систематично ображає інших користувачів пишучи їм образливі коментарі. В крайніх випадках, адміністратор може навіть заблокувати профіль користувача.

Проте, ще однією задачею адміністратора на вебресурсі може бути управління процесами, до яких не мають доступу звичайні користувачі. Оскільки, вебресурс для обміну літературними творами використовуватиме метод підбору подібних за настроєм коротких художніх текстів засобами природної мови, то взаємодіяти із цим методом зможе лише адміністратор. Таке

обмеження, в першу чергу, пов'язане із тим, що користувачі можуть не знати як працює метод, а не правильне користування може призвести до того, що система не зможе коректно працювати.

Отже, адміністратор має доступ як до літературних творів, якими діляться користувачі вебресурсу, так і до навчальних текстів, необхідних для роботи методу підбору подібних за настроєм коротких художніх текстів засобами природної мови. Адміністратор може переглядати навчальні тексти та вносити в них зміни, якщо це потрібно. Також, адміністратор може видаляти тексти із БД або додавати нові. Проте, основною функцією адміністратора є можливість перенавчити систему після внесення змін у навчальні тексти. Даний функціонал є корисним, якщо розглядати можливе розширення бази навчальних текстів для інформаційної системи з метою отримання кращих результатів.

Отже, для адміністратора доступні два модулі: робота з методом та робота із даними користувачів. Діаграма взаємодії адміністратора з доступними йому системами зображена на рисунку 2.5.

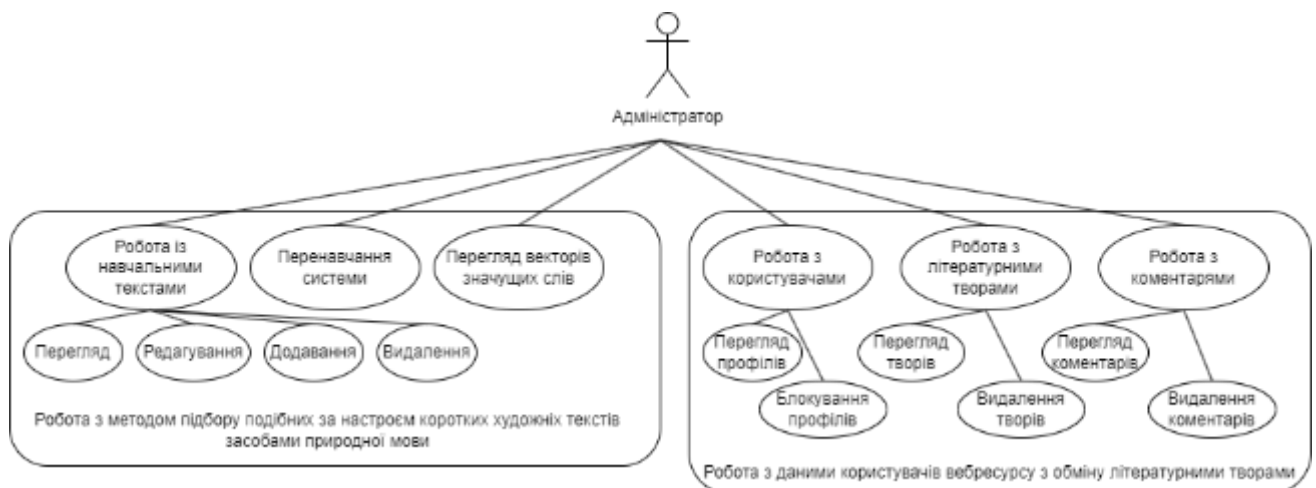


Рисунок 2.5 – Діаграма взаємодії адміністратора з вебресурсом з обміну літературними творами

Користувачів вебресурсу з обміну літературними творами можна поділити на два типи: гість та зареєстрований користувач. Гість – це користувач не зареєстрований на вебресурсі. В таких користувачів дещо обмежений функціонал. Зазвичай, гості можуть лише переглядати вміст вебсайту. В той час

як зареєстровані користувачі можуть наповнювати вебсайт контентом, наприклад, публікувати дописи або залишати коментарі чи реакції на дописи інших користувачів.

Функціонал користувачів вебресурсу з обміну літературними творами полягає в тому, що користувачі мають змогу зареєструватись на вебресурсі і публікувати власні художні твори (Рисунок 2.6). Також, користувачі можуть брати участь в обговореннях творів інших користувачів, пишучи коментарі під їхніми дописами.

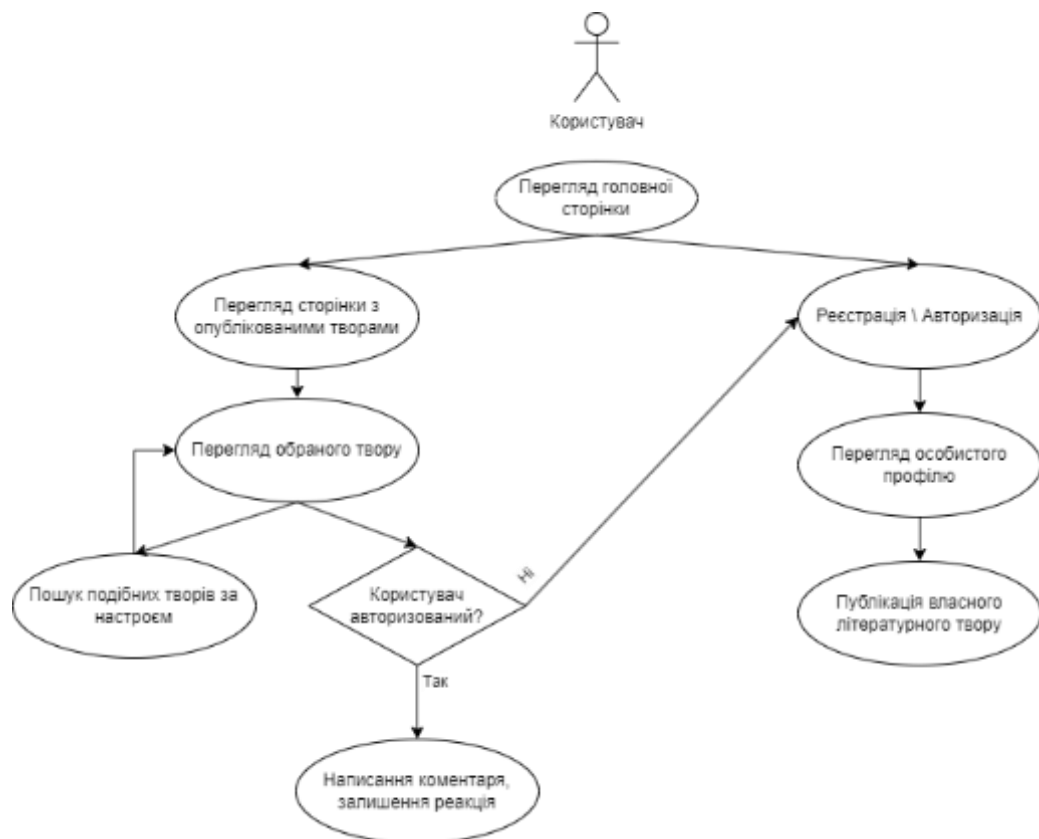


Рисунок 2.6 – Діаграма взаємодії користувача з вебресурсом з обміну літературними творами

Отже, завдяки створенню декількох користувацьких ролей можна розділити функціонал програмного продукту відповідно до ролей користувачів. Таким чином, функціонал пов'язаний з управлінням даними вебресурсу та роботою з методом підбору подібних за настроєм коротких художніх текстів засобами природної мови доступний лише для адміністратора. В той час як для користувачів доступний функціонал пов'язаний з обміном літературних творів.

2.3 Функціональна структура вебресурсу з обміну літературними творами

Функціональна структура вебресурсу з обміну літературними творами складається з декількох груп функцій, які об'єднують в собі функціонал, пов'язаний з одним бізнес-процесом. Спершу, варто розглянути групи функцій для адміністратора вебресурсу.

Перша група функцій для адміністратора пов'язана із роботою з методом підбору подібних за настроєм коротких художніх текстів засобами природної мови (Рисунок 2.7).

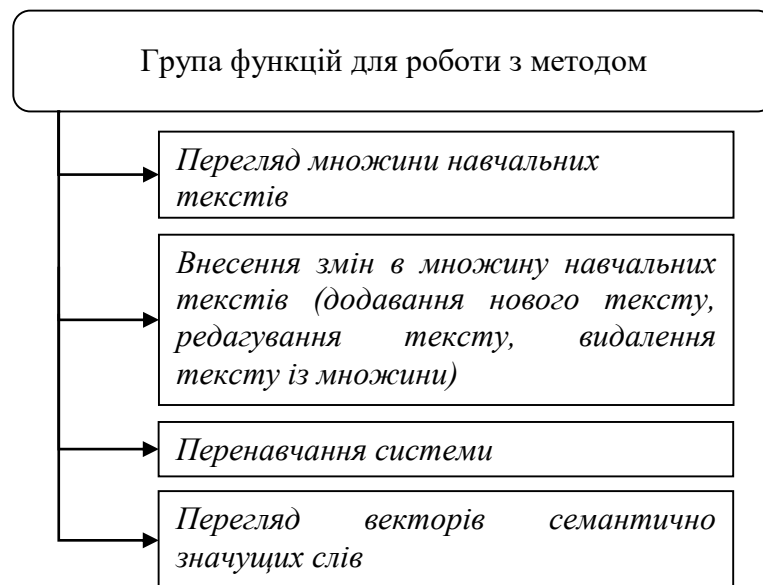


Рисунок 2.7 – Група функцій адміністратора для роботи з методом підбору подібних за настроєм коротких художніх текстів засобами природної мови

Функціонал, який доступний в межах цієї групи, включає в себе роботу з множиною навчальних текстів, які можна переглядати, редагувати, видаляти або додавати нові тексти. Оскільки, навіть найменші зміни навчального датасету впливають на стан системи, то для адміністратора доступна функція перенавчання системи. Також, для перегляду результатів навчання системи адміністратор має доступ до перегляду векторів семантично значущих слів для кожного типу настрою.

Наступна група функцій для адміністратора включає в себе роботу із даними користувачів вебресурсу з обміну літературними творами (Рисунок 2.8). Сюди відноситься перегляд інформації про користувачів, перегляд дописів, опублікованих користувачами, перегляд коментарів до дописів. Крім цього, в адміністратора є можливість блокувати профіль користувача, видаляти його дописи та видаляти коментарі користувачів.

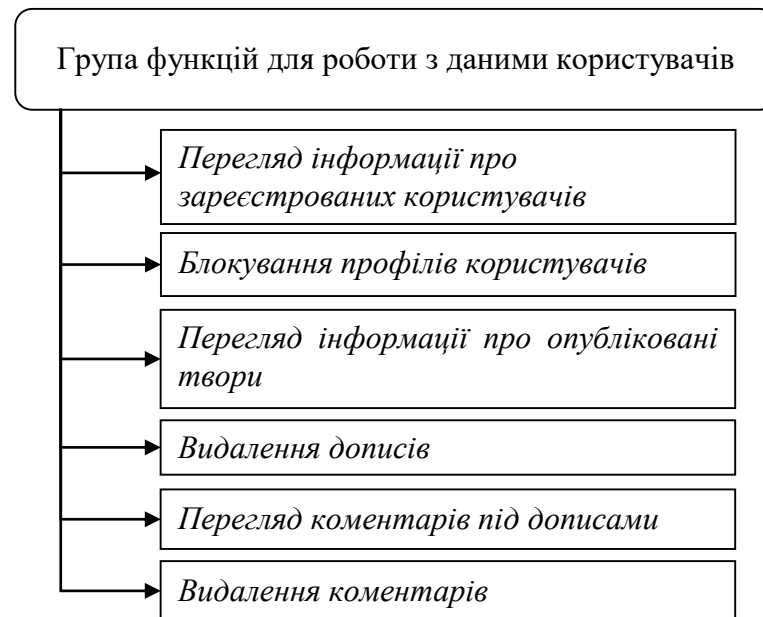


Рисунок 2.8 – Група функцій адміністратора для роботи з даними користувачів

Для користувача вебресурсу з обміну літературними творами передбачена група функцій, пов'язана з публікацією дописів (Рисунок 2.9). Користувач може переглядати дописи, опубліковані іншими користувачами на окремій сторінці вебресурсу, де доступні функції сортування дописів за датою публікації, емоційним настроєм та автором. На цій сторінці твори подані в скороченій формі, тому, щоб переглянути повний текстовий зміст, користувачу необхідно обрати допис. На сторінці обраного допису користувач може прочитати твір повністю, і за бажанням, залишити коментар чи лайк. Також, користувач може шукати подібні за емоційним настроєм твори до обраного твору.

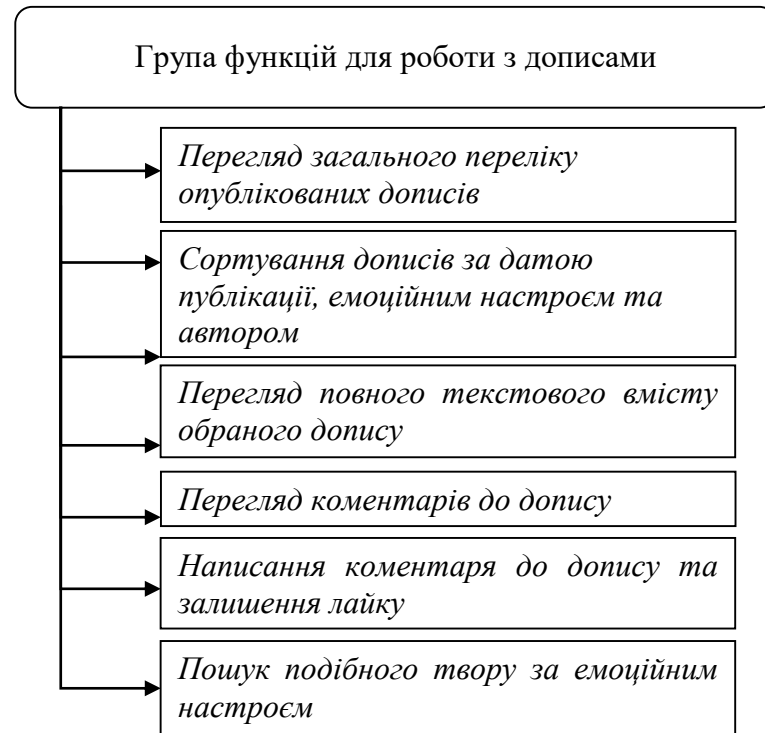


Рисунок 2.9 – Група функцій користувача для роботи з дописами

Для користувачів вебресурсу важливо реалізувати можливість реєстрації з подальшою роботою з особистим профілем (Рисунок 2.10). Зареєстровані користувачі можуть редагувати особисту інформацію, яка відображається на їхньому профілі. Також, на сторінці особистого профілю доступна можливість опублікувати власний літературний твір. Крім цього, на особистому профілі відображається перелік опублікованих творів у хронологічному порядку від найновішого. За бажанням, користувачі можуть видаляти опубліковані раніше твори. Також, користувач може видалити свій профіль, що призведе до видалення усіх його коментарів та дописів. Для того, щоб завершити роботу з вебресурсом з обміну літературними творами користувачу достатньо вийти з особистого профілю.

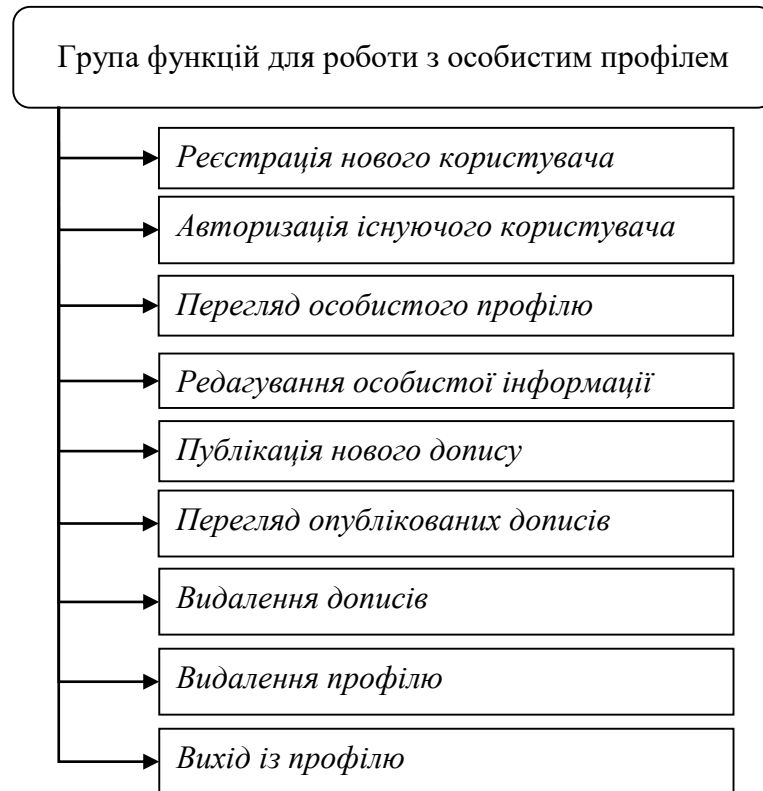


Рисунок 2.10 – Група функцій користувача для роботи з особистим профілем

Таким чином, поділ функціоналу вебресурсу з обміну літературними творами на окремі групи функцій, враховує потреби двох користувацьких ролей та надає їм доступ до необхідного функціоналу. Зокрема, адміністратор системи має доступ до управління даними користувачів та методом підбору подібних за настроєм коротких художніх текстів засобами природної мови. В той час, як для користувачів доступний функціонал пов'язаний із переглядом опублікованих творів, пошуком подібних творів за емоційним настроєм, обговоренням творів у коментарях та публікацією власних творів.

2.4 Проектна архітектура вебресурсу з обміну літературними творами та взаємозв'язок компонентів

Вебресурс з обміну літературними творами, який використовую метод подібних за настроєм коротких художніх текстів засобами природної мови складається з трьох взаємопов'язаних підсистем та бази даних (Рисунок 2.11).



Рисунок 2.11 – Схема структури вебресурсу з обміну літературними творами

Дві підсистеми є прикладними, адже в них зосереджений функціонал, пов'язаний із взаємодією користувачів з вебресурсом. Перша прикладна підсистема містить в собі функціонал для звичайних користувачів вебресурсу. Тобто, тут зосереджені функції пов'язані із реєстрацією та авторизацією користувачів, роботою з особистим профілем користувача, публікацією власних творів та переглядом творів опублікованих іншими користувачами. Дана підсистема пов'язана із БД, адже усі особисті дані користувачів зберігаються в базі даних. Також, кожен опублікований допис зберігається в БД разом із обчисленими оцінками прояву емоційного настрою, які відображаються на сторінці кожного твору. Оскільки, публікація допису передбачає автоматизоване визначення прояву емоційних настроїв у творі, то ця підсистема, також, пов'язана й з підсистемою, яка використовує реалізований метод підбору подібних творів.

Також, коли користувач проводить пошук подібних за настроєм творів, дана прикладна підсистема звертається до підсистеми методу, яка проводить усі необхідні обчислення та видає результат.

Наступна прикладна підсистема вміщує в себе функціонал пов'язаний і роллю адміністратора вебресурсу. Оскільки, адміністратор вебресурсу виконує функції пов'язані з управлінням даними то ця підсистема взаємопов'язана із БД, адже адміністратор може не лише переглядати дані а й вносити в них зміни та видаляти. Також, адміністратор може керувати роботою методу підбору подібних за настроєм коротких художніх текстів засобами природної мови через внесення змін до множини навчальних текстів. Після внесення змін, які можуть бути у вигляді редагування наявних текстів, додавання нових або видалення існуючих, адміністратор може перенавчити систему, що призведе до отримання інших результатів пошуку подібних за настроєм текстів.

Остання підсистема підбору подібних за настроєм коротких художніх текстів засобами природної мови, яка використовує метод описаний в п.п 2.1 містить в собі послідовну реалізацію кроків методу. Відповідно до методу, підсистема отримує з БД множину навчальних текстів та використовуючи словники настроїв формує вектори семантично значущих слів для кожного типу настрою і зберігає цю інформацію в БД. Далі, при кожній публікації користувачами своїх літературних творів, у цій підсистемі відбувається обчислення оцінок прояву настроїв для кожного типу емоційного настрою. Ці оцінки також зберігаються в БД та відображаються на сторінках дописів. Також, коли користувачі проводять пошук подібних за емоційним настроєм творів у цій підсистемі відбувається обчислення оцінок подібності та формування переліку подібних творів, які відображаються користувачеві через графічний інтерфейс. Ці дані не зберігаються у БД, адже потребують динамічного перерахунку щоразу, коли це потрібно.

Необхідність існування БД зумовлюється потребою в збереженні великої кількості даних та інформації. Оскільки, ці дані потребують не лише збереження

а й можливості з ними працювати, то БД повинна мати чітку структуру, яка надаватиме доступ до даних.

Отже, відповідно до спроектованої структури інформаційної системи можна побачити, як окремі компоненти взаємодіють між собою утворюючи цілісну систему. Кожна з підсистем має набір функціоналу, який необхідний для її існування та зв'язку з іншими підсистемами. В результаті описаної взаємодії підсистем вебресурс з обміну літературними творами надає можливість користувачам обмінюватись власними творами та шукати подібні за емоційним настроєм твори використовуючи метод підбору подібних за настроєм коротких художніх текстів засобами природної мови.

2.5 Проектування даталогічної моделі бази даних вебресурсу з обміну літературними творами

Для реалізації методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови у вигляді вебресурсу з обміну літературними творами необхідно спроектувати структуру даталогічної моделі бази даних з метою збереження усіх необхідних даних для правильної роботи інформаційної системи. Спроектована даталогічна модель БД вебресурсу з обміну літературними творами складається з десяти таблиць: Users, User type, Sex, Posts, Likes, Comments, Evaluation of emotionality, Emotional moods, Text for learning, Vectors (Рисунок 2.12).

Умовно, БД можна поділити на дві частини. До першої частини можна віднести таблиці, які зберігають інформацію про користувачів вебресурсу та їхні опубліковані літературні твори. Друга частина БД зберігає вибірки навчальних текстів та пов'язану із ними обробку даних з метою уникнення проведення навчання системи при кожному запуску застосунку.

Оскільки, вебресурс з обміну літературними творами включає в себе функціонал пов'язаний із навчанням системи, який не має бути доступний звичайним користувачам вебресурсу, виникає необхідність розділення

користувачів за типами. В залежності від типу, користувач буде наділений різними правами доступу до функціоналу вебресурсу. Наприклад, адміністратор матиме доступ до модулю навчання системи та зможе перенавчити систему, розширивши базу навчальних текстів. Для реалізації цього задуму в БД створено таблицю для збереження інформації про типи користувачів з назвою «User type» (Таблиця 2.1).

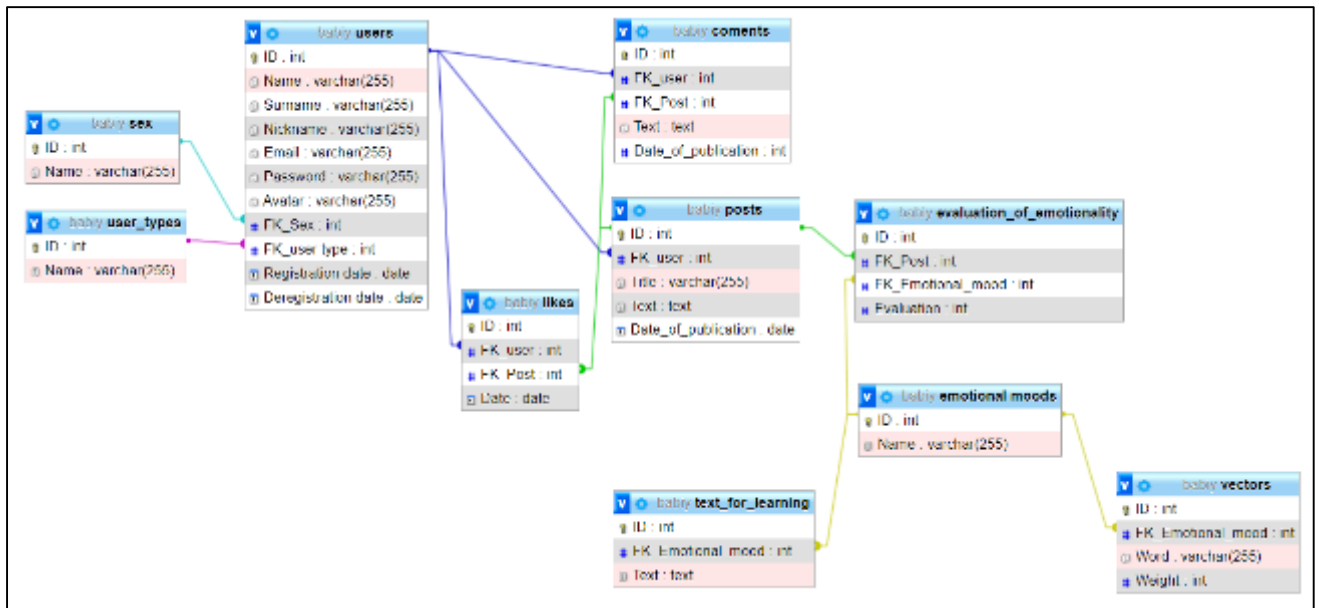


Рисунок 2.12 – Даталогічна модель БД вебресурсу з обміну літературними творами

Таблиця 2.1 – Опис атрибутів таблиці «User type»

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

Основою вебресурсу з обміну літературними творами є наявність зареєстрованих користувачів, які зможуть здійснювати обмін літературними творами, шляхом публікації творів на своїх персональних сторінках. Тому,

виникає необхідність у створенні таблиці «Users», яка зберігатиме персональну інформацію зареєстрованих користувачів (Таблиця 2.2).

Таблиця 2.2 – Опис атрибутів таблиці «Users»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ.
2.	Name	text	Ім'я користувача
3.	Surname	text	Прізвище користувача
4.	Nickname	text	Нікнейм користувача
5.	Email	text	Електронна пошта користувача
6.	Avatar	text	Фото користувача
7.	Password	text	Пароль для входу на особистий профіль
8.	FK_sex	int	Вторинний ключ для збереження інформації про стать користувача
9.	FK_user_type	int	Вторинний ключ для збереження інформації про тип користувача
10.	Registration_date	date	Дата реєстрації користувача на вебресурсі.
11.	Deregistration_date	date	Дата видалення профілю користувача з вебресурсу.

Стать користувачів є важливим параметром для ідентифікації особистості та обробки даних вебресурсу з обміну літературними творами, тому виникає необхідність створення окремої таблиці «Sex» (Таблиця 2.3).

Таблиця 2.3 – Опис атрибутів таблиці «Sex»

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

Обмін літературними творами здійснюється шляхом публікації цих творів на особистому профілі користувача. Для цього створено таблицю «Posts» з набором атрибутів, метою яких є збереження даних про літературний твір (Таблиця 2.4).

Таблиця 2.4 – Опис атрибутів таблиці «Posts»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ.
2.	FK_user	int	Вторинний ключ для збереження інформації про автора допису
4.	Title	text	Назва допису
3.	Text	text	Текстовий вміст допису
5.	Date_of_publication	date	Дата публікації допису

Для взаємодії користувачів із дописами передбачена можливість написання коментарів та залишення реакцій у вигляді лайків. Для цього створено дві окремі таблиці: «Comments» (Таблиця 2.5) та «Likes» (Таблиця 2.6).

Таблиця 2.5 – Опис атрибутів таблиці «Comments»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ.
2.	FK_user	int	Вторинний ключ для збереження інформації про автора коментаря
4.	FK_post	int	Вторинний ключ для збереження інформації про прокоментований допис
3.	Text	text	Текстовий вміст коментаря
5.	Date	date	Дата публікації коментаря

Таблиця 2.6 – Опис атрибутів таблиці «Likes»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ.
2.	FK_user	int	Вторинний ключ для збереження інформації про власника лайку
3.	FK_post	int	Вторинний ключ для збереження інформації про допис, який лайкнули
4.	Date	date	Дата залишення реакції

Після публікації допису для нього автоматично визначаються оцінки прояву емоційних настроїв, які варто одразу ж зберігати в БД, щоб не перераховувати при кожному запуску програми. Для цього створено таблицю «Evaluation of emotionality» (Таблиця 2.7).

Таблиця 2.7 – Опис атрибутів таблиці «Evaluation of emotionality»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ.
2.	FK_post	int	Вторинний ключ для збереження інформації про допис
3.	FK_emotional_mood	int	Вторинний ключ для збереження інформації клас емоційного настрою
4.	Evaluation	int	Оцінка прояву обраного емоційного настрою

Для збереження класу емоційного настрою створено таблицю «Emotional moods» з такими атрибутами: ID, Name. (Таблиця 2.8).

Таблиця 2.8 – Опис атрибутів таблиці «Emotional moods»

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

Збереження множини навчальних текстів дає змогу швидко отримувати інформацію про ці тексти, змінювати її, видаляти чи додавати нову. Тому, важливо створити окрему таблицю з назвою «Text for learning» та атрибутами: ID, FK_emotional_mood, Text (Таблиця 2.9).

Для збереження сформованих векторів значущих слів для кожного класу емоційних настроїв створено таблицю «Vectors» (Таблиця 2.10).

Таблиця 2.9 – Опис атрибутів таблиці «Text for learning»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ.
2.	FK_emotional_mood	int	Вторинний ключ для збереження інформації про емоційний настрій тексту
3.	Text	text	Текстовий вміст навчального тексту

Таблиця 2.10 – Опис атрибутів таблиці «Vectors»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ.
2.	FK_emotional_mood	int	Вторинний ключ для збереження інформації про емоційний настрій вектору значущих слів
3.	Word	text	Слово
4.	Weight	int	Вага слова

Отже, було спроектовано даталогічну модель бази даних вебресурсу з обміну літературними творами, яка складається з десяти таблиць, що зберігають усю необхідну інформацію про користувачів та опубліковані ними літературні твори. Крім цього, у БД зберігаються навчальні тексти, необхідні для роботи методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови.

2.6 Підготовка словників для подання категорій настрою художніх текстів

Оскільки, метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови ґрунтується на використанні словників настроїв, то було знайдено відповідний словник тональності слів для української мови «Tone dict uk» [21]. Словник складається із переліку слів з визначеною дискретною тональністю в діапазоні від -2 до 2, де слова із тональністю -2, -1 – це слова із песимістичним емоційним настроєм, 0 – нейтральні слова, 1, 2 – слова із оптимістичним настроєм (Рисунок 2.13). Особливістю та перевагою даного словника є те, що всі слова приведені до базової форми.

361	безталанний	-1	-1	0
362	безтолковий	-2	-1,666666666666667	0,471404520791032
363	безтурботний	0	0	0,816496580927726
364	безтурботний	0	-0,4444444444444444	0,955813918560292
365	безтурботний	1	0,666666666666667	0,816496580927726
366	безумовний	0	0	0
367	безумство	-1	-1	0
368	безуспішний	-1	-1	0
369	безхитрісний	0	0	0,816496580927726
370	безхребетний	-1	-1	0
371	безхребетний	0	0	0
372	безцеремонний	-1	-1,3333333333333333	0,471404520791032
373	безцільний	-1	-1,3333333333333333	0,471404520791032
374	безцінний	2	1,666666666666667	0,471404520791032

Рисунок 2.13 – Фрагмент вигляду словника тональності слів [21]

Для зручного використання словника тональності слів, потрібно провести його попередню обробку. Словник містить два зайві поля, які не використовуватимуться в роботі методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови, тому ці поля будуть видалені. Також, оцінка емоційного настрою слів має два ступені градації, а оскільки, для роботи методу це надлишкова інформація, то усі слова із значенням -2 та 2 будуть змінені на значенням -1 та 1, відповідно. Таким чином, буде отримано новий словник тональності в якому є лише перелік слів та їхня оцінка емоційного настрою в діапазоні від -1 до 1.

Загалом в словнику міститься 6859 слів, з яких песимістичними є 2085, нейтральними – 2672, оптимістичними – 2102 (Рисунок 2.14).

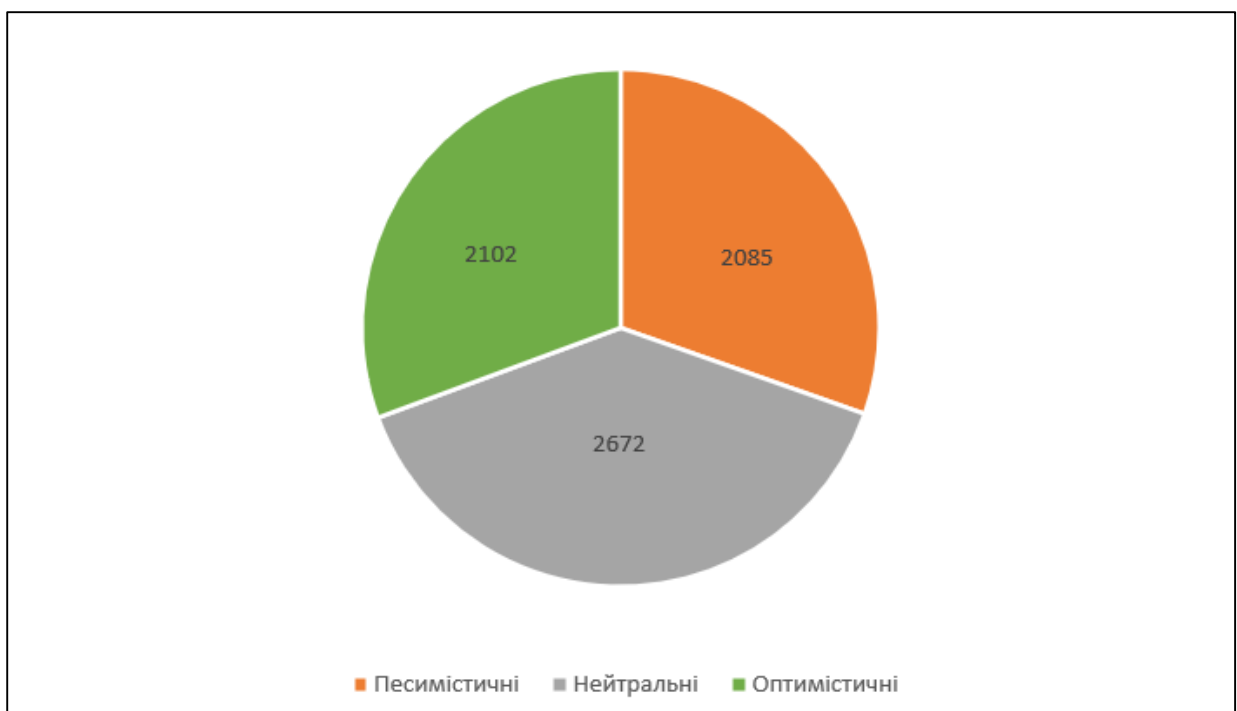


Рисунок 2.14 – Розподіл слів за емоційним настроєм

Сформований словник тональності слів буде використано на першому етапі методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови, який полягає у формуванні векторів семантично значущих слів для песимістичного, нейтрального та оптимістичного типів настроїв для коротких художніх текстів.

2.7 Особливості використання спеціалізованих програмних компонентів

Для того щоб пришвидшити та спростити процес реалізації методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови у вигляді вебресурсу з обміну літературними творами буде використано бібліотеку PHP ML.

Бібліотека PHP ML – це спеціалізований інструмент для використання методів та алгоритмів машинного навчання у середовищі PHP [22]. Бібліотека дозволяє використовувати в проєктах такі методи машинного навчання, як класифікація (SVC, KNearestNeighbors, NaiveBayes), кластеризація (KMeans, DBSCAN), регресія (LeastSquares, SVR). Використання PHP ML для розробки вебресурсу з обміну літературними творами зумовлене тим, що метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови потребує використання статистичного методу TF-IDF для підрахунку значущості слів у текстах. Бібліотека PHP ML включає в себе реалізацію методу TF-IDF, який за вхідними даними у вигляді переліку унікальних слів з кількістю їхніх повторів, визначає вагу слів [23].

Отже, для реалізації методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови у вигляді вебресурсу з обміну літературними творами буде використано бібліотеку PHP ML для обчислення значущості слів у текстах методом TF-IDF.

2.8 Визначення шляхів дослідження та засобів створення програмного забезпечення

Для дослідження роботи ефективності розробленого методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови прийнято рішення створити вебресурс з обміну літературними творами.

Перш за все вебресурс повинен забезпечувати виконання таких функцій, як реєстрація та авторизація користувачів у системі. У вебресурсі повинно бути реалізовано поділ функціоналу відповідно до користувацьких ролей. Для адміністратора необхідно реалізувати можливість керування даними, які зберігаються в БД та управління роботою методу підбору подібних текстів. Для користувачів необхідно забезпечити можливість перегляду та публікації дописів, написання коментарів та пошуку подібних за настроєм текстів.

Для реалізації вебресурсу з обміну літературними творами потрібно провести вибір засобів розробки, зокрема, вибір мови програмування, фреймворку та СКБД, які підходять для розробки веб застосунків.

Враховуючи функціонал, який повинен бути реалізований у вебресурсі, необхідно розділити програмний код на взаємопов'язані компоненти. Для веб застосунків, поділ функціоналу відбувається за шаблоном MVC, який дозволяє розділити логіку програми та реалізацію методів від графічного інтерфейсу та бази даних.

Для реалізованого вебресурсу з обміну літературними творами необхідно провести тестування створюючи тест-кейси. Даний метод тестування, перш за все, націлений на перевірку правильності роботи програмного застосунку. Створюючи тест-кейси можна відслідкувати, які дії користувачів призводять до виникнення помилок. Це дозволяє краще продумати шляхи виправлення помилок та покращити досвід використання програми для користувачів.

Особливу увагу варто приділити дослідженню ефективності роботи методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної. Для цього, можна порівнювати результати роботи методу із оцінками отриманими від експерта. В даному випадку, прийнято рішення обрати за експерта ChatGPT, оскільки це добре навчена нейронна мережа, яка показує досить точні результати. Порівняння відбуватиметься за показником переважаючого емоційного настрою тексту. Таким чином можна буде встановити наскільки точно створена система здатна визначати емоційний настрій текстів. Таке дослідження варто провести для кожного класу емоційного

настрою, з метою встановлення здатності системи розрізняти емоційні настрої у текстах.

Таким чином, тестування вебсервісу з обміну літературними творами дасть змогу дослідити чи коректно працює система та чи доступні користувачам функції, яких вони потребують. А дослідження ефективності роботи методу покаже наскільки точні результати надає вебресурс.

2.9 Висновки до розділу 2

Метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови призначений для визначення емоційного настрою тексту та пошуку подібних за визначеним настроєм текстів. Розроблений метод поділяється на три послідовні етапи. На першому етапі відбувається формування векторів значущих слів для кожного типу настрою методом TF-IDF з використанням множини навчальних текстів та словників тональності для кожного з типів настроїв. Другий етап полягає в визначенні оцінок прояву кожного типу настрою в користувацьких текстах. Призначенням третього етапу є формування переліку подібних текстів за емоційним настроєм на основі тестового тексту.

Вебресурс з обміну літературними творами є прикладною програмною реалізацією методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної. У вебресурсі передбачено поділ функціоналу відповідно до користувацьких ролей. Адміністратор вебресурсу має доступ до управління даними, які зберігаються в БД, а також, може вносити зміни до множини навчальних текстів та перенавчати систему. Інші користувачі вебресурсу мають змогу зареєструватись та наповнювати вебресурс контентом, публікуючи власні літературні твори та обговорюючи твори інших користувачів у коментарях. Також, для користувачів буде доступний функціонал пошуку подібних текстів за емоційним настроєм.

Оскільки, для того щоб вебресурс з обміну літературними творами міг забезпечувати правильне виконання функцій, необхідно спочатку забезпечити можливість збереження даних. Для цього, було спроектовано структуру бази даних, яка складається з десяти таблиць. У БД зберігаються дані не лише про користувачів вебресурсу та їхні дописи, а й множина навчальних текстів та стан навченої системи з метою оптимізації використання програми. Такий підхід дозволяє зберегти вектори значущих слів та оцінки прояву настроїв один не раз і не проводити величезні обрахунки після кожного запуску програмного продукту.

Оскільки, метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови ґрунтується на використанні словників тональності, було знайдено відповідний словник слів поділений на три класи емоцій: оптимістичні, нейтральні та песимістичні. Для використання знайденого словника було проведено його попередню обробку з метою видалення надлишкових даних. А для автоматизації використання методу TF-IDF, яким формуються вектори значущих слів для кожного типу настрою, використовуватиметься спеціалізована бібліотека PHP ML, яка включає реалізацію даного методу.

Отже, для прикладної реалізації методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови потрібно створити вебресурс з обміну літературними творами, який забезпечуватиме виконання зазначених функцій. Використовуючи розроблений застосунок необхідно провести дослідження ефективності роботи методу. Також, необхідно провести тестування програмного продукту.

Розділ 3 Програмна реалізація вебресурсу з підбору подібних за настроєм художніх текстів

3.1 Вибір засобів розробки вебресурсу з підбору подібних за настроєм художніх текстів

Для програмної реалізації методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови у вигляді вебресурсу з обміну літературними творами потрібно використати набір таких засобів розробки, як: мова програмування, фреймворк та СКБД.

Оскільки, програмна реалізація методу полягає в створенні вебресурсу з обміну літературними творами, було прийнято рішення використати мову програмування PHP, адже це одна з найпопулярніших мов для веб розробки. PHP має велику активну спільноту, що сприяє швидкому пошуку програмних рішень та отриманні допомоги від професіоналів. Також, для PHP існує декілька хороших фреймворків, які значно прискорюють та полегшують розробку веб застосунків [24].

Laravel – це один з найпопулярніших фреймворків для розробки веб застосунків на мові програмування PHP. Однією з основних переваг Laravel є його простота використання, адже бібліотека має хорошу та зрозумілу структуру. Фреймворк працює на основі шаблону проектування MVC, використання якого допомагає структурувати програмний код застосунку, розділивши його на окремі взаємопов'язані компоненти. Також, фреймворк має безліч вбудованих інструментів, які автоматизовують розробку таких модулів як створення системи автентифікації користувачів. Оскільки, вебресурс з обміну літературними творами передбачає процес реєстрації користувачів, то використання інструментів які надає фреймворк значно пришвидшать процес реалізації цього функціоналу. Ще однією перевагою фреймворку Laravel є те, що він постійно розвивається та вдосконалюється, випускаючи нові версії. Все це супроводжується детальною документацією від розробників фреймворку, яка містить приклади використання компонентів бібліотеки [25].

Зважаючи на те, що розробка вигляді вебресурсу з обміну літературними творами передбачає роботу з великою кількістю даних, виникає необхідність правильного збереження цих даних з метою отримання до них доступу в потрібний момент. Для цього використовуються спеціалізовані інструменти для збереження та керування даними, які називаються СКБД. Одним із популярних та безкоштовних варіантів СКБД є MySQL, який легко інтегрується до програмних застосунків, створених мовою програмування PHP на фреймворку Laravel. MySQL є досить гнучкою системою та простою в налаштуванні під власні потреби. MySQL надає високу продуктивність, що дозволяє швидко отримувати доступ до великої кількості даних, що особливо актуально при розробці вебресурсу з обміну літературними творами, який може включати велику кількість користувачів [26].

Отже, для програмної реалізації методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови у вигляді вебресурсу з обміну літературними творами прийнято рішення використати такий набір інструментів, як: мова програмування PHP, фреймворк Laravel та СКБД MySQL. Вибір мови зумовлений тим, що необхідно створити веб застосунок, а PHP відмінно для цього підходить, адже обрана мова програмування повністю орієнтована на веб розробку. Фреймворк Laravel є хорошим рішенням для реалізації вебресурсу з обміну літературними творами, адже надає готові рішення для автоматизованого створення системи автентифікації користувачів вебресурсу. СКБД MySQL є гнучким та зручним інструментом збереження даних, який легко під'єднується до веб застосунків, зокрема, через фреймворк Laravel.

3.2 Структура та функціональне призначення програмних складових вебресурсу з підбору подібних за настроєм художніх текстів

При створенні вебресурсу з обміну літературними творами використовується фреймворк Laravel, який базується на шаблоні MVC.

Застосування шаблону MVC дозволяє розділити інтерфейс користувача і програмну логіку додатку. Шаблон MVC складається з трьох основних компонентів, кожен з яких виконує певні функції в інформаційній системі. Models або моделі відповідають за взаємодію з базою даних. Views або представлення зберігають структуру сторінок вебсайту і відповідають за інтерфейс користувача. Controllers або контролери діють як посередники між моделями та представленнями: вони отримують дані з моделей, обробляють їх і передають у представлення для відображення користувачу [27].

Згідно із шаблоном проектування MVC, та спроектованої в п.п 2.5 даталогічної моделі БД вебресурсу, в проекті створено десять моделей:

- User – модель для збереження даних про зареєстрованих користувачів на вебресурсі з обміну літературними творами;
- UserType – модель для збереження даних про типи користувачів (адміністратор або звичайний користувач);
- Sex – модель для збереження інформації про стать зареєстрованих користувачів;
- Post – модель для збереження інформації про опубліковані твори користувачів;
- Like – модель для збереження інформації про лайки, які користувачі можуть залишати до дописів;
- Comment – модель для збереження інформації про коментарі, які користувачі можуть писати під дописами;
- EvaluationOfEmotionality – модель для збереження оцінок прояву емоцій для опублікованих творів;
- EmotionalMood – модель для збереження інформації про типи емоційного настрою, які використовуються в програмі;
- TextForLearning – модель для збереження навчальних текстів;
- Vector – модель для збереження векторів значущих слів для кожного типу настрою.

Для реалізації інтерфейсу користувача, через який відбувається взаємодія користувача із вебресурсом, в проєкті було створено такі файли представлень:

- AuthLogin – сторінка для авторизації;
 - AuthRegister – сторінка для реєстрації;
 - LayoutsMain – загальна розмітка сторінок, яка включає шапку вебсайту та підвал;
 - PostsIndex – сторінка з переліком усіх дописів;
 - PostsShow – сторінка обраного допису;
 - Home – особиста сторінка авторизованого користувача;
 - Index – головна сторінка вебресурсу з обміну літературними творами.
 - AdminTextIndex – сторінка для перегляду множини навчальних текстів в панелі адміністратора;
 - AdminTextShow – сторінка для перегляду обраного тексту в панелі адміністратора;
 - AdminTextCreate – сторінка з формою для додавання нового навчального тексту в панелі адміністратора;
 - AdminVectorIndex – сторінка для перегляду векторів значущих слів для типів настроїв;
 - AdminUserIndex – сторінка для перегляду переліку зареєстрованих користувачів в панелі адміністратора;
 - AdminUserShow – сторінка для перегляду обраного користувача в панелі адміністратора;
 - AdminPostIndex – сторінка для перегляду переліку опублікованих дописів в панелі адміністратора;
 - AdminPostShow – сторінка для перегляду обраного допису в панелі адміністратора;
- Також, було створено такі контролери:
- HomeController – контролер для роботи із головною сторінкою вебресурсу з обміну літературними творами (перевіряє чи користувач авторизований та відображає його персональну сторінку);

- AdminController – контролер в якому реалізований функціонал для адміністратора, зокрема, робота з методом та даними про користувачів вебресурсу;
- PostsController – контролер, який відповідає за відображення, взаємодію та створення дописів зареєстрованими користувачами вебресурсу;
- ComentsController – контролер, який відповідає за написання та відображення коментарів до дописів;
- LikesController – контролер, який відповідає за можливість залишати лайки на дописах;
- RegisterController – контролер, який відповідає за реєстрацію користувачів на вебресурсі з обміну літературними творами;
- LoginController – контролер, який відповідає за авторизацію користувачів на вебресурсі з обміну літературними творами.

На діаграмі 3.1 можна побачити взаємозв'язок деяких створених компонентів програмного застосунку (повна діаграма наведена в додатку В). З діаграми видно, що контролери є посередниками для обміну даними між моделями та представленнями. В моделях реалізовано доступ до даних, які зберігаються в БД. Також, тут створенні зв'язки між моделями для отримання додаткової інформації про сутності. В контролерах реалізовано усі основні методи, необхідні для роботи вебресурсу. В представленнях міститься відображення результатів взаємодії між контролером та моделлю у вигляді графічного інтерфейсу користувача.

Якщо в програмному продукті необхідно реалізувати додаткові методи та алгоритми, які можуть використовуватись в різних частинах програми, необхідно використати інструменти під назвою Helpers або помічники. Це окремі файли, які дозволяють відділити певні алгоритми та викликати їх у будь-якому місці програми. Оскільки, для роботи вебресурсу з обміну літературними творами необхідно реалізувати метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови, то реалізацію методу краще

за все відокремити в окремому файлі помічника. Тому, було створено файл з назвою «Method» для реалізації методу.

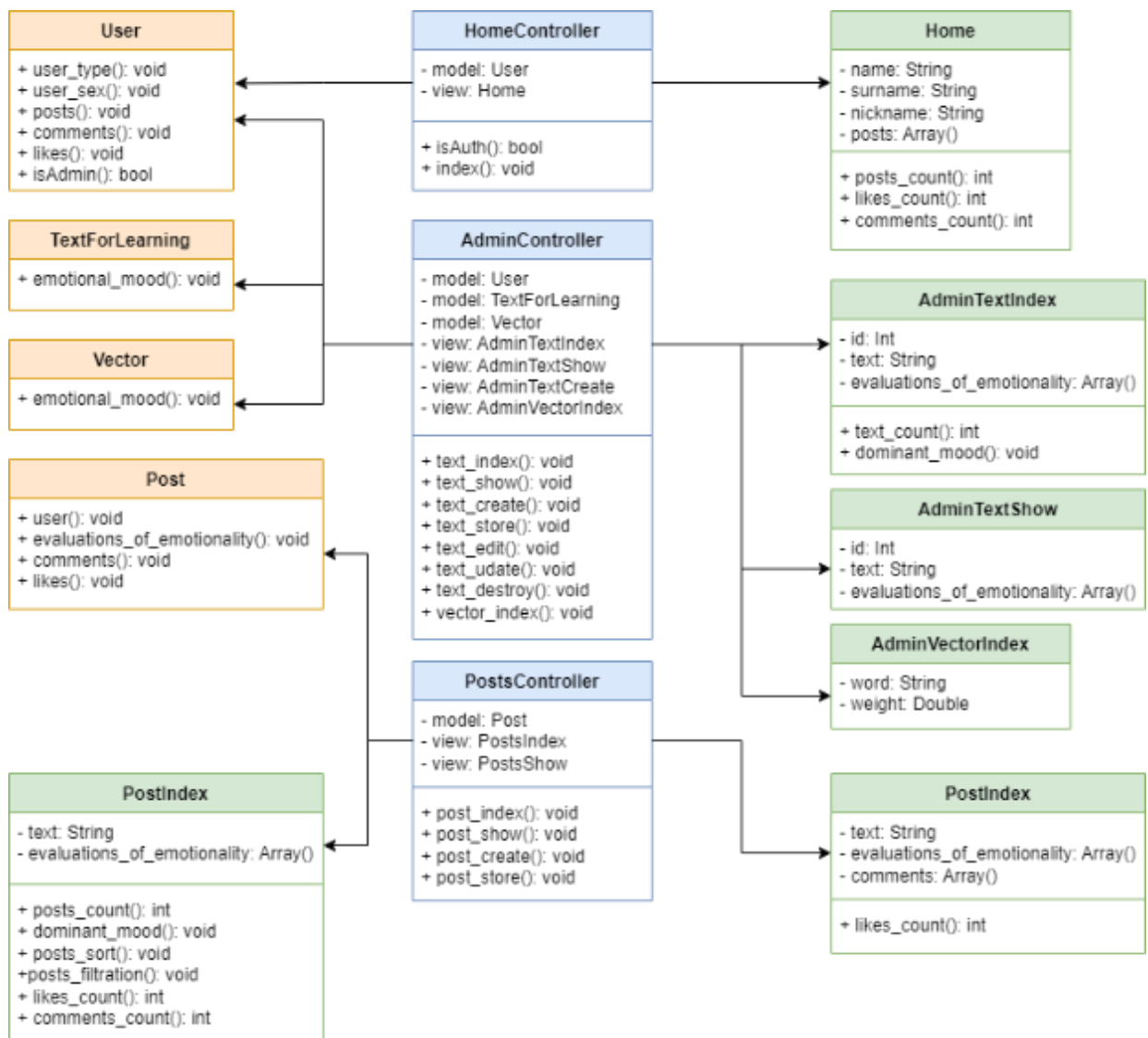


Рисунок 3.1 – Діаграма класів створеного вебресурсу з обміну літературними творами

Реалізована структура модулів вебресурсу з обміну літературними творами відповідно до шаблону MVC складається із десяти моделей, п'ятнадцяти представлень, семи контролерів та одного файлу помічника в якому реалізовано метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови. Такий підхід дає змогу розділити програмний код на різні взаємопов'язані компоненти, що значно пришвидшує і полегшує

розробку програмного застосунку, а також, його подальшу підтримку та оновлення.

3.3 Особливості реалізації програмних складових вебресурсу з підбору подібних за настроєм художніх текстів

Графічний інтерфейс користувача, який надає доступ до функціоналу програмного продукту, можна розділити на дві частини: інтерфейс для адміністратора вебресурсу та інтерфейс для користувачів. Інтерфейс для адміністратора вебресурсу, перш за все, надає доступ до функціоналу пов'язаного з управлінням методом підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови (Рисунок 3.2).

Панель адміністратора вебресурсу

Вихід

Модуль для навчання системи

База даних вебресурсу

Додати текст

Переглянути тексти

Перезавантажити систему

Варто показуємо 1 - 5 з 1275

№	Текстовий зміст	Емоційний настрій
1	Сьогоднішній ранок був наповнений теплим посміслом моїх батьків, які ніжно прошептали кращі фантазії і створили мені незалежне лапідарне світло. Пташки за вікном співали свої радостні пісні, віддаючи до цього дня особливе тепло-П. Природа прокинулася після довгої зими, разом з нами керує...	Оптимістичний
2	Цей вечір здався неспівзвучним догмам і порожнечі. За вікном ніколи дождь, а якщо випадки вранці, що бігли по снігу, то ми піддавалися відсутній світлості. Діжурні були теж, і я відчув, як тіні заповнюють простір навколо мене. Сьогодні я розумів, що самотність може бути надзвичайно важкою...	Песимістичний
3	Після довгого роздуму нарешті настав той день, коли ми з'їхалися всією родиною. Дітвора і дорослі обійми, шалені танцювальні реверси розігнали всі наші думки. Для всіх було моментом відпочинку нові звички і нові звички, а нові плани та мрії об'єднали нас ще більше. Це був день, сповнений...	Оптимістичний
4	Сьогодні в кімнаті пролунали перли. Покази були свіжими, а дерева стояли в тиші, створюючи прекрасну атмосферу для роздумів. Я бачив знайомі стовпи, але перли були як листки, що повільно падали на землю. Прогуків дитинства мені розповіли і викликали від...	Нейтральний
5	Сьогодні в уявній, що одна з моїх найбільших мрій розвивається на дрібні шматочки. На мій зусилля, навіть і плем'ям втрачати сенс. Відсуття розчарування та безпорадності накрили мене хмелем, яка зусилля зачухавалися у всьому, що я робив. Це був день, коли все...	Песимістичний

Наступна сторінка

Рисунок 3.2 – Перелік навчальних текстів в панелі адміністратора

На цій сторінці адміністратор може переглядати навчальні тексти на яких відбувається навчання системи. Кожен текст має заздалегідь визначений емоційний настрій. Для економії місця на сторінці, тексти відображаються в скороченому вигляді, проте, натиснувши на текст адміністратор може переглянути його повний текстовий вміст та відредагувати за потреби. Крім цього, адміністратор може видалити текст із бази даних або додати новий текст. Це дозволяє змінювати множину навчальних текстів з метою отримання інших результатів роботи методу.

Після внесення змін, адміністратор може перенавчити систему натиснувши на відповідну кнопку. Це запускає в роботу перший крок методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови, який полягає в формуванні векторів значущих слів використовуючи метод TF-IDF, словники тональності та навчальні тексти.

Для початку, використовуючи бібліотеку RNP-ML обчислюються ваги слів для кожного тексту із множини навчальних текстів. Потім, переліки цих слів порівнюють із словниками тональності, і якщо певне слово є в обох переліках, то це слово потрапляє в вектор значущих слів для кожного з настроїв. Таким чином, формується три вектори значущих слів (оптимістичний, нейтральний, песимістичний). Для того, щоб зберегти стан навченої системи, переліки слів із їхньою вагою зберігаються в базі даних. Також, це дозволяє проводити описані розрахунки лише при зміні множини навчальних текстів. Реалізація описаного функціоналу знаходить в файлі помічнику з назвою «Method».

Адміністратор має змогу переглянути, які слова входять у кожен з векторів натиснувши на кнопку «Переглянути вектори» (Рисунок 3.3). Для отримання цього переліку програма звертається до БД та отримує збереженні в ній дані про слова та їхню вагу. Такі переліки формуються для кожного типу настрою. Після кожного перенавчання системи, якщо вносились зміни в множину навчальних текстів, вектори значущих слів оновлюватимуться. Адміністратор не може змінювати ці переліки бо йому доступна лише функція перегляду.



Рисунок 3.3 – Перелік значущих слів для оптимістичного типу настрою

Для користувачів доступ до функціоналу методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови здійснюється в кількох місцях вебресурсу. Перш за все, на сторінці з переліком опублікованих творів, користувачі можуть побачити оцінку емоційного настрою, який переважає в тексті (Рисунок 3.4). Даний алгоритм реалізовано за рахунок того, що при публікації твору система автоматично визначає оцінки прояву кожного типу настрою. Ці дані зберігаються в БД і оновлюються лише якщо адміністратор перенавчає систему.

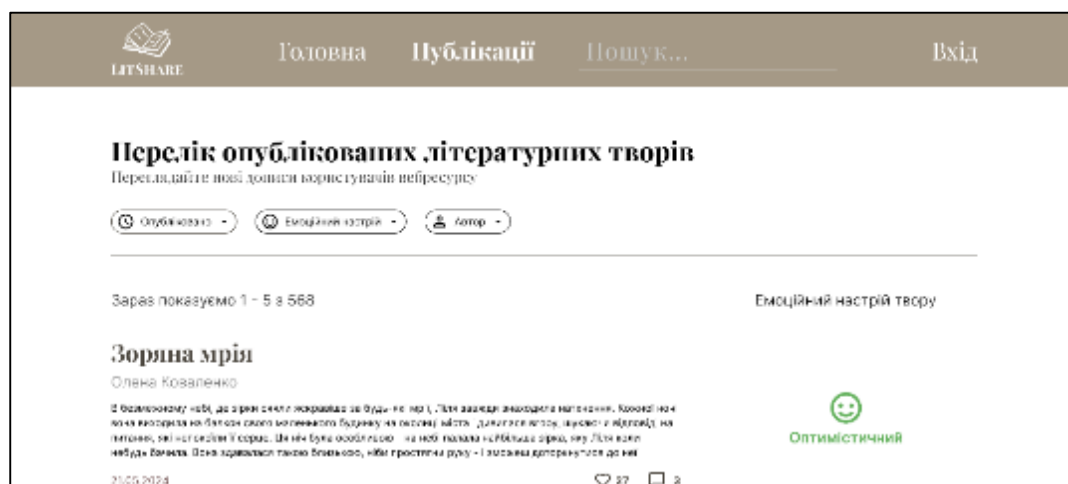


Рисунок 3.4 – Сторінка з переліком опублікованих літературних творів

Для того, щоб прочитати текстовий вміст допису повністю та переглянути оцінки прояву інших типів настрою, користувачу потрібно натиснути на обраний допис. Після цього відкриється нова сторінка на якій можна переглянути зазначену інформацію (Рисунок 3.5).

The screenshot shows a web interface for a text post. At the top, there is a navigation bar with 'Головна', 'Публікації', 'Пошук...', and 'Вхід'. The post title is 'Зоряна мрія' by 'Слена Коваленко'. The main content is a text excerpt. To the right of the text, there are three mood analysis statistics: 'Оптимістичність 61%' (Optimism 61%), 'Нейтральність 35%' (Neutrality 35%), and 'Песимістичність 4%' (Pessimism 4%). At the bottom right, there is a button labeled 'Пошук подібних' (Search similar). The date '25.05.2024' and engagement icons (likes and comments) are visible at the bottom left.

Рисунок 3.5 – Додаткова інформація про прояви настрою в тексті

Для зручності сприйняття оцінки прояву настроїв показуються у відсотковому форматі, хоча в БД вони зберігаються в числовому. Даний алгоритм ґрунтується на тому, що загальна оцінка прояву настрою – це 100%. Відповідно до цього, обчислюється відсоткове співвідношення оцінок для кожного з трьох типів настрою, які в сумі складають 100%.

Далі користувач може провести пошук подібних творів за емоційним настроєм серед числа опублікованих творів на вебресурсі. Щоразу, як користувач звертається до цього функціоналу, відбувається виклик методу пошуку подібних творів за емоційним настроєм. Цей метод порівнює оцінки прояву настроїв для обраного тексту та усіх інших користувацьких текстів та формує новий показник – оцінку подібності текстів. Проте, щоб алгоритм не рахував оцінку подібності для усіх текстів, до уваги беруться лише ті тексти у

яких переважаючий емоційний настрій збігається із переважаючим емоційним настроєм обраного тексту. Відповідно до отриманих оцінок, формується відсортований перелік творів, який починається твором із найвищою оцінкою подібності (Рисунок 3.6).



Рисунок 3.6 – Перелік подібних творів

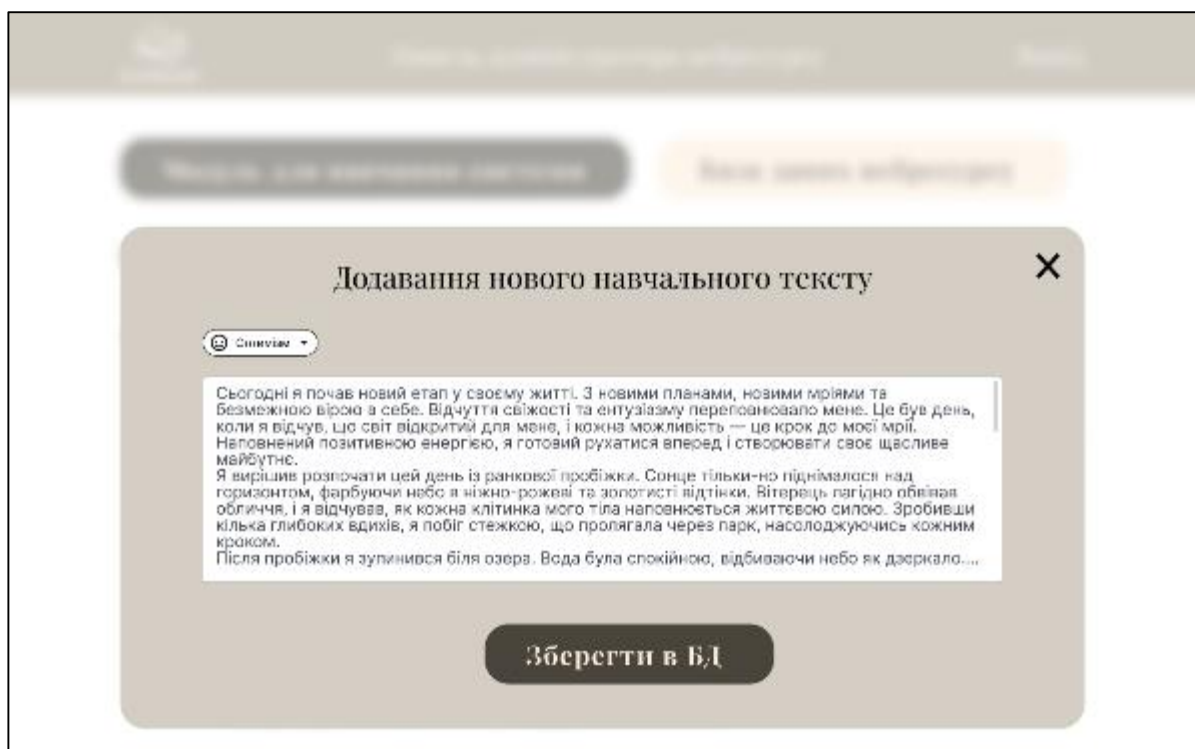
Таким чином, створений вебресурс з обміну літературними творами надає користувачам доступ до використання методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови через можливість перегляду оцінок проявів емоційного настрою для опублікованих творів та пошуку подібних творів до обраного. Також, реалізовано можливість розширення бази даних навчальних текстів адміністратором вебресурсу з метою покращення отриманих результатів.

3.4 Тестування вебресурсу та вимоги до розгортання

Для того, щоб дослідити наскільки правильно працює створений програмний продукт необхідно провести тестування його функціоналу. Для цього буде створено 4 тест-кейси для перевірки функціоналу адміністратора та

zareєстрованих користувачів вебресурсу. Для перевірки правильної роботи функцій доступних адміністратору буде створено тест-кейс, який перевірить можливість додавання нового навчального тексту та подальше перенавчання системи. Для zareєстрованих користувачів, перш за все, буде перевірено можливість реєстрації користувачів на вебресурсі, публікації власного літературного твору та написанні коментарів під дописом іншого користувача.

Щоб перевірити функціональні можливості адміністратора, спочатку потрібно авторизуватись під вказаною користувацькою роллю. Після цього відкриється сторінка з назвою «Панель адміністратора» на якій буде відображатись перелік текстів із навчальної множини. Для того щоб додати новий текст потрібно натиснути на кнопку «Додати текст», що призведе до відкриття спливаючого вікна з формою для додавання тексту (Рисунок 3.7). Щоб додати текст потрібно заповнити форму та натиснути на кнопку «Зберегти в БД».



Додавання нового навчального тексту

Сьогодні я почав новий етап у своєму житті. З новими планами, новими мріями та безмежною вірою в себе. Відчуття свіжості та ентузіазму переполювало мене. Це був день, коли я відчув, що світ відкритий для мене, і кожна можливість — це крок до моєї мрії. Наповнений позитивною енергією, я готовий рухатися вперед і створювати своє щасливе майбутнє.

Я вирішив розпочати цей день із ранкової пробіжки. Сонце тільки-но піднімалося над горизонтом, фарбуючи небо в ніжно-рожеві та золотисті відтінки. Вітерець лгідно обвівав обличчя, і я відчував, як кожна клітинка мого тіла наповнюється життєвою силою. Зробивши кілька глибоких вдихів, я побіг стежкою, що пролягала через парк, насолоджуючись кожним кроком.

Після пробіжки я зупинився біля озера. Вода була спокійною, відбиваючи небо як дзеркало....

Зберегти в БД

Рисунок 3.7 – Процес заповнення форми для додавання нового навчального тексту

Для перевірки процесу збереження нового навчального тексту в БД створено перший тест-кейс (Таблиця 3.1).

Таблиця 3.1 – Тест-кейс АТ0001

Тест-кейс ID: АТ0001	Пріоритет: 1	Створено: 24.05.2024,А. Бабій
Назва: Перевірка процесу додавання нового навчального тексту		
Вхідні дані: Текстовий вміст навчального тексту, визначений емоційний клас тексту		
Кроки	Очікуваний результат	
<p><i>Передумова:</i> користувач повинен бути авторизованим у системі під роллю адміністратора</p> <p>Відкрити вебресурс</p> <p>Авторизуватись в ролі адміністратора</p> <p>Натиснути на кнопку «Модуль для навчання системи»</p> <p>Натиснути на кнопку «Додати текст»</p> <p>Обрати емоційний настрій тексту</p> <p>Ввести текстовий вміст навчального тексту у форму</p> <p>Натиснути на кнопку «Зберегти»</p> <p>Порівняти фактичний результат з очікуваним</p>	<p>Форма для додавання нового навчального тексту закрилась. Сторінка з переліком навчальних текстів перезавантажилась.</p> <p>Першим текстом в переліку відображається щойно доданий текст.</p> <p>В БД в таблиці «Text_for_learning» з'явився новий запис з інформацією про щойно доданий навчальний текст.</p>	
Результат виконання тест-кейсу: пройдено успішно		

В результаті відтворення зазначених кроків у БД в таблиці «Text_for_learning» з'явився новий запис, який зберігає текстовий вміст нового

навчального тексту та його емоційний настрій (Рисунок 3.8). Отже, функціонал адміністратора пов'язаний із розширенням множини навчальних текстів працює вірно.

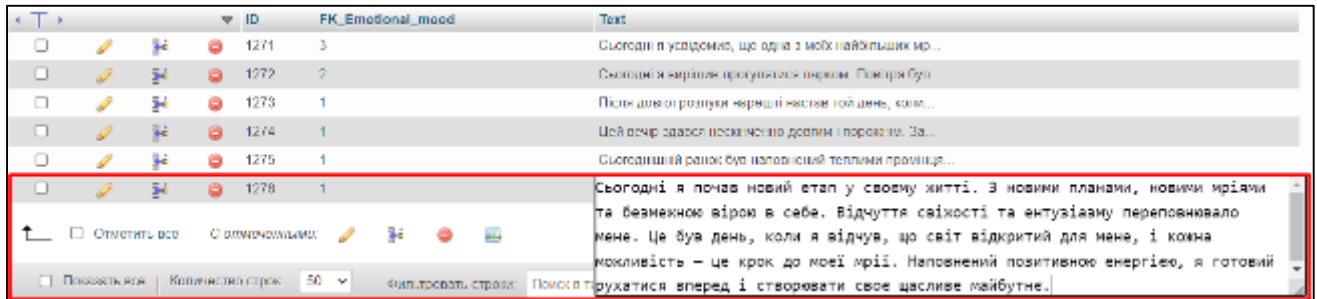


Рисунок 3.8 – Підтвердження збереження нового навчального тексту до БД

Для того щоб користувачі мали доступ до повного функціоналу вебресурсу з обміну літературними творами, необхідно перевірити можливість реєстрації нових користувачів. Для цього, користувачу потрібно заповнити форму реєстрації особистими даними. Для перевірки процесу реєстрації нових користувачів створено другий тест-кейс (Таблиця 3.2)

Таблиця 3.2 – Тест-кейс АТ0002

Тест-кейс ID: АТ0002	Пріоритет: 1	Створено: 24.05.2024, А. Бабій
Назва: Перевірка процесу реєстрації нового користувача		
Вхідні дані: Прізвище = «Сидоренко», Ім'я = «Андрій», Нікнейм = «andrew», Email = «andrews@gmail.com», Номер телефону = «+380989590921», Дата народження = «15.04.2000», Стать = «чоловік», Пароль = «andrews», Аватар = «avatar.png»		
Кроки	Очікуваний результат	
<p><i>Передумова:</i> користувач повинен бути не зареєстрованим у системі</p> <p>Відкрити вебресурс</p> <p>Натиснути кнопку «Реєстрація»</p>	<p>Сторінка з формою для реєстрації відкрилась, відкрився особистий профіль користувача.</p>	

<p>Ввести необхідні дані в поля</p> <p>Натиснути кнопку «Зареєструватись»</p> <p>Порівняти фактичний результат з очікуваним</p>	<p>В БД у таблиці «Users» з'явився новий запис з інформацією про зареєстрованого користувача.</p>
<p>Результат виконання тест-кейсу: пройдено успішно</p>	

В результаті проведеної реєстрації нового користувача у БД в таблиці «Users» з'явився новий запис, який підтверджує успішність проведеної операції (Рисунок 3.9).

	id	nickname	surname	name	avatar	phone_number	email	user_type_id
<input type="checkbox"/>	36	viktorina123	Дрозд	Вікторія	uploads/cXg18xQnHsG6KvBF6nDvQ9l8WVnha8a8FWFu p	+380985754783	vikadina@gmail.com	2
<input type="checkbox"/>	38	nika_alex	Павлова	Олексія	uploads/14PheVUuy4q1LULyU0228SUs96XsbA2u0IKUC...	+380985940347	nika_alex@gmail.com	2
<input type="checkbox"/>	39	gorbalenko	Горбатово	Анатолій	uploads/4k18ra1Bcl9UMr22k2ByeCjoo6/VpHoNG3zq5...	+380985735274	gorbalenko1@gmail.com	2
<input type="checkbox"/>	40	yu.maestro	Шутов	Юрій	uploads/Y3oush1P288HrWfC0uyKQV0KjEipuzY17Nvj...	+38098423422	maestro@gmail.com	2
<input type="checkbox"/>	42	_didzcy	Діженко	Діана	uploads/UyFuD6uzdFmgOUHxqHiNqdfvHJFvMvYNNxk...	+380984832948	diana@gmail.com	2
<input type="checkbox"/>	43	viktorina	Посталує	Вікторія	uploads/c6KIK3QpR6W6p2HxvUPYCMsxyIPVlwFR07g3h...	+380984830784	viktorinas@gmail.com	2
<input type="checkbox"/>	44	tankevich_vlad	Танкевич	Владислав	uploads/bd51aus7RknWkaaWkWF7GBIA7W5L7xGqilINR p	+380948382912	vladtankevich@gmail.com	2
<input type="checkbox"/>	48	blazhenko.2020	Блаженко	Олексія др	uploads/abDc5dUB47YZUddtHsCM11B0KnoV8JmV9kk...	+380984932873	blazhenko2020@gmail.com	2
<input type="checkbox"/>	47	pasichnyk	Пасичник	Марія	uploads/1Z3jpaYVJZQV6ccLs11Lmo8sXKLyFWIen93...	+380984938294	pasichnyk@gmail.com	2
<input type="checkbox"/>	48	petlenko	Петренко	Сергія	uploads/HH9p5uJbnZrLrVVK8v5pEK5opu7Q88NHmCqY...	+380984782940	petlenko@khmu.edu.ua	2
<input type="checkbox"/>	51	alex	Ліщук	Олександр	uploads/DQ8d4jQJCh8DFtJ9bYH9vpFchVsb466nmFgRaj...	+380984938290	alex@gmail.com	2
<input type="checkbox"/>	53	nadina_snezhana	Рудіна	Сніжана	uploads/gA793RprS18NJC6OP62IwGK7elgtpv5CicGnR...	+380984938783	nadinasn@gmail.com	2
<input type="checkbox"/>	55	dima_honcharenko	Гончаренко	Дмитро	uploads/qysajlFmPgDrCVPMfDuIBRYY3O3AIs2S1JmC4F7e...	+380984948430	honcharenko@gmail.com	2
<input type="checkbox"/>	58	panasyuk.sasha	Панасюк	Олексія ддр	uploads/HyYQRgK96IowVkuDXrHsRbqYucOsar0y5Lb...	+380984990257	panasyuk@gmail.com	2
<input type="checkbox"/>	58	taras_trojan	Троян	Тарас	uploads/DY1qsNa1UbpFHl1Qq18GVZhu6Kz1Lxpo1gLn...	+380984937286	trojan@gmail.com	2
<input type="checkbox"/>	61	lesia_new	Лесюк	Леся	uploads/HoYVEU4qG7S212gDpNoJNY68OKNFC0kQkLTVp...	0885753384	lesia@gmail.com	2
<input type="checkbox"/>	62	andrew	Сидоренко	Андрій	avatar.png	+38098590921	andrews@gmail.com	2

Рисунок 3.9 – Підтвердження збереження інформації про нового користувача

Крім цього, цю інформацію можна переглянути на особистому профілі користувача (Рисунок 3.10). Отже, перевірку функціоналу для реєстрації нових користувачів успішно пройдено.

На особистій сторінці користувача є форма для публікації допису, скориставшись якою користувач має змогу опублікувати власний літературний твір. Тому, наступний тест-кейс перевірятиме можливість публікації дописів користувачами вебресурсу з обміну літературними творами (Таблиця 3.3).

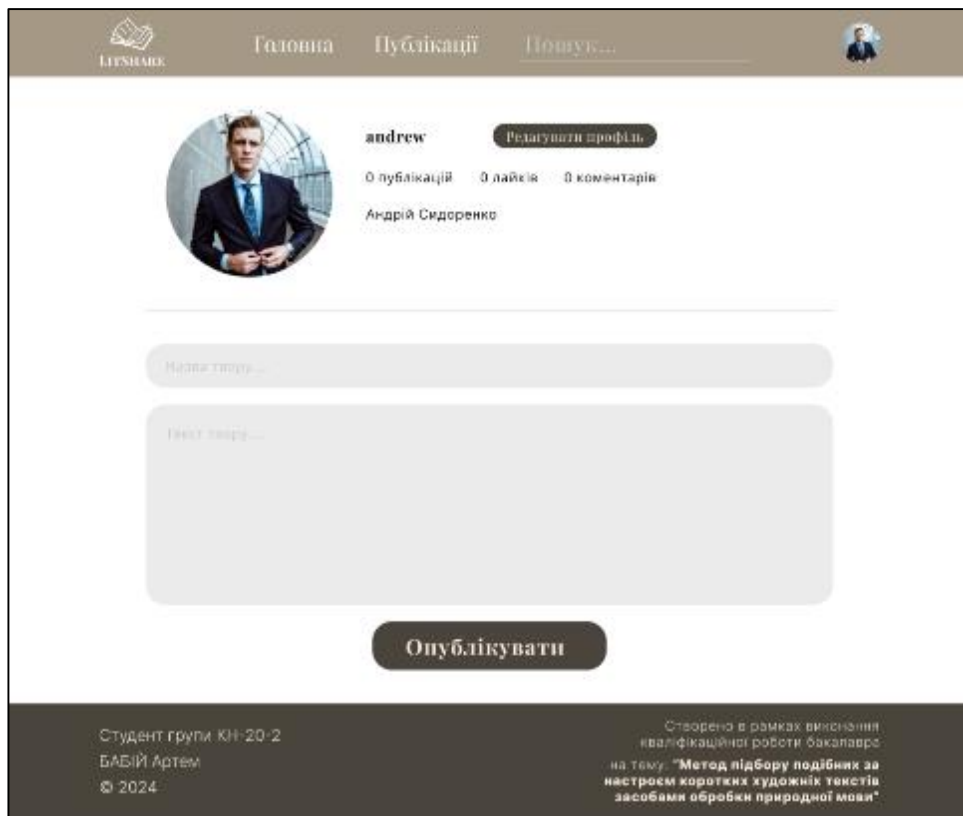


Рисунок 3.10 – Особистий профіль користувача

Таблиця 3.3 – Тест-кейс АТ0003

Тест-кейс ID: АТ0003	Пріоритет: 1	Створено: 24.05.2024, А. Бабій
Назва: Перевірка процесу публікації дописів		
Вхідні дані: Текст твору, назва твору		
Кроки	Очікуваний результат	
<p><i>Передумова:</i> користувач повинен бути авторизованим у системі</p> <p>Відкрити вебресурс</p> <p>Ввійти в особистий профіль користувача</p> <p>Відкрити особисту сторінку</p> <p>Ввести текстовий вміст твору у відповідне поле</p> <p>Ввести назву твору у відповідне</p>	<p>Особиста сторінка користувача перезавантажилась. На сторінці відображається новий допис з визначеним емоційним настроєм, який переважає в творі.</p> <p>Очікуване визначення емоційного настрою: оптимістичне.</p> <p>Фактичне визначення емоційного</p>	

<p>поле</p> <p>Натиснути на кнопку «Опублікувати»</p> <p>Порівняти фактичний результат з очікуваним</p>	<p>настрою: оптимістичне.</p> <p>В БД в таблиці «Post» з'явився новий запис з інформацією про допис.</p>
<p>Результат виконання тест-кейсу: пройдено успішно</p>	

Після публікації допису, особиста сторінка користувача перезавантажується, після чого опублікований допис відображається першим у переліку дописів користувача (Рисунок 3.11).

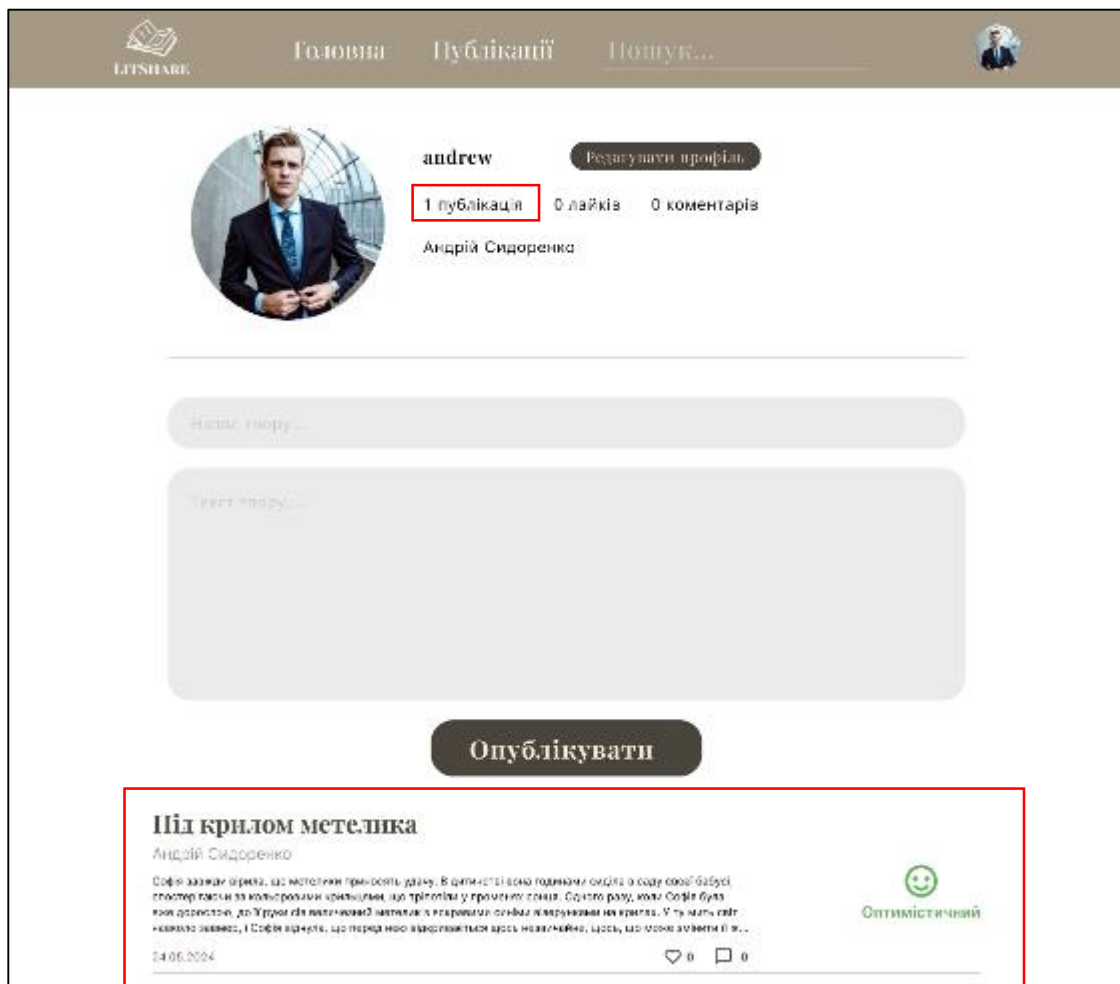


Рисунок 3.11 – Підтвердження публікації допису користувачем

Поряд з дописом відображається його переважаючий емоційний настрій. В даному прикладі, як і очікувалось, система визначила, що цей твір належить до

творів оптимістичного емоційного настрою. Отже, перевірку процесу публікації дописів пройдено успішно.

Крім публікації дописів, користувачі можуть брати участь в обговоренні дописів інших користувачів пишучи їм коментарі. Тому, останній тест-кейс буде перевіряти чи дійсно в користувачів є змога залишати коментарі (Таблиця 3.4).

Після того, як користувач обере допис, який він хоче прокоментувати та залише коментар у формі під дописом, сторінка з обраним дописом перезавантажиться і новий коментар буде відображатись першим у переліку коментарів до допису (Рисунок 3.12). Таким чином, доведено, що у користувачів є можливість залишати коментарі до дописів інших користувачів.

Таблиця 3.4 – Тест-кейс АТ0004

Тест-кейс ID: АТ0004	Пріоритет: 2	Створено: 24.05.2024, А. Бабій
Назва: Перевірка процесу написання коментарів		
Вхідні дані: Обраний допис для коментування, текстовий вміст коментаря		
Кроки	Очікуваний результат	
<p><i>Передумова:</i> користувач повинен бути авторизованим у системі</p> <p>Відкрити вебресурс</p> <p>Ввійти в особистий профіль користувача</p> <p>Відкрити сторінку з переліком опублікованих творів</p> <p>Обрати бажаний твір</p> <p>Написати коментар у формі під дописом</p> <p>Натиснути на кнопку «Відправити»</p> <p>Порівняти фактичний результат з очікуваним</p>	<p>Сторінка обраного допису перезавантажилась. У переліку коментарів, щойно написаний коментар відображається першим.</p> <p>В БД в таблиці «Comments» з'явився новий запис з інформацією про коментар.</p>	

Результат виконання тест-кейсу: пройдено успішно

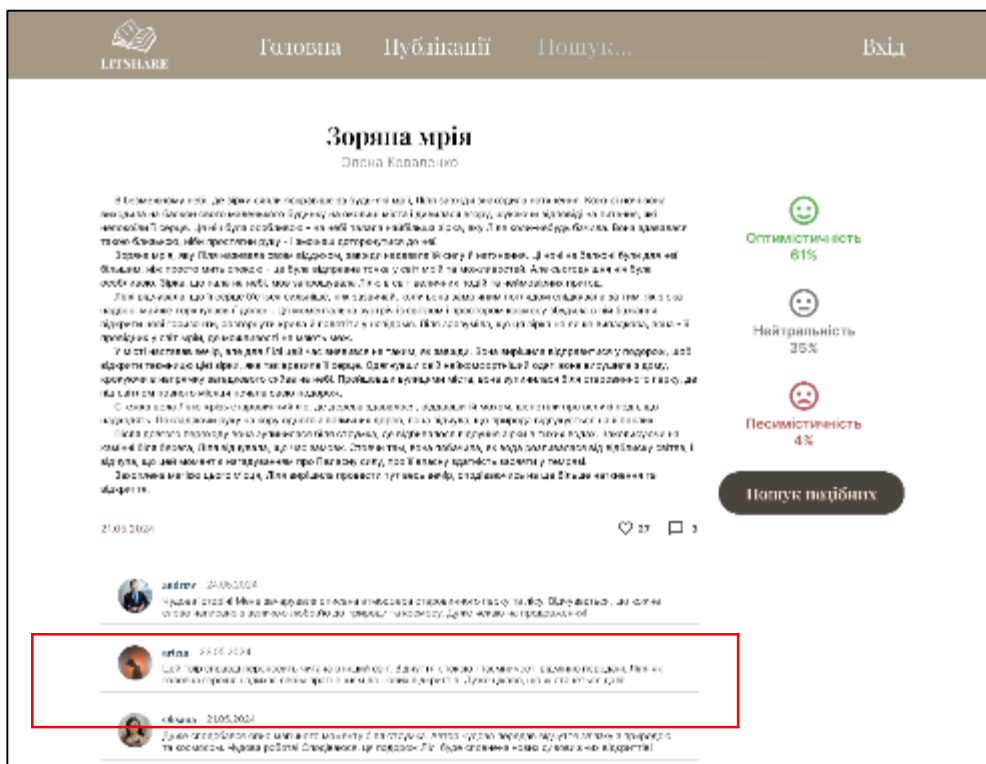


Рисунок 3.12 – Відображення нового коментаря під дописом

Проведене тестування реалізованого вебресурсу з обміну літературними творами доводить, що функціонал для користувачів працює вірно. Адміністратор системи може розширювати множину навчальних текстів шляхом додавання нових текстів та збереження їх у БД. В той час, як користувачі вебресурсу можуть реєструватись та наповнювати особисті профілі контентом, публікуючи власні літературні твори. Також, користувачі мають змогу брати участь в обговоренні творів інших користувачів через коментарі під дописами.

3.6 Аналіз функціональності вебресурсу з підбору подібних за настроєм художніх текстів

Знайомство користувача, незалежно від його ролі, відбувається з головної сторінки вебресурсу з обміну літературними творами (Рисунок 3.13).

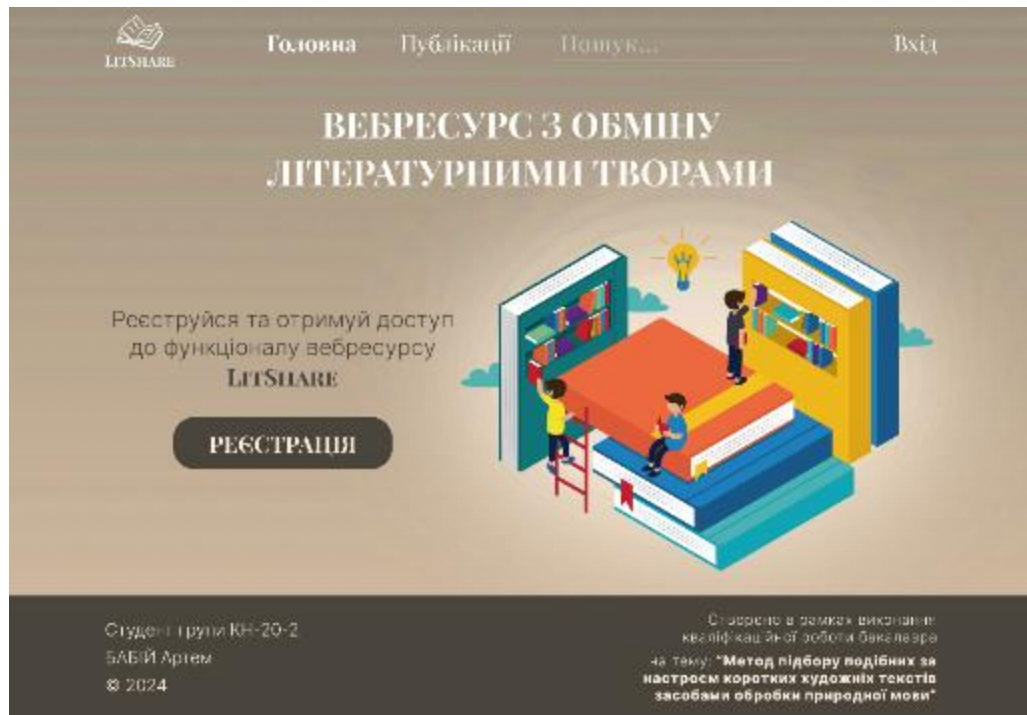


Рисунок 3.13 – Головна сторінка вебресурсу з обміну літературними творами

Далі, користувач може переглядати дописи, які опубліковані на вебресурсі, але без можливості коментувати їх. Для цього користувач може увійти в особистий профіль, якщо він існує або зареєструватись. Щоб увійти необхідно натиснути на кнопку «Вхід» в головному меню вебресурсу. Після цього відкриється нова сторінка з формою для авторизації. Користувачу потрібно ввести його нікнейм та пароль, які він використовував при реєстрації профілю (Рисунок 3.14).

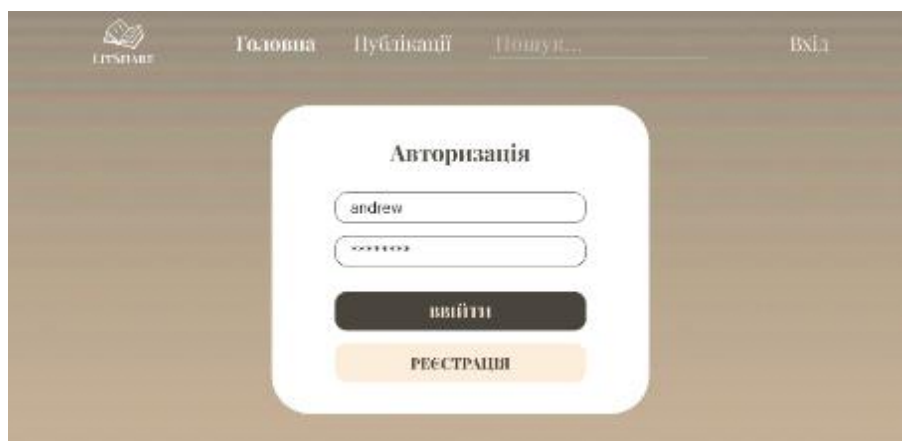


Рисунок 3.14 – Процес авторизації на вебресурсі

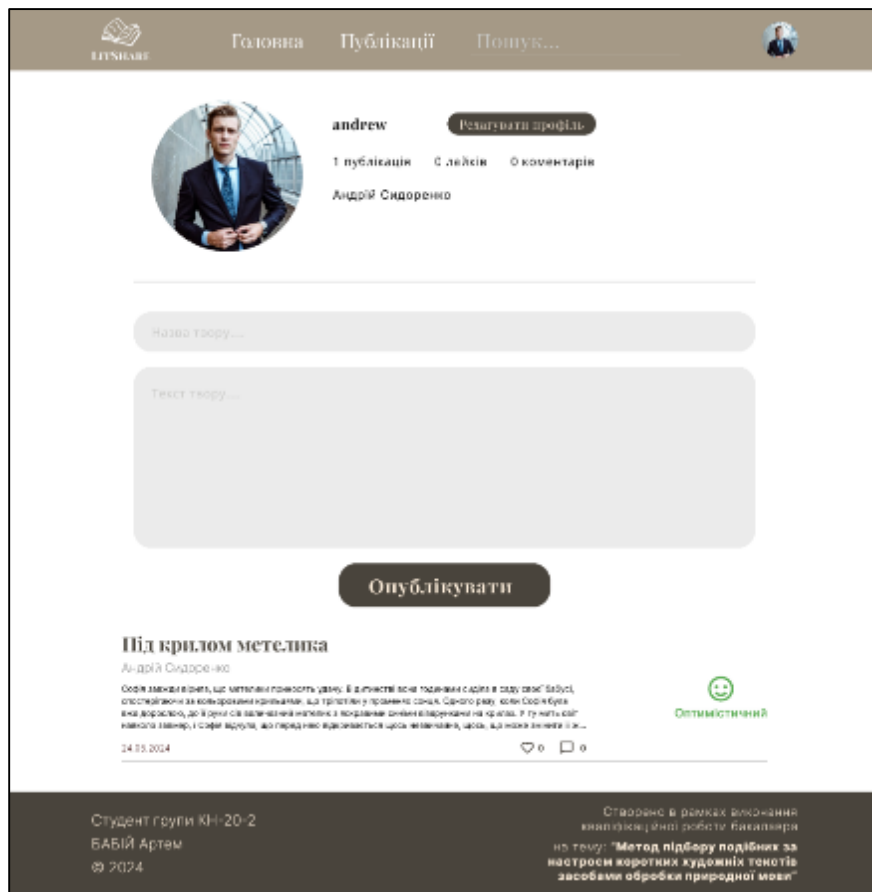


Рисунок 3.15 – Особистий профіль користувача

У випадку успішної авторизації користувач потрапляє на свій особистий профіль, на якому він може побачити інформацію про себе та опубліковані твори (Рисунок 3.15).

Користувач може редагувати особисту інформацію натиснувши на кнопку «Редагувати профіль». Також, користувач може публікувати нові твори, скориставшись формою. Для цього, потрібно ввести назву твору у відповідне поле та текстовий вміст твору (Рисунок 3.16).

Після цього потрібно натиснути на кнопку «Опублікувати», що призведе до перезавантаження сторінки. Новий допис буде відображатись першим в переліку дописів користувача.

Користувач може переглядати дописи інших користувачів натиснувши на кнопку в головному меню з назвою «Публікації». Це призведе до відкриття нової сторінки, де дописи відображаються в хронологічному порядку (Рисунок 3.17).

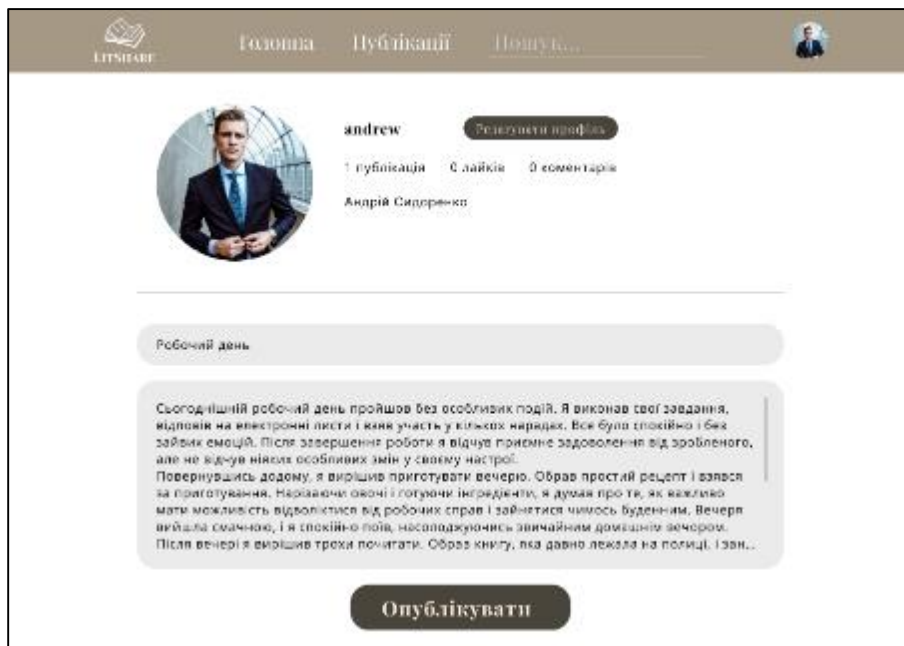


Рисунок 3.16 – Процес публікації нового допису

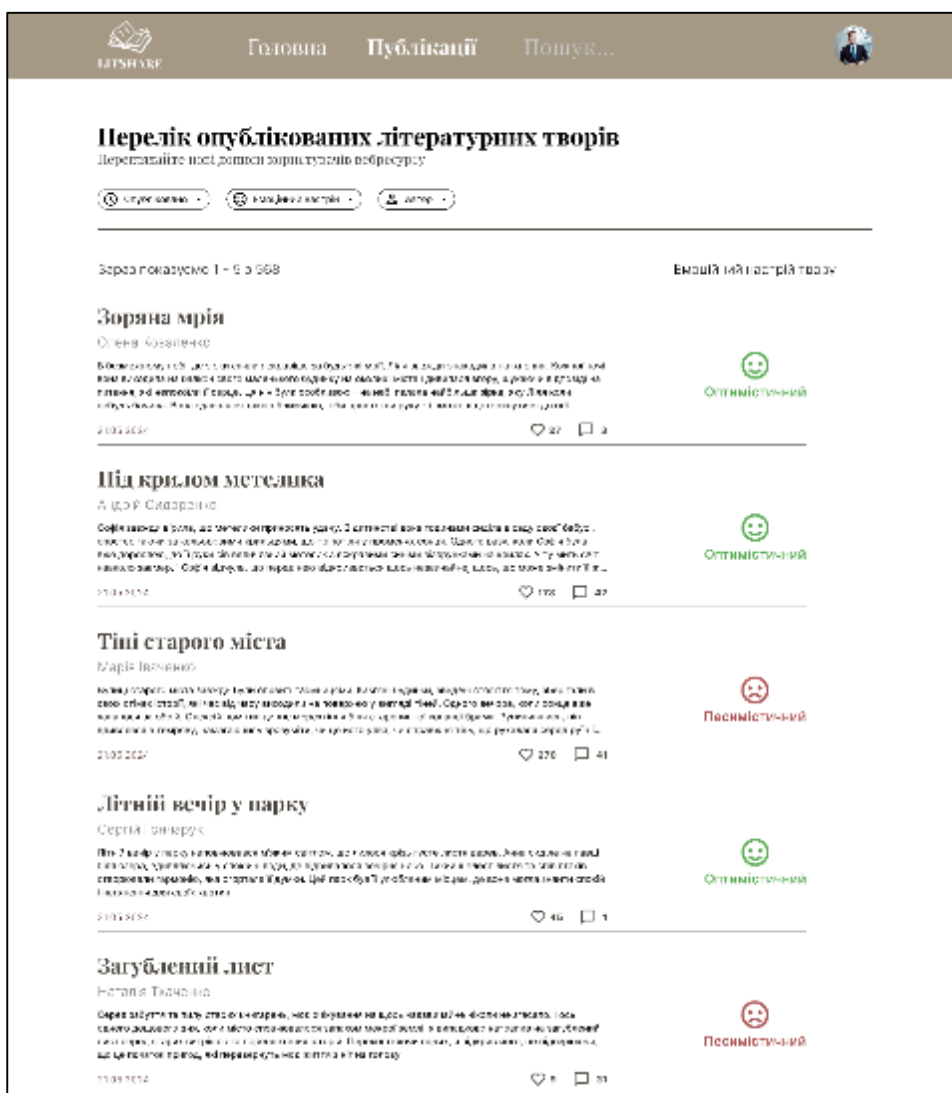


Рисунок 3.17 – Перелік опублікованих літературних творів

Біля кожного допису, користувач може побачити оцінку емоційного настрою твору, яка відображається у вигляді смайлика та напису. Також, для кращого візуального сприйняття, кожен емоційний настрій виділяється своїм кольором (оптимістичний – зелений, нейтральний – сірий, песимістичний – червоний). Настрій який відображується на цій сторінці є настроєм, який переважає в тексті.

Крім цього, користувач може скористатись функцією сортування та фільтрації. Користувач може відсортувати дописи за датою публікації, та відфільтрувати за переважаючим емоційним настроєм та автором.

Для того щоб переглянути повний текст твору та додаткову інформацію про оцінку емоційного настрою твору, користувачу потрібно натиснути на бажаний допис. Після цього, відкриється нова сторінка з обраним дописом на якій можна переглянути повний текст твору, відсоткову оцінку прояву настроїв та перелік коментарів (Рисунок 3.18).

The screenshot shows a post on the LITSHARE website. At the top, there are navigation links: 'Головна', 'Публікації', and 'Пошук...'. The post title is 'Зоряна мрія' by 'Олена Коваленко'. The main text of the post is partially visible, discussing a dream and a journey. To the right of the text is a mood analysis bar with three categories: 'Оптимістичність 61%' (green), 'Нейтральність 35%' (grey), and 'Песимістичність 4%' (red). Below the text, there are three comments from users 'anisa' (dated 24.08.2024), 'anisa' (dated 23.03.2024), and 'Olga' (dated 21.05.2024). At the bottom of the page, there is a footer with the text: 'Студент групи КН-20-2 БАБІЙ Артем © 2024' and 'Створено в рамках виконання кваліфікаційної роботи бакалавра на тему: "Метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови"'. There is also a button labeled 'Пошук подібних'.

Рисунок 3.18 – Сторінка обраного твору

Саме на цій сторінці користувач має змогу провести пошук подібних за настроєм творів. Для цього потрібно натиснути на кнопку «Пошук подібних», яка знаходить з правого боку від тексту під переліком оцінок прояву настроїв. Після цього відкриється нова сторінка з переліком творів, подібних до обраного (Рисунок 3.19).

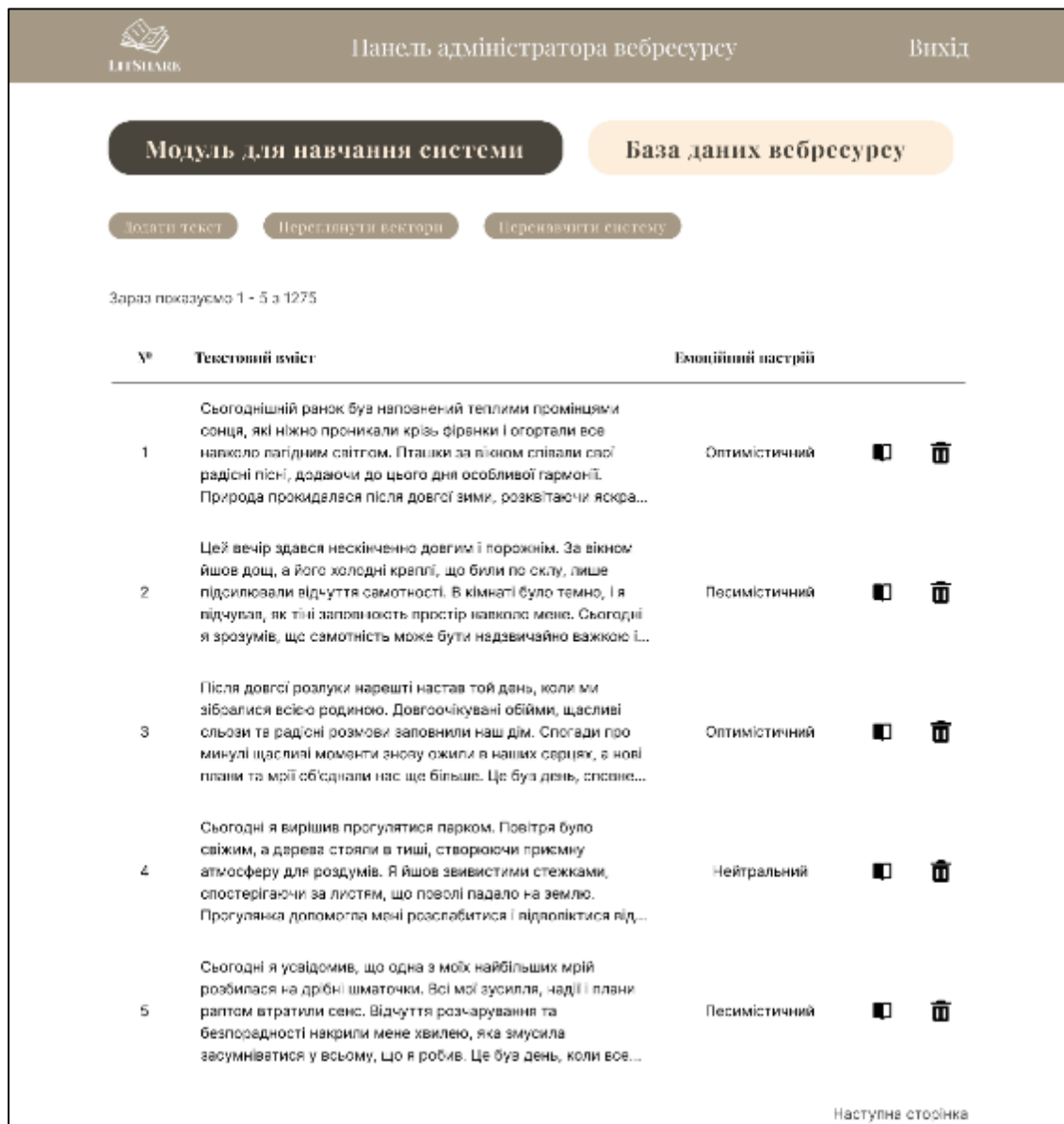
Ступінь подібності творів можна побачити з правого боку в форматі відсоткового числа. Крім цього, на цій сторінці користувач також може скористатись сортуванням дописів за датою та ступенем подібності. За замовчуванням дописи відображуються від найбільш подібного.

The screenshot shows the 'LITSHARE' website interface. At the top, there are navigation links: 'Головна', 'Публікації', and 'Пошук...'. A user profile picture is visible in the top right corner. The main heading is 'Результати пошуку подібних текстів за смодійним настроєм'. Below this, it says 'Пошук здійснюється для твору "Зорина мрія"'. There are two filter buttons: 'Сортування' and 'Ступінь подібності'. The search results are displayed in a list format, showing the title of the poem, the author's name, a short excerpt, and the similarity percentage. Each result also includes the date of publication, the number of likes, and the number of comments.

Твір	Автор	Степінь подібності творів
Під крилом метелика	Андрій Сидоренко	77.3%
Літній вечір у парку	Сергій Гончарук	72.7%
Мелодія дощу	Дмитро Шевченко	69.8%
Птахи на горизонті	Оксана Павчанка	64.2%
Спогади про майбутнє	Володимир Поліщук	61.4%

Рисунок 3.19 – Сторінка з переліком подібних творів

Для того щоб мати доступ до панелі адміністратора користувач повинен авторизуватись на профілі з правами адміністратора. Після авторизації адміністратор потрапляє на сторінку з переліком навчальних текстів на яких відбувається навчання методу підбору подібних за настроєм коротких художніх текстів (Рисунок 3.20).



Панель адміністратора вебресурсу

Вихід

Модуль для навчання системи

База даних вебресурсу

Додати текст

Переглянути вектори

Перенавчити систему

Зараз показуємо 1 - 5 з 1275

№	Текстовий зміст	Емоційний настрій
1	Сьогоднішній ранок був наповнений теплими промінцями сонця, які ніжно проникали крізь фіранки і огортали все навколо лагідним світлом. Пташки за віном співали свої радісні пісні, додаючи до цього дня особливої гармонії. Природа прокидалася після довгої зими, розквітаючи яскра...	Оптимістичний
2	Цей вечір здався нескінченно довгим і порожнім. За вікном йшов дощ, а його холодні краплі, що били по склу, лише підсилювали відчуття самотності. В кімнаті було темно, і я відчував, як тіні заповнюють простір навколо мене. Сьогодні я зрозумів, що самотність може бути надзвичайно важкою і...	Песимістичний
3	Після довгої розлуки нарешті настав той день, коли ми зібралися всією родиною. Довгоочікувані обійми, щасливі сльози та радісні розмови заповнили наш дім. Спогади про минулі щасливі моменти знову ожили в наших серцях, а нові плани та мрії об'єднали нас ще більше. Це був день, сповне...	Оптимістичний
4	Сьогодні я вирішив прогулятися парком. Повітря було свіжим, а дерева стояли в тиші, створюючи приємну атмосферу для роздумів. Я йшов звивистими стежками, спостерігаючи за листям, що поволі падало на землю. Прогулянка допомогла мені розслабитися і відволіктися від...	Нейтральний
5	Сьогодні я усвідомив, що одна з моїх найбільших мрій розбилася на дрібні шматочки. Всі мої зусилля, надії і плани раптом втратили сенс. Відчуття розчарування та безпорадності накрили мене хвилею, яка змусила засумніватися у всьому, що я робив. Це був день, коли все...	Песимістичний

Наступна сторінка

Рисунок 3.20 – Вигляд сторінки з переліком навчальних текстів

Адміністратор може видаляти тексти із БД натиснувши на відповідну іконку. Також, адміністратор може відкрити текст та переглянути його повністю або відредагувати. З метою розширення множини навчальних текстів для

адміністратора реалізовано можливість додавання нових навчальних текстів. Для цього потрібно натиснути на кнопку «Додати текст», що призведе до відкриття модального вікна з формою для додавання тексту (Рисунок 3.21). Для того щоб додати текст потрібно вибрати його емоційний настрій у випадяючому списку та додати текстовий вміст у поле нижче. Далі, адміністратору потрібно натиснути на кнопку «Зберегти до БД». Після цього модальне вікно закриється а текст збережеться до БД та буде відображатись в переліку навчальних текстів.

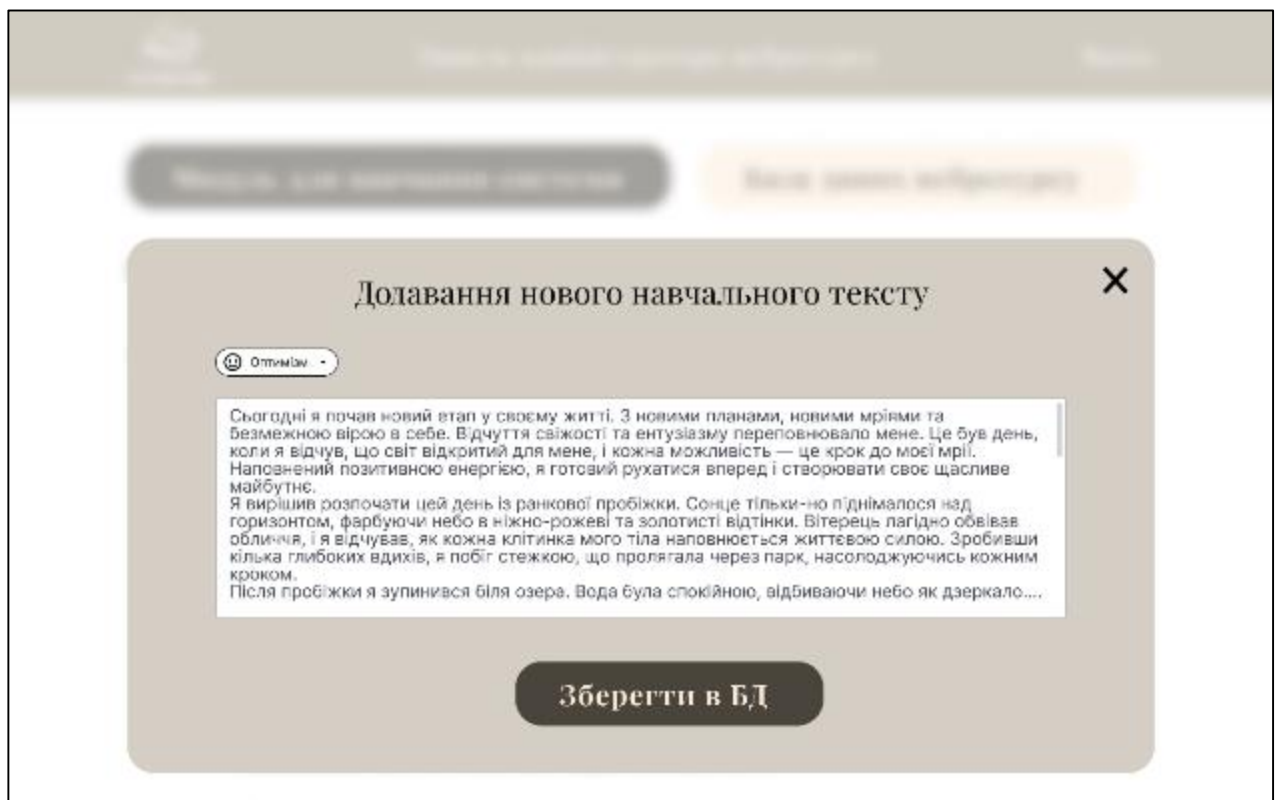


Рисунок 3.21 – Додавання нового навчального тексту

Після додавання нового тексту адміністратор може перенавчити систему. Для цього потрібно натиснути на кнопку «Перенавчити систему». Також, адміністратор може переглянути сформовані вектори значущих слів для кожного типу настрою. Для цього необхідно натиснути на кнопку «Переглянути вектори», що призведе до відкриття нової сторінки. На цій сторінці буде відображатись перелік слів і їхня вага, окремо для кожного з настроїв (Рисунок 3.22).

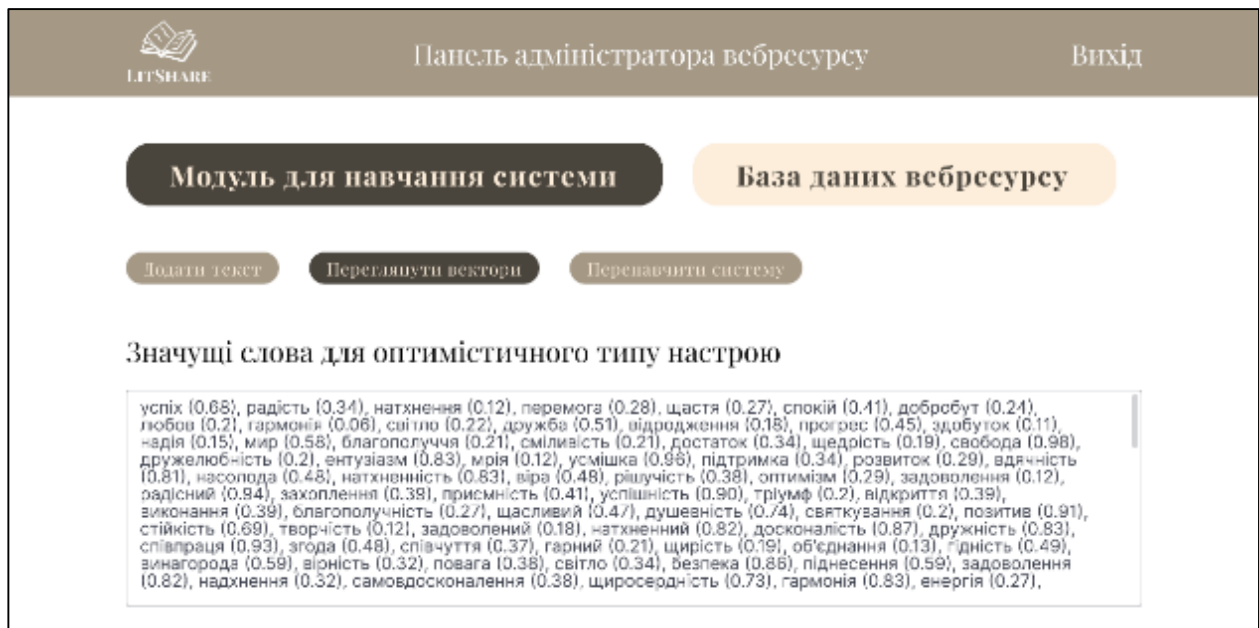


Рисунок 3.22 – Перелік значущих слів для оптимістичного типу настрою

Аналіз функціональності вебресурсу з обміну літературними творами показує, що реалізований програмний застосунок виконує поставленні завдання та демонструє роботу методу підбору подібних за настроєм коротких художніх текстів.

3.7 Результати досліджень ефективності методу підбору подібних за настроєм коротких художніх текстів

Використовуючи створений вебресурс з обміну літературними творами буде досліджено ефективність роботи методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови. Дослідження ефективності буде відбуватись у два етапи. Спочатку буде досліджено наскільки правильно програма здатна визначати емоційний настрій творів. Після цього буде досліджено другу частину методу, яка полягає в підборі подібних за настроєм художніх текстів.

Для дослідження буде використано набір із 150 текстів (50 оптимістичного настрою, 50 нейтральних та 50 песимістичних). Тексти для цього набору будуть згенеровані з використанням ChatGPT [20]. Для генерації

тексту достатньо написати запит в якому вказати емоційний настрій тексту (Рисунок 3.23). Нова модель GPT-4o здатна генерувати якісні твори, які важко відрізнити від творів створених людиною. Тому, згенеровані твори через ChatGPT відмінно підходять для задачі дослідження ефективності роботи методу.

Отже, спершу буде визначено домінуючий емоційний настрій згенерованих творів використовуючи розроблений метод. Якщо визначення емоційного настрою через реалізовану програму збігатиметься із емоційним настроєм твору, який вказувався при його генерації, то такий випадок визначення емоційного настрою буде вважатись вірним. В іншому ж випадку, визначення емоційного настрою буде вважатись помилковим.

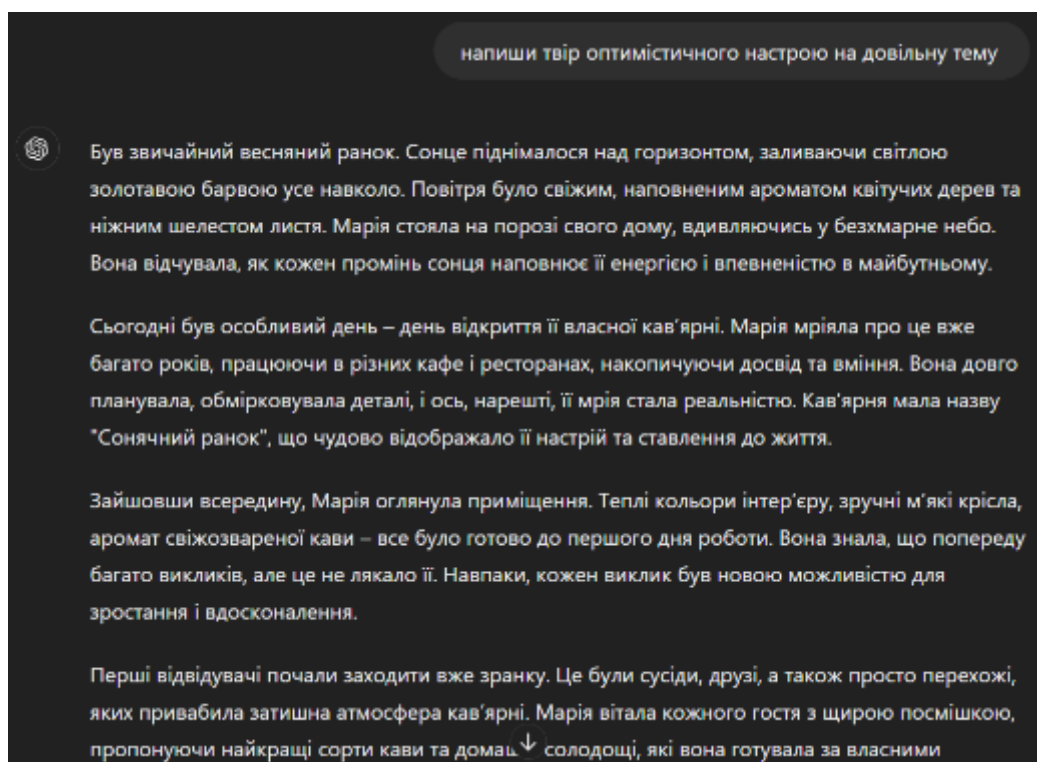


Рисунок 3.23 – Приклад генерація твору оптимістичного настрою через ChatGPT

Результати визначення емоційного настрою творів за допомогою створеної програми та порівняння із вказаним емоційним настроєм при генерації творів через ChatGPT наведено в таблиці 3.5.

Таблиця 3.5 – Результати визначення емоційного настрою творів

	Оптимістичний настрій	Нейтральний настрій	Песимістичний настрій
Кількість правильно визначених творів за емоційним настроєм	45	38	47
Кількість неправильно визначених творів за емоційним настроєм	5	12	3
Разом творів:	50	50	50

Відповідно до отриманих результатів можна визначити точність визначення емоційного настрою з використанням розробленого методу (Таблиця 3.6).

Таблиця 3.6 – Точність визначення емоційного настрою

	Оптимістичний настрій	Нейтральний настрій	Песимістичний настрій
Точність	90%	76%	94%

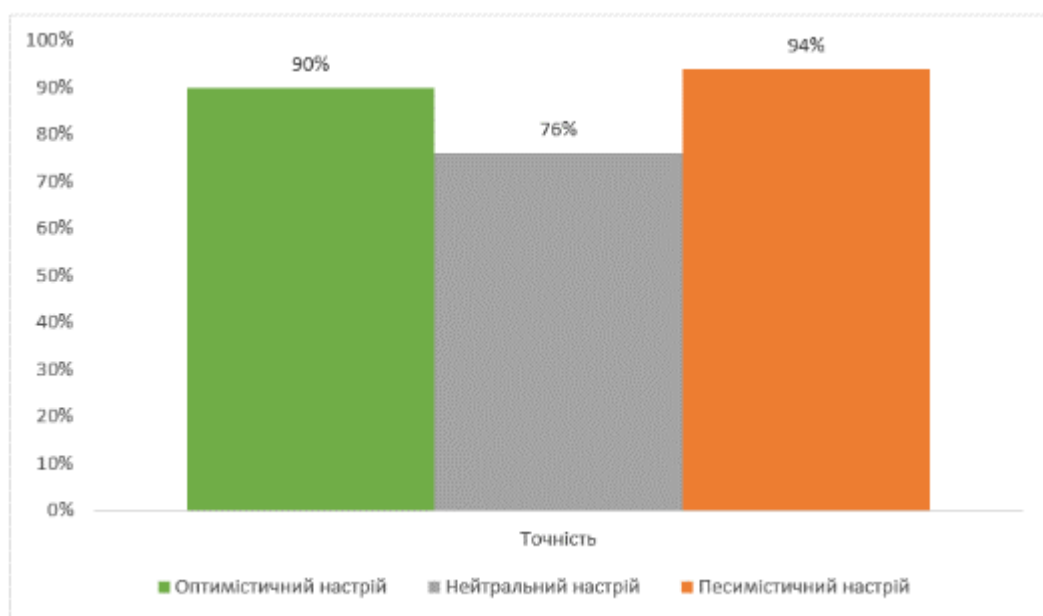


Рисунок 3.24 – Діаграма точності визначення емоційного настрою творів

Як видно з таблиці 3.6 та діаграми зображеної на рисунку 3.24 найкраще програма змогла визначити емоційний настрій в творах песимістичного характеру. Найгірші результати програма показує з творами нейтрального характеру. Це може бути пов'язано з тим, що ідентифікувати нейтральний емоційний настрій досить важко, адже важко виділити ключові слова, які цей настрій характеризують. Тому, для покращення результатів потрібно розширити множину навчальних дописів для нейтрального типу настрою.

Для визначення точності підбору подібних за настроєм текстів буде використано ту саму тестову вибірку творів. Для кожного твору буде проведено пошук п'яти подібних творів за емоційним настроєм. У випадку, якщо емоційні настрої в творах збігатимуться, то пошук подібних творів вважатиметься вірним.

В результаті проведеного дослідження пошуку подібних творів було отримано результати наведені в таблиці 3.7.

Таблиця 3.7 – Результати пошуку подібних за емоційним настроєм творів

	Для творів оптимістичного настрою	Для творів нейтрального настрою	Для творів песимістичного настрою
Кількість правильно підібраних творів	38	29	42
Кількість неправильно визначених творів за емоційним настроєм	12	21	8
Разом творів:	50	50	50

Відповідно до отриманих результатів можна визначити точність підбору творів за емоційним настроєм з використанням розробленого методу (Таблиця 3.8).

Таблиця 3.8 – Точність визначення емоційного настрою

	Для творів оптимістичного настрою	Для творів нейтрального настрою	Для творів песимістичного настрою
Точність	76%	58%	84%

Для візуалізації отриманих результатів створено діаграму, яка зображена на рисунку 3.25. Як видно з отриманих результатів, можна зробити висновок, що результати підбору подібних за емоційним настроєм творів залежать від якості визначення оцінок емоційних настроїв цих творів.

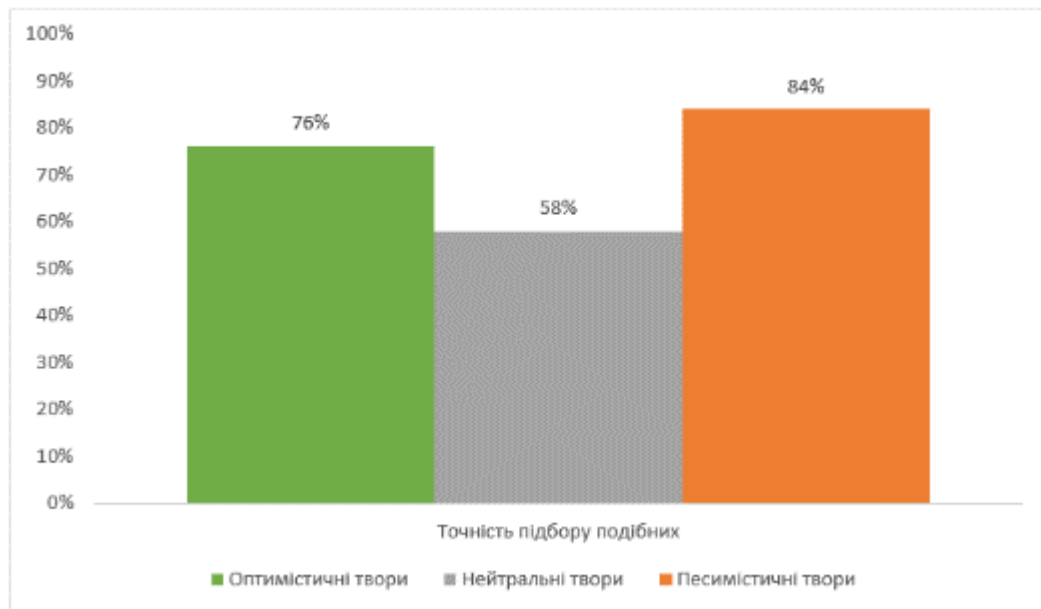


Рисунок 3.25 – Діаграма точності підбору подібних творів за емоційним настроєм

Оскільки, точність визначення емоційного настрою для нейтральних текстів була найнижчою, то якість підбору текстів нейтрального типу також значно гірша ніж оптимістичного та песимістичного. Це знову доводить той факт, що необхідно покращити множину навчальних текстів для нейтрального типу настрою.

Найкращі результати отримано при підборі подібних творів песимістичного характеру. Точність підбору цих текстів склала 84%, що можна

вважати досить хорошим результатом, проте при розширенні множини навчальних текстів можливе покращення показників точності підбору.

Отже, дослідження ефективності роботи методу подібних за настроєм коротких художніх текстів засобами обробки природної мови, показує, що реалізований метод здатний добре працювати з текстами, які мають чітко виражений оптимістичний або песимістичний тип настрою. Проте, при роботі з текстами нейтрального настрою результати значно гірші, що можна пояснити тим, що досить важко побудувати якісний вектор значущих слів для нейтрального типу настрою. Тому, при практичному застосуванні даного методу, можна спробувати обмежитись визначенням лише оптимістичної та песимістичної оцінки прояву емоційного настрою.

3.8 Висновки до розділу 3

Для програмної реалізації методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови у вигляді вебресурсу з обміну літературними творами було використано наступний набір інструментів: мову програмування PHP, фреймворк Laravel та СКБД MySQL. Вибір мови програмування PHP обумовлений тим, що ця мова чудово підходить для створення веб застосунків, оскільки повністю орієнтована на веб розробку. Фреймворк Laravel був обраний через те, що він надає готові рішення для автоматизованого створення системи автентифікації користувачів вебресурсу та допомагає структурувати програмний код застосунку. СКБД MySQL є гнучким та зручним інструментом для збереження даних, який легко інтегрується з веб застосунками, зокрема, через фреймворк Laravel.

Структуру інформаційної системи було реалізовано відповідно до шаблону MVC, яка включає в себе десять моделей, чотирнадцять представлень, сім контролерів та один файл-помічник, де реалізовано метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови. Такий підхід дозволяє розділити програмний код на різні взаємопов'язані компоненти.

Використовуючи зазначений набір інструментів було створено програмну реалізацію методу подібних за настроєм коротких художніх текстів засобами обробки природної мови у вигляді вебресурсу з обміну літературними творами. Створений вебресурс дозволяє користувачам реєструвати особисті профілі та публікувати власні літературні твори для яких автоматично визначаються оцінки прояву емоційних настроїв. Користувачі вебресурсу мають змогу проводити пошук подібних творів за емоційним настроєм та обговорювати опубліковані твори у коментарях. Для управління методом підбору подібних творів, зокрема, для можливості редагування множини навчальних текстів створено панель адміністратора, яка надає доступ до зазначеного функціоналу.

Проведене тестування реалізованої інформаційної системи через тест-кейси доводить, що користувачі вебресурсу з обміну літературними творами мають доступ до потрібного їм функціоналу та мають змогу його використовувати з метою отримання результатів визначення оцінок прояву емоційних настроїв для текстів та пошуку подібних текстів.

Дослідження ефективності роботи реалізованого методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови показало такі показники точності підбору подібних текстів: для оптимістичних – 76%, для нейтральних – 58%, для песимістичних – 84%. Найгірші результати отримано при роботі з текстами нейтрального типу настрою, адже задача визначення цього типу настрою є значно складнішою ніж визначення оптимістичного чи песимістичного типу настрою. Тому, для покращення отриманих результатів потрібно розширити множину навчальних текстів, особливо для текстів нейтрального типу настрою.

Загальні висновки

Кваліфікаційна робота бакалавра присвячена розробці методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови з метою спрощення роботи з літературними сервісами за рахунок автоматизованого підбору подібних за настроєм текстів.

Під час виконання кваліфікаційної роботи бакалавра було досліджено предметну область, яка стосується методів та підходів для визначення емоційного настрою в текстах та способів їх порівняння. Також, було досліджено існуючі програмні рішення та наукові публікації з метою визначення напрямків подальшої роботи. Проведене дослідження предметної області допомогло встановити актуальність даної теми, яка полягає в тому, що збільшення кількості текстів в мережі призводить до потреби їх автоматизованої обробки та пошуку потрібних за певними критеріями, зокрема, за емоційним настроєм.

Для досягнення поставленої мети, яка полягає в спрощенні роботи з літературними сервісами за рахунок автоматизованого підбору подібних за настроєм текстів було розроблено метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови. Розроблений метод складається з трьох послідовних етапів. На першому етапі відбувається формування векторів значущих слів для кожного типу настрою за вхідними даними у вигляді множини навчальних текстів та словників тональності для кожного типу настрою. Далі, використовуючи сформовані вектори значущих слів відбувається обчислення оцінок прояву емоційних настроїв для користувацьких текстів. На останньому третьому етапі відбувається пошук подібних текстів, враховуючи обчислені оцінки прояву настроїв.

Для програмної реалізації розробленого методу було створено вебресурс з обміну літературними творами. Вебресурс створено з використанням мови програмування PHP, фреймворку Laravel, СКБД MySQL та спеціалізованої бібліотеки PHP ML. Створений вебресурс дозволяє користувачам реєструвати

особисті профілі та публікувати власні літературні твори, для яких автоматично визначаються оцінки емоційних настроїв. Також, користувачі вебресурсу можуть шукати подібні твори за емоційним настроєм та обговорювати опубліковані твори у коментарях. Для управління методом підбору подібних творів, зокрема можливості внесення змін в множину навчальних текстів, створено панель адміністратора, яка надає доступ до цього функціоналу. Крім цього, адміністратор може переглядати вектори семантично значущих слів для кожного типу емоційного настрою, що допомагає відслідковувати зміни в системі при редагуванні множини навчальних текстів.

Тестування створеної інформаційної системи для розробленого методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови доводить, що вебресурс використовує розроблений метод та надає користувачам доступ до функціоналу, пов'язаного із використанням можливостей методу.

Дослідження ефективності роботи методу показало, що точність підбору подібних творів оптимістичного настрою становить 76%, нейтрального – 58% та песимістичного – 84%. Отримані результати залежать від множини навчальних текстів та словників тональності, на основі яких формуються вектори значущих слів для кожного типу настрою. Тому для покращення отриманих результатів необхідно розширювати навчальну множину.

Створений вебресурс з обміну літературними творами є навчальним проєктом, мета створення якого полягає в дослідженні ефективності роботи методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови. Оскільки, в ході дослідження отримано досить хороші результати, то розроблений метод може бути використаний для аналізу текстів на прояв емоційних настроїв та пошуку подібних текстів.

Перелік посилань

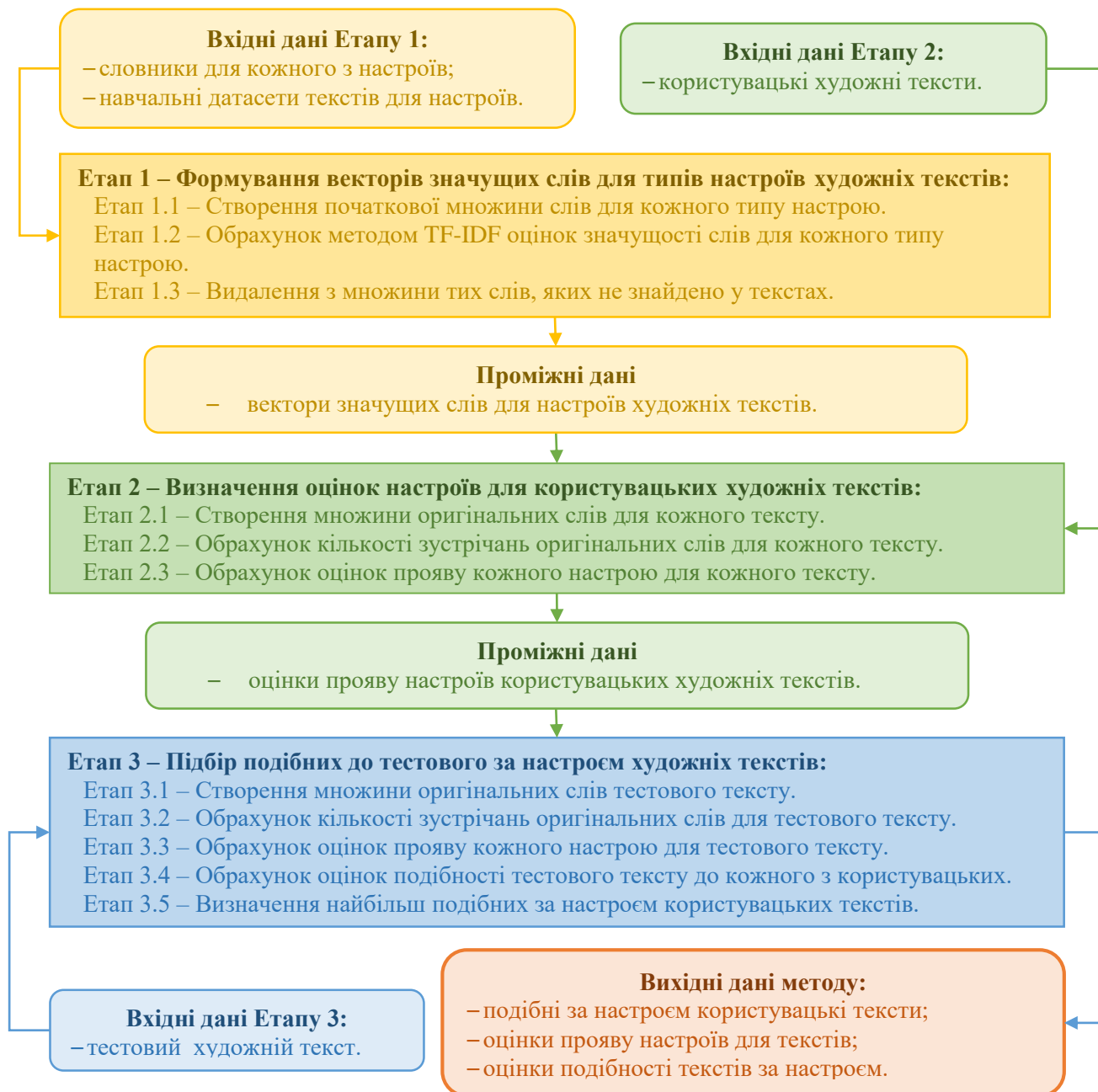
1. Освіта. Аспекти теорії тексту. Твір і текст. Реферат. URL: <https://osvita.ua/vnz/reports/journalism/25024/>
2. Грещук В. В., Грещук О. Б. Словотвір і текст: монографія. Івано-Франківськ: Прикарпатський національний університет імені Василя Стефаника, 2022. 223 с.
3. Allref. Ознаки тексту, його види та форми реалізації. URL: https://allref.com.ua/uk/skachaty/Priznaki_teksta_ego_vidy_i_formy_realizacii
4. Вікіпедія. Стилїстика української мови. URL: https://uk.wikipedia.org/wiki/Стилїстика_української_мови
5. Яблонська Т. М. Лінгвістичний аналіз художнього тексту. Одеса, 2019. 325 с.
6. Психологер. Оптимізм: хто такі оптимісти, переваги та недоліки. URL: <https://psychologer.com.ua/optymizm-khto-taki-optymisty/>
7. Український психологічний хаб. Настрій це. URL: <https://www.psykholoh.com/post/настрій-це>
8. Психологер. Песимізм: ознаки, вплив, переваги та ризики. URL: <https://psychologer.com.ua/pesymizm/>
9. Тихоніна С. І. Лінгвальні засоби відтворення емоцій в тексті художнього твору. Закарпатські філологічні студії. 2023. № 30. С. 115–118.
10. IBM. What is NLP. URL: <https://www.ibm.com/topics/natural-language-processing>
11. GeeksforGeeks. Naive Bayes Classifiers. URL: <https://www.geeksforgeeks.org/naive-bayes-classifiers/>
12. IBM. How SVM Works? URL: <https://www.ibm.com/docs/en/spss-modeler/saas?topic=models-how-svm-works>
13. GeeksforGeeks. Introduction to Recurrent Neural Network. URL: <https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/>

14. GeeksforGeeks. Introduction to Convolution Neural Network. URL: <https://www.geeksforgeeks.org/introduction-convolution-neural-network/>
15. Оленич І., Притула М., Сінькевич О, Хамар О. Система автоматизованого визначення тональності тексту. Електроніка та інформаційні технології. 2021. №15. С. 16–23.
16. Вікіпедія. TF-IDF. URL: <https://uk.wikipedia.org/wiki/TF-IDF>
17. Рябишев О. В. Дослідження методів аналізу емоційного забарвлення тексту українською мовою. Харків. Харківський національний університет радіоелектроніки, 2021. 52 с.
18. Біленький В. С., Кобзев В. Г., Матвеев Д. І. Методи оцінювання тонального забарвлення тексту. Матеріали Міжнародної науково-технічної конференції здобувачів вищої освіти та молодих вчених. Івано-Франківськ, 2020. С. 89–96.
19. MonkeyLearn. Sentiment Analyzer. URL: <https://monkeylearn.com/sentiment-analysis-online/>
20. OpenAI. ChatGPT. URL: <https://chatgpt.com>
21. GitHub. Tone-dict-uk. URL: <https://github.com/lang-uk/tone-dict-uk>
22. PHP-ML. URL: <https://php-ml.readthedocs.io/en/latest/>
23. PHP-ML. TF-idf Transformer. URL: <https://php-ml.readthedocs.io/en/latest/machine-learning/feature-extraction/tf-idf-transformer/>
24. Code Institute. What is PHP? Uses & Introduction. URL: <https://codeinstitute.net/global/blog/what-is-php-programming/>
25. Laravel. URL: <https://laravel.com/>
26. Oracle. What is MySQL? URL: <https://www.oracle.com/mysql/what-is-mysql/>
27. Освіта. Що таке MVC? URL: <https://uk.education-wiki.com/3886982-what-is-mvc>

ДОДАТКИ

Додаток А

Схема методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови



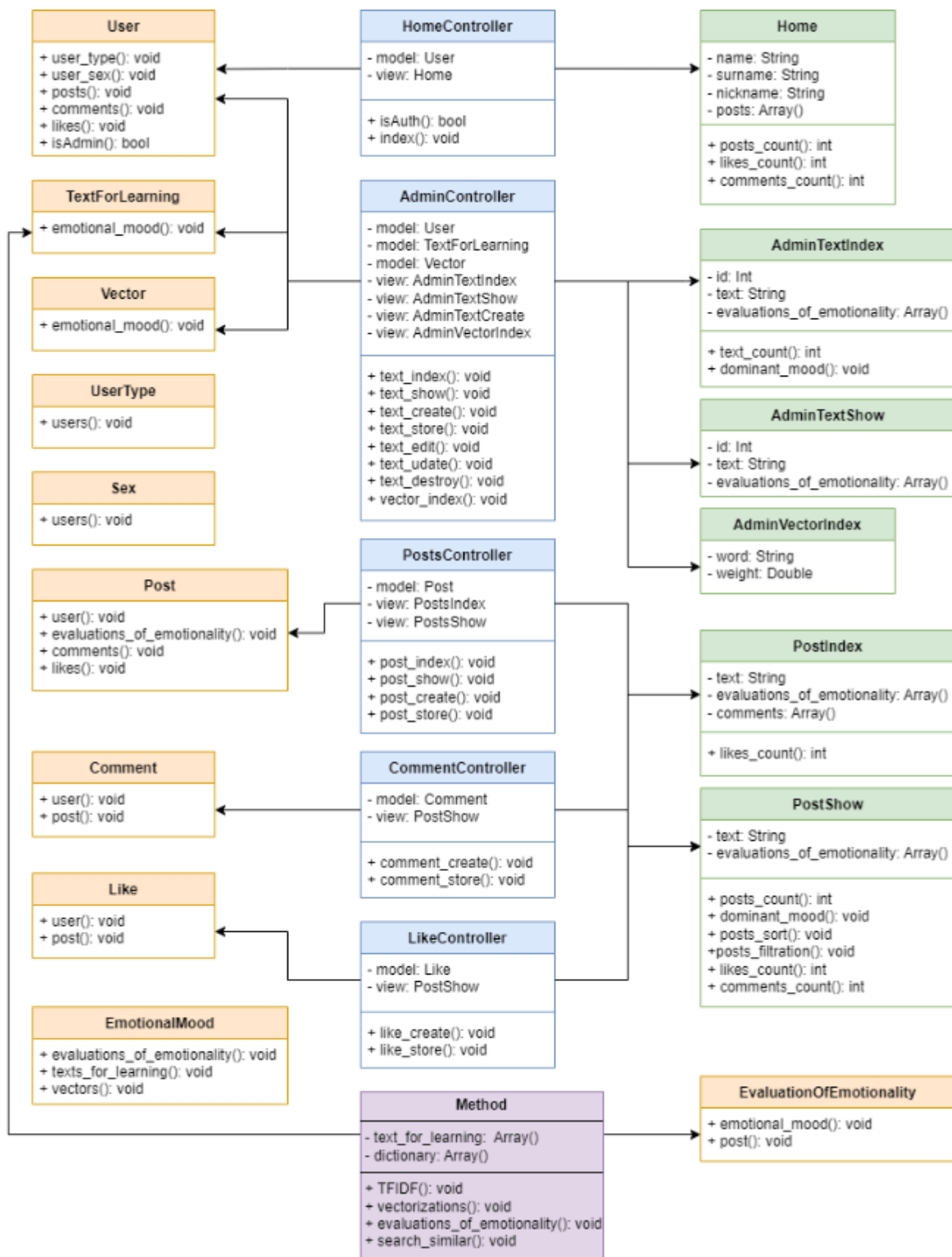
Додаток Б

Схема структури вебресурсу з обміну літературними творами



Додаток В

Діаграма класів вебресурсу з обміну літературними творами



Додаток Г

Презентаційний матеріал

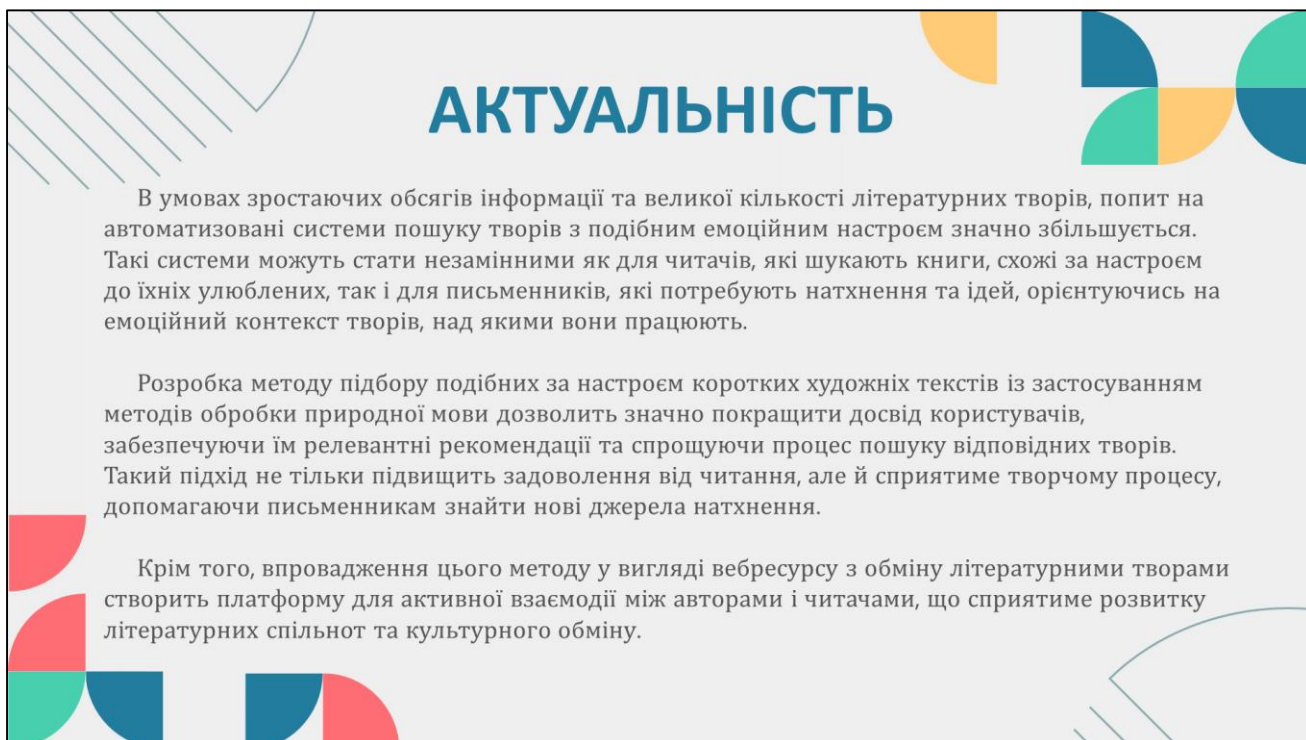


КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

МЕТОД ПІДБОРУ ПОДІБНИХ ЗА НАСТРОЄМ КОРОТКИХ ХУДОЖНІХ ТЕКСТІВ ЗАСОБАМИ ОБРОБКИ ПРИРОДНОЇ МОВИ

Виконав:
студент групи КН-20-2
Бабій Артем

Керівник:
к.т.н., доцент кафедри КН
Пасічник Олександр



АКТУАЛЬНІСТЬ

В умовах зростаючих обсягів інформації та великої кількості літературних творів, попит на автоматизовані системи пошуку творів з подібним емоційним настроєм значно збільшується. Такі системи можуть стати незамінними як для читачів, які шукають книги, схожі за настроєм до їхніх улюблених, так і для письменників, які потребують натхнення та ідей, орієнтуючись на емоційний контекст творів, над якими вони працюють.

Розробка методу підбору подібних за настроєм коротких художніх текстів із застосуванням методів обробки природної мови дозволить значно покращити досвід користувачів, забезпечуючи їм релевантні рекомендації та спрощуючи процес пошуку відповідних творів. Такий підхід не тільки підвищить задоволення від читання, але й сприятиме творчому процесу, допомагаючи письменникам знайти нові джерела натхнення.

Крім того, впровадження цього методу у вигляді вебресурсу з обміну літературними творами створить платформу для активної взаємодії між авторами і читачами, що сприятиме розвитку літературних спільнот та культурного обміну.

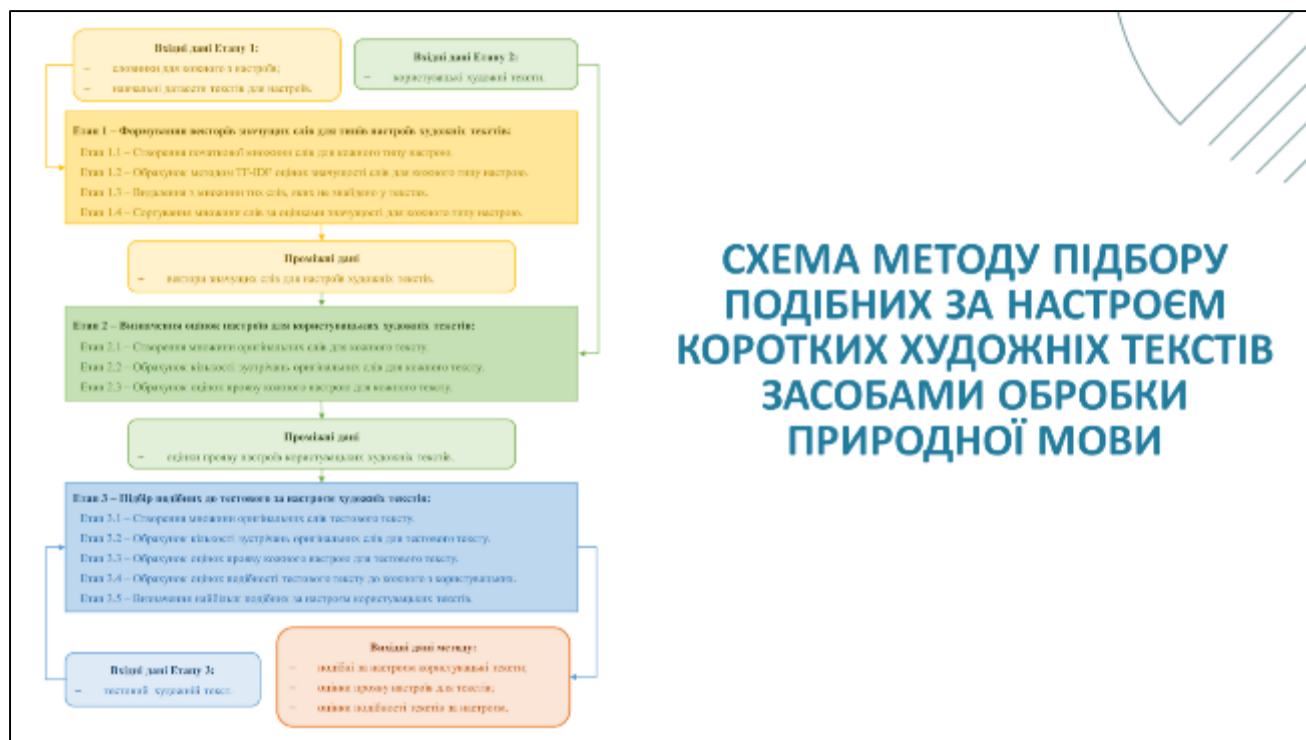
МЕТА РОБОТИ

Мета кваліфікаційної роботи бакалавра – спрощення роботи з літературними сервісами за рахунок автоматизованого підбору подібних за настроєм текстів.

Для досягнення поставленої мети необхідно розробити метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови та програмно реалізувати у вигляді вебресурсу з обміну літературними творами.

ПОСТАВЛЕНІ ЗАВДАННЯ





РОЗПОДІЛ СЛІВ ЗА ЕМОЦІЙНИМ НАСТРОЄМ У СЛОВНИКАХ ТОНАЛЬНОСТІ

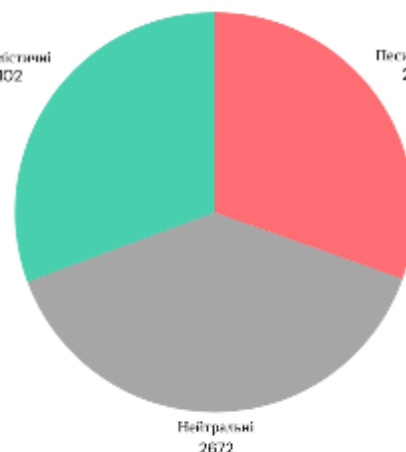
ОПТИМІСТИЧНІ

Оптимістичні слова характеризуються позитивними емоціями, надією, впевненістю, радістю та енергійністю. Вони часто асоціюються з успіхом, щастям і перспективами на краще майбутнє.

НЕЙТРАЛЬНІ

Нейтральні слова характеризуються відсутністю виразних позитивних або негативних емоцій. Вони зазвичай є об'єктивними, фактичними та не викликають сильних емоційних реакцій.

Оптимістичні
2102



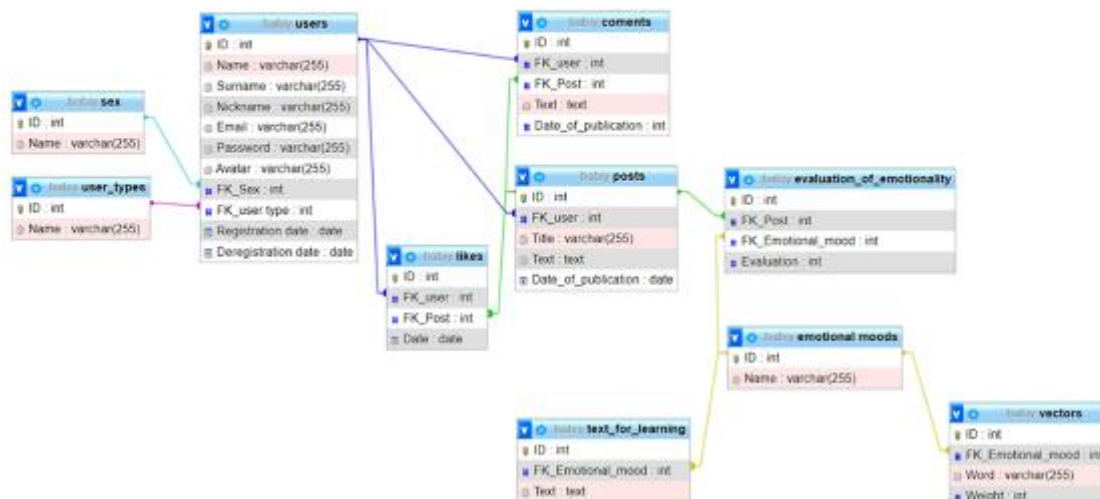
Песимістичні
2065

ПЕСИМІСТИЧНІ

Песимістичні слова характеризуються негативними емоціями, зневірою, сумом, розчаруваннями та тривогою. Вони часто асоціюються з невдачею, труднощами та безнадією.

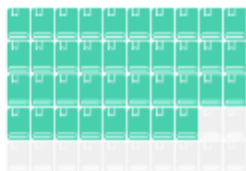


ДАТАЛОГІЧНА МОДЕЛЬ БД ВЕБРЕСУРСУ З ОБМІНУ ЛІТЕРАТУРНИМИ ТВОРАМИ



ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ РОБОТИ МЕТОДУ

Результати пошуку подібних творів за емоційним настроєм



38\50

Точність пошуку подібних творів для оптимістичного типу настрою становить 76%



29\50

Точність пошуку подібних творів для нейтрального типу настрою становить 58%



42\50

Точність пошуку подібних творів для песимістичного типу настрою становить 84%

ВИСНОВКИ

- Під час виконання кваліфікаційної роботи бакалавра було досліджено предметну область, яка стосується методів та підходів для визначення емоційного настрою в текстах та способів їх порівняння.
- Для досягнення поставленої мети, яка полягає в спрощенні роботи з літературними сервісами за рахунок автоматизованого підбору подібних за настроєм текстів було розроблено метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови.
- Для програмної реалізації розробленого методу було створено вебресурс з обміну літературними творами з використанням мови програмування PHP, фреймворку Laravel, СКБД MySQL та спеціалізованої бібліотеки PHP ML.
- Дослідження ефективності роботи методу показало, що точність підбору подібних творів оптимістичного настрою становить 76%, нейтрального – 58% та песимістичного – 84%.

Створений вебресурс з обміну літературними творами є навчальним проєктом, мета створення якого полягає в дослідженні ефективності роботи методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови. Оскільки, в ході дослідження отримано досить хороші результати, то розроблений метод може бути використаний для аналізу текстів на прояв емоційних настроїв та пошуку подібних текстів.

Tue Jun 18 11:44:31 EEST 2024, Петровський Сергій Степанович, Хмельницький національний університет, ХНУ

Anti-Plagiarism v-15.257

Максимальне співпадіння з одним документом 2.0%

Словники перевірки: en_US, ru_RU, ua_UA. Помилко в документах: 8%

ID: 131259 Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови Додано в БД: 2024-06-18 Автора: Артем БАБІЙ Керівники: Олександр ПАСІЧНИК Консультанти: Опоненти:	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	93255	1422	4672 (5%)	76 (5%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

Ім'я користувача:

Кафедра КН

ID перевірки:

1016370656

Дата перевірки:

18.06.2024 08:12:15 EEST

Тип перевірки:

Doc vs Internet + Library

Дата звіту:

18.06.2024 11:18:00 EEST

ID користувача:

100005671

Назва документа: КН-20-2 Бабій_ЗАПИСКА

Кількість сторінок: 88 Кількість слів: 14710 Кількість символів: 116968 Розмір файлу: 4.40 MB ID файлу: 1016177781

Виявлено модифікації тексту (можуть впливати на відсоток схожості)

8.63%**Схожість**

Найбільша схожість: 3.52% з джерелом з Бібліотеки (ID файлу: 1015021615)

5.91% Джерела з Інтернету

423

Сторінка 90

7.36% Джерела з Бібліотеки

151

Сторінка 94

0% Цитат

Вилучення цитат вимкнено

Вилучення списку бібліографічних посилань вимкнено

0%**Вилучень**

Немає вилучених джерел

Модифікації

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

Підозріле форматування

23

сторінки

**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 підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови

Автор: студент групи КН-20-2 Артем Бабій

Спеціальність: 122 – Комп'ютерні науки

Освітня програма: освітньо-професійна

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

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 Артема Бабія, не є плагіатом, оскільки: запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; до запозичень входять фрагменти програмного коду, що не мають авторства і містять поширені конструкції; серед запозичень знаходяться загальновідомі терміни, скорочення та визначення.

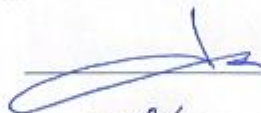
Обсяг запозичень, визначений системами виявлення збігів/ідентичності/схожості, складає:

- за системою Anti-Plagiarism: 2%;

- за системою Unichesk: 8,6 %,

що є допустимими запозиченнями.

Керівник роботи



Олександр ПАСІЧНИК

Гарант ОП



Олександр МАЗУРЕЦЬ

Завідувач кафедри КН



Олександр БАРМАК



ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ

Кафедра комп'ютерних наук



ВІДГУК НАУКОВОГО КЕРІВНИКА на кваліфікаційну роботу бакалавра

студента гр. КН-20-2 Бабія Артема Вікторовича

за темою Метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови

1. Актуальність теми

Тема підбору подібних за настроєм коротких художніх текстів є надзвичайно актуальною з кількох причин. По-перше, у сучасному світі, де обсяг інформації стрімко зростає, читачам важко знаходити твори, які відповідають їхнім емоційним уподобанням. Автоматизовані системи, що здатні аналізувати настрої текстів, допомагають швидко й ефективно знаходити літературні твори, які резонують з настроєм читача, що покращує його читачський досвід. По-друге, для письменників такі системи є джерелом натхнення та ідей, адже вони можуть легко знайти твори з подібною атмосферою, що може стимулювати творчий процес і допомогти у створенні нових текстів.

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

За стандартом, а саме описом предметної області, об'єктом роботи є процес підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови. Метою роботи є спрощення роботи з літературними сервісами за рахунок автоматизованого підбору подібних за настроєм текстів. При вирішенні поставленої задачі використано методи машинного навчання для роботи з текстовою інформацією. Тому результати виконання кваліфікаційної роботи бакалавра відповідають стандарту бакалавра спеціальності 122 – Комп'ютерні науки.

3. Професійні та особистісні якості бакалавра

Студент Бабій Артем Вікторович показав себе з кращого боку, демонструючи необхідні професійні та особистісні якості під час виконання кваліфікаційної роботи. Він виявив хороші знання у темі роботи, успішно застосовуючи їх для вирішення поставлених завдань. Студент був відповідальним і наполегливим, що дозволило йому ретельно виконати всі етапи дослідження.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Бабій Артем Вікторович виявив високий ступінь самостійності під час виконання кваліфікаційної роботи. Він самостійно проводив дослідження, аналізував інформацію та реалізовував завдання роботи.

5. Ступінь оволодіння методами дослідження

Студент продемонстрував високий ступінь оволодіння методами дослідження. Він успішно застосовував сучасні інструменти та методи для аналізу та обробки даних, показавши здатність самостійно проводити дослідження.

6. Повнота та якість розкриття теми роботи

Тема кваліфікаційної роботи повністю і якісно розкрита. Студент чітко сформулював мету та завдання роботи, успішно виконав їх, продемонструвавши всебічний підхід до вивчення теми. Якість викладення матеріалу і висновків на високому рівні, що забезпечило повне розкриття теми.

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

Матеріал кваліфікаційної роботи має логічність, послідовність і аргументованість у викладенні. Текст чітко структурований, зрозумілий і логічно обґрунтований на основі наукових джерел. Також має високий рівень літературної грамотності.

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

Метод підбору подібних за настроєм коротких художніх текстів з використанням засобів обробки природної мови має значний потенціал для практичного застосування. Наведений в роботі підхід може бути корисним для розробки рекомендаційних систем в літературних сервісах, де важливо враховувати емоційний настрій індивідуальних користувачів.

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 «задовільно».

Керівник  к.т.н., доц. каф. КН Олександр ПАСІЧНИК



ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ

Кафедра комп'ютерних наук



РЕЦЕНЗІЯ

на кваліфікаційну роботу бакалавра

студента *гр. КН-20-2 Бабія Артема Вікторовича*

за темою: Метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови

1. Актуальність обраної теми

У сучасних умовах зростання обсягу інформації та кількості літературних творів існує попит на автоматизовані системи, що можуть знаходити подібні твори за емоційним настроєм, що може бути корисним як для читачів, які шукають книги, схожі на їхні улюблені, так і для письменників, що шукають натхнення та ідеї для своїх творів. Розробка методу підбору коротких художніх текстів з подібним настроєм за допомогою засобів обробки природної мови покращить користувацький досвід під час пошуку схожих творів та отримання релевантних рекомендацій.

2. Повнота розкриття мети та завдань роботи

В роботі автор повністю досягнув поставленої мети, виконавши всі завдання, передбачені в роботі. Було проведено всебічне дослідження, зокрема, детально проаналізовано предметну область, описано кроки запропонованого методу та розроблено інформаційну систему для оцінки ефективності методу підбору подібних за настроєм коротких художніх текстів.

3. Зміст кожного розділу роботи

В роботі наведено актуальну інформацію, адже автор зсилається на останні сучасні джерела. У розділі 1 подано характеристику предметної області семантичного аналізу текстів за настроєм. У розділі 2 створено вебресурс з підбору подібних за настроєм художніх текстів та наведено кроки методу підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови. У розділі 3 розкрито особливості програмної реалізації вебресурсу з підбору подібних за настроєм художніх текстів та досліджено ефективність методу.

4. Оцінка розробленої інформаційної системи, її практична цінність

Розроблений метод підбору подібних за настроєм коротких художніх текстів засобами обробки природної мови допоможе покращити досвід користувача під час пошуку подібних творів та релевантних рекомендацій, а тому має практичну цінність.

5. Якість оформлення кваліфікаційної роботи бакалавра

В кваліфікаційній роботі бакалавра автор дотримався чіткого й логічного викладу матеріалу, забезпечив правильне оформлення всіх розділів, таблиць, графіків і рисунків. Також оформив роботу грамотно, коректно цитуючи джерела. Якісне оформлення свідчить про уважність і відповідальність автора.

6. Недоліки кваліфікаційної роботи бакалавра

Невірно оформлено скорочення в тексті записки. Не розглянуто наукові статті закордонних авторів по темі кваліфікаційної роботи бакалавра. Проте це не впливає на отриманий результат роботи.

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 «задовільно».

Рецензент

Валерій Мартинюк

