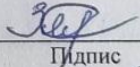
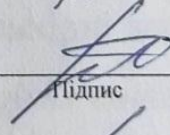
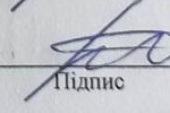


КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

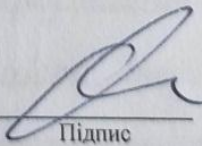
на тему Метод визначення рівня соціальної прийнятності текстового
контенту дописів за їх семантичним аналізом

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

Виконав: студентка 3 курсу, група КНс-20-1  Я.С. Загуровська
Курс, група виконавця Підпис Ініціали, прізвище
Керівник: к.т.н., доцент кафедри КН  Р.О. Багрій
Науковий ступінь, посада Підпис Ініціали, прізвище
Нормоконтроль: к.т.н., доцент кафедри КН  Р.О. Багрій
Науковий ступінь, посада Підпис Ініціали, прізвище

До захисту допускаю:

Зав. кафедри КН, д.т.н., професор

 О.В. Бармак
Підпис Ініціали, прізвище

05 06 2023 р.

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет інформаційних технологій

Кафедра комп'ютерних наук

Освітній ступінь бакалавр

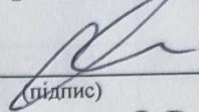
Галузь знань 12 – Інформаційні технології

Спеціальність 122 – Комп'ютерні науки

Освітня програма освітньо-професійна програма підготовки бакалавра

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних наук



(підпис)

д.т.н., професор О.В. Барма

« 06 » 03 2023 року

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА

1. Тема кваліфікаційної роботи бакалавра: «Метод визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом»

2. Завдання видано студентці Загуровській Яні Сергіївні

(прізвище, ім'я, по батькові)

3. Керівник роботи доцент кафедри КН Багрій Руслан Олександрович

(посада, прізвище, ім'я, по батькові)

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

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

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

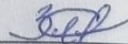
Виконати аналіз предметної області оцінювання рівня соціальної прийнятності текстового контенту дописів, розробити метод визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом. Обрати архітектуру нейронної мережі для визначення емоційного окрасу текстових дописів. Спроекувати структуру інформаційної системи на базі методу визначення рівня соціальної прийнятності текстового контенту дописів та виконати її програмну реалізацію. Провести тестування створеного програмного засобу на базі розробленого методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом.

7. Календарний план виконання кваліфікаційної роботи бакалавра:

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

Виконавець: студентка 3 курсу, група КНс-20-1

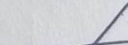
Курс, група виконавця


Підпис

Я.С. Загуровська
Ініціали, прізвище

Керівник:

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


Підпис

Р.О. Багрій
Ініціали, прізвище

Анотація

Тема кваліфікаційної роботи бакалавра: Метод визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом

Виконавець кваліфікаційної роботи бакалавра: студентка групи КНс-20-1 Загуровська Яна Сергіївна

Керівник кваліфікаційної роботи бакалавра: к.т.н., доцент кафедри КН Багрій Руслан Олександрович.

Кваліфікаційна робота бакалавра містить:

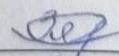
Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
62	33	4	26	5

Метою кваліфікаційної роботи бакалавра є розробка методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом та відповідного програмного забезпечення.

Результатом виконання кваліфікаційної роботи бакалавра є створений програмний засіб для автоматизованого визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом, що може використовуватися у задачах виявлення аб'юзивного контенту. При реалізації методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом для роботи з масивами даних було застосовано бібліотеку NumPy.

Ключові слова: інтернет-контент, RNN, рівень прийнятності допису, семантичний аналіз

Виконавець: студентка 3 курсу, група КНс-20-1
Курс, група виконавця


Відпис

Я.С. Загуровська
Ініціали, прізвище

Зміст

Перелік скорочень	4
Вступ.....	5
Розділ 1 Характеристика предметної області та постановка задачі	7
1.1 Проблема автоматизованого визначення соціальної неприйнятності текстового контенту.....	7
1.2 Підходи до оцінки соціальної прийнятності текстового контенту.....	11
1.3 Методи штучного інтелекту для визначення рівня соціальної прийнятності текстового контенту.....	12
1.4 Аналіз існуючих програмних реалізацій та публікацій в області визначення рівня соціальної прийнятності текстового контенту.....	15
1.5 Постановка задачі.....	18
Висновок до першого розділу.....	19
Розділ 2 Метод визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом.....	20
2.1 Схема та етапи методу визначення рівня соціальної прийнятності текстового контенту дописів.....	20
2.2 Структура інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів.....	23
2.3 Формування набору даних для визначення рівня соціальної прийнятності текстового контенту дописів.....	27
2.3.1 Набір даних для числової оцінки рівня соціальної прийнятності текстового контенту дописів	27
2.3.2 Набір навчальних даних нейромережі для оцінки тональності допису .	29
2.4 Застосування спеціалізованих програмних розширень для реалізації методу.....	29
Висновок до другого розділу	32
Розділ 3 Програмна реалізація методу визначення рівня соціальної прийнятності текстового контенту дописів.....	34

3.1 Архітектура інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів.....	34
3.2 Програмна реалізація складових інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів.....	36
3.3 Тестування інформаційної системи	42
3.4 Інструкція користувача для програмної реалізації методу визначення рівня соціальної прийнятності текстового контенту дописів.....	48
3.5 Дослідження ефективності впливу параметрів навчання на точність класифікації	56
Висновок до третього розділу.....	57
Висновки	59
Перелік посилань.....	61
Додатки	

Перелік скорочень

Скорочення, термін, позначення	Пояснення
КРБ	Кваліфікаційна робота бакалавра
КН	Комп'ютерні науки
ШІ	Штучний інтелект
RNN	Рекурентні нейронні мережі
LSTM	Long Short-Term Memory
SVM	Support Vector Machine

Вступ

Кваліфікаційна робота бакалавра присвячена розробці методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом та відповідного програмного забезпечення, що буде використовувати розроблений метод.

Актуальність. Кількість користувачів соціальних мереж у світі постійно зростає, що свідчить про значний обсяг контенту, що був опублікований і публікується щодня в цих мережах. У зв'язку з цим зростає і важливість забезпечення якості цього контенту. Зокрема, важливим аспектом є соціальна прийнятність контенту, оскільки несанкціонована, образлива або неправдива інформація може викликати шкоду та негативні наслідки для людей, груп чи суспільства загалом.

Тому визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом може бути корисним інструментом для оцінки рівню соціальної прийнятності цифрового текстового контенту, який публікується в соціальних мережах, та для запобігання поширенню шкідливої чи образливої інформації. Також це може сприяти поліпшенню якості комунікації та підвищенню рівня соціальної взаємодії в цілому.

Об'єкт дослідження – цифровий текстовий контент дописів, який публікується в соціальних мережах.

Предмет дослідження – моделі, методи та засоби машинного навчання, які дозволяють визначити рівень соціальної прийнятності цифрового текстового контенту на основі його семантичного аналізу.

Мета кваліфікаційної роботи бакалавра полягає в розробці методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом та відповідного програмного забезпечення.

Завдання кваліфікаційної роботи бакалавра – провести дослідження предметної області для задачі визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом; розробити метод визначення рівня соціальної прийнятності текстового контенту дописів; обрати

архітектуру нейронної мережі для визначення емоційного окрасу текстового допису; розробити інформаційну систему на базі методу визначення рівня соціальної прийнятності текстового контенту дописів; провести тестування створеного програмного засобу.

Розділ 1 Характеристика предметної області та постановка задачі

1.1 Проблема автоматизованого визначення соціальної неприйнятності текстового контенту

Соціальні мережі як засіб інформаційних технологій для спілкування на сьогодні посідають чільне місце у житті людей. Найпопулярніші соціальні мережі в усьому світі станом на січень 2023 року за кількістю активних користувачів щомісяця проілюстровано на рисунку 1.1 [1].

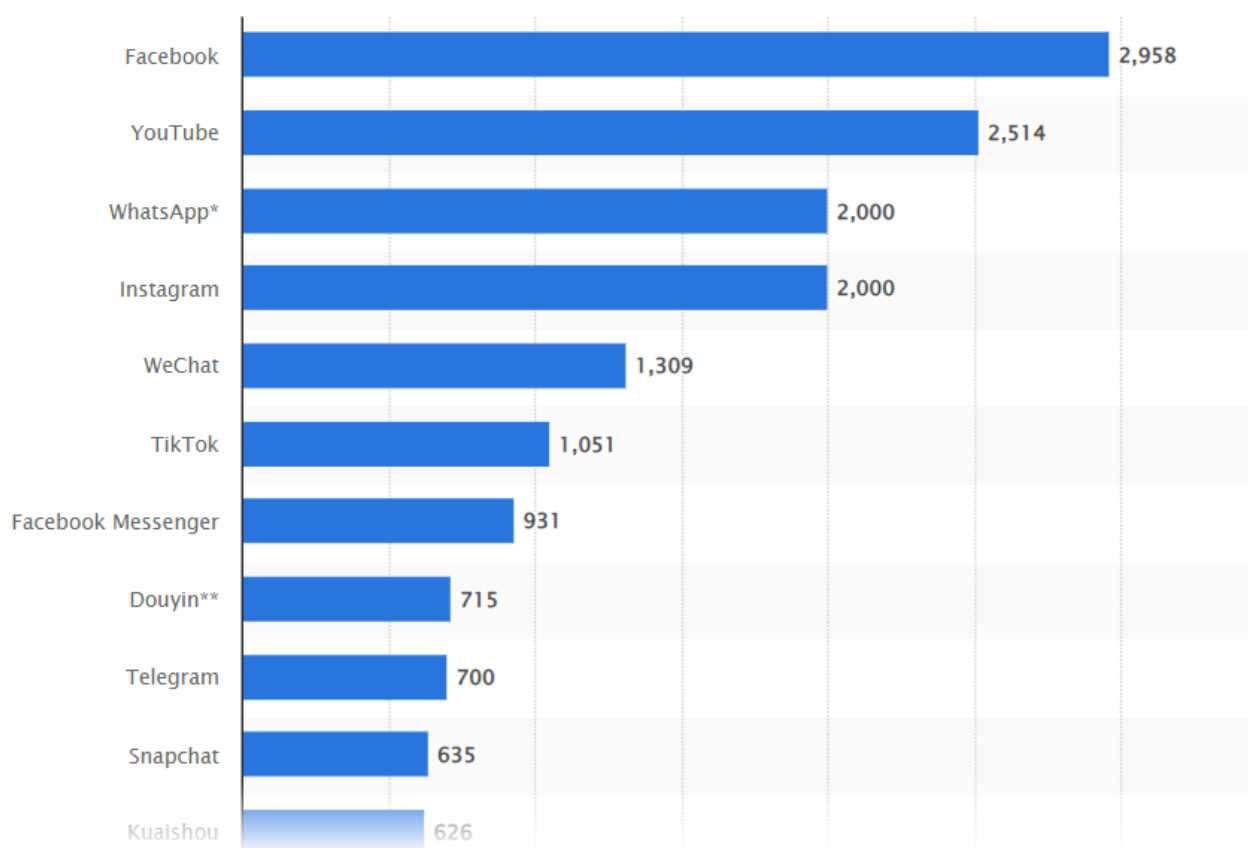


Рисунок 1.1 – Найпопулярніші соціальні мережі за кількістю активних користувачів щомісяця [1]

У правилах переважної більшості соціальних мереж чітко прописані вимоги до контенту, а також прийнятні норми поведінки користувачів. За порушення таких норм та правил соцмережі зазвичай блокують або контент, або ж самі акаунти. Стандарти спільнот Facebook, Instagram, Twitter, YouTube, та

інших подібних платформ забороняють публікувати контент терористичного характеру, розпал ворожнечі на різних рівнях, не допускати порушення авторських прав, скоординованої неавтентичної поведінки, проявів булінгу та розміщення порнографічних матеріалів. Також соціальні мережі ведуть боротьбу із фейковими сторінками, а деякі платформи видаляють навіть неактивні акаунти [2].

Контент, який вважається соціально прийнятним, може залежати від різних факторів, таких як культурні та моральні норми, вікові групи, географічне розташування та інші. Однак, загалом, соціально прийнятний контент є таким, який не порушує правила соціальної мережі та не викликає відразу чи обурення в масштабах значної частини аудиторії.

Соціально прийнятний контент може включати такі елементи як інформаційні повідомлення, гумор, музику, фотографії, відео, мистецтво, спорт та інші види змісту, які не порушують права людей, закони, здоровий глузд та норми етики. До прикладу, за правилами спільноти Facebook, заборонено публікувати контент, що містить загрози, які здатні призвести до смерті (та інших форм насильства високого ступеня тяжкості), або ж визнання факту вчинення в минулому насильства, спрямованого на адресу людей або місць, при цьому загроза визначається як будь-яка з наступних [3]:

- контент, який містить висловлення про намір вчинити насильство важкого характеру, включаючи випадки, коли об'єкт насильства не згадується конкретно словами, а вказується символічно, або ж містить зображення зброї або способу вчинення насильства;

- заклики до насильства високого ступеня тяжкості, включно з контентом, у якому об'єкт не позначений словесно, проте представлений символом чи містить візуальне зображення зброї або обставин вчинення насильства;

- затвердження на підтримку насильства важкого характеру;

- мотиваційні або умовні заяви про скоєння насильства важкого характеру;

– твердження, які містять визнання того, що було скоєне насильство важкого характеру, за винятком ситуацій, коли контент оприлюднюється з метою рятування життя або самозахисту, або коли подібні дії були вчинені працівниками правоохоронних органів, збройних сил або ж служб державної безпеки.

Також заборонено публікувати контент, який пропонує або визнає послуги, пов'язані з насильством високого ступеня тяжкості, такі як наймані вбивці, оренда зброї або насильницьке видалення статевих органів, або заохочує до використання таких послуг. Контент, що містить твердження про намір або схвалення вчинення насильства високого ступеня тяжкості, заклики до дії та мотиваційні чи умовні заяви, що стосуються захоплення або викрадення людей, а також пропагандування, підтримку чи схвалення полону або викрадення.

Опис полону або викрадення, і який не поширюється жертвою або її родиною, як прохання про допомогу, і не має за мету інформування, засудження або підвищення обізнаності. Загрози вчинення насильства високого ступеня тяжкості проти живих людей, і в яких використовуються зображення, створені або змінені за допомогою цифрових засобів, зображено зброю, способи скоєння насильства або розчленовані тіла.

Також заборонено публікувати інструкції з виготовлення та використання вибухових матеріалів, якщо відсутній контекст, що явно вказує на призначення контенту для мети, що не пов'язана з насильством (є частиною комерційних відеоігор або ж носить очевидний науковий чи освітній характер).

Соціальна мережа Facebook має політики щодо запобігання знущанням і приниженням користувачів. Ці політики забороняють будь-яку форму цькування, від простих погроз до розкриття особистих даних. Соцмережа бореться з жорстокими та цілеспрямованими нападами на публічних осіб, коли їх безпосередньо позначають у публікаціях і коментарях. Публічні особи визначаються як урядовці державного й національного рівня, політичні кандидати на такі посади, особи із понад мільйоном прихильників або читачів у соціальних мережах і осіб, новини про яких суттєво висвітлюються [4].

Facebook також має політики щодо захисту приватних осіб. Контент, покликаний принизити або присоромити, наприклад, заяви про дії сексуального характеру, видаляється. Крім того, Facebook забезпечує посилений захист користувачів у віці від 13 до 18 років, оскільки цькування та переслідування ще сильніше впливають на неповнолітніх.

Публікований контекст і наміри мають значення, тому користувачі можуть публікувати та поширювати контент, якщо чітко зрозуміло, що його поширено з метою засудження чи привернення уваги до проблем цькування та переслідування.

Щодо політик безпеки Google ADS, тут теж наведено свої правила для публікації прийняттого контенту. Зокрема, забороняється публікувати контент, який пропагує ненависть, дискримінацію або зневажає осіб або групи осіб на основі їх раси, етнічного походження, віросповідання, обмежених можливостей, віку, громадянства, статусу учасника бойових дій, сексуальної орієнтації, статі, гендерної належності або будь-яких інших характеристик, пов'язаних з систематичною дискримінацією або маргіналізацією. Також заборонено публікувати контент, який містить образи, цькування або ж утиски осіб або групи осіб, або ж контент, який містить погрози чи пропагує завдання фізичної або моральної шкоди собі чи іншій людині [5].

Ще однією категорією соціально-неприйняттого контенту є контент, спрямований на експлуатацію інших осіб. Сюди ж належить шантаж, вимагання неправомірної винагороди тощо.

Отже, з урахуванням великих обсягів інформації, яка щодня публікується у дописах на різних платформах, й потенційних загроз які вона несе, визначення соціальної неприйнятності текстового контенту є важливою задачею інформаційних технологій, адже належить до соціально-важливих задач.

1.2 Підходи до оцінки соціальної прийнятності текстового контенту

Зважаючи на важливість автоматизованого визначення соціальної неприйнятності текстового контенту, на сьогодні існують кілька підходів до визначення рівня соціальної прийнятності текстового контенту дописів в соціальних мережах. Одним із основних підходів є аналіз емоційного окрасу тексту.

Під емоційним аналізом розуміють процес перетворення тексту написаною природною мовою на бінарну форму (позитивний або негативний відгук). Оцінка емоційного відгуку може бути проведена як для цілого документу, так і для окремих речень та характеристик об'єкта, про який йдеться [6]. Такий підхід базується на використанні алгоритмів машинного навчання для визначення емоційного тону тексту. Він може включати оцінку наявності позитивних та негативних слів у тексті, визначення інтенсивності емоційного забарвлення тощо.

Іншим ключовим підходом є аналіз змісту [7]. Такий підхід зосереджений на вивченні змісту тексту та наявності в ньому образливих або неприйнятних висловлювань. Він може включати аналіз ключових слів, наявних в тексті, а також оцінку їх зв'язку з конкретною темою, допомагає розуміти зміст тексту та виявляти в ньому ключові елементи. Цей метод дозволяє визначити, чи відповідає текст певній темі, чи містить образливі або неприйнятні висловлювання, та оцінити настрій автора. Аналіз змісту може використовувати різні техніки, включаючи інструменти машинного навчання. До прикладу, можна використовувати методи статистичного аналізу, щоб виявити ключові слова та теми, що пов'язані з текстом. Також можуть бути застосовані методи глибинного навчання, щоб автоматично визначити настрій тексту та виявити образливі або ж неприйнятні висловлювання.

Аналіз змісту є важливим інструментом для багатьох галузей, включаючи маркетинг, журналістику та наукові дослідження. Цей метод допомагає

зрозуміти думки та настрої цільової аудиторії, а також виявити проблемні місця та можливість для поліпшення відповідних продуктів та послуг.

Наступним ключовим напрямком є аналіз контексту [8]. Аналіз контексту в задачі визначення рівня соціальної прийнятності текстового контенту включає в себе оцінку контекстуальних факторів, які можуть впливати на сприйняття тексту аудиторією. Це можуть бути фактори, такі як тема, мета, культурний та соціальний контекст, відтінок та стиль тексту. Одне і те ж висловлювання може мати різний рівень соціальної прийнятності в залежності від контексту, якщо це висловлювання зроблено в жартівливій формі серед друзів, то воно може бути вважатися прийнятним, але якщо воно з'явиться в публічному просторі і викличе обурення, то його можна вважати неприйнятним.

Аналіз контексту може включати в себе використання природньої мови та машинного навчання, а також залучення експертів в галузі мовленнєвої етики та соціології.

Крім того, визначення рівня соціальної прийнятності текстового контенту може використовувати комбінацію цих підходів або інших методів аналізу, які допомагають зрозуміти сприйняття та реакцію аудиторії на контент в соціальних мережах.

Отже, у межах роботи доцільно використати комбінований підхід на основі визначення тональності допису та аналізу змісту для контролю кількості слів образливого контенту.

1.3 Методи штучного інтелекту для визначення рівня соціальної прийнятності текстового контенту

Як було вище визначено, у роботі буде застосовано комбінований підхід на основі визначення тональності допису та підхід аналізу змісту для контролю кількості слів образливого контенту.

Визначення тональності текстового контенту є одним із способів визначення його соціальної прийнятності. Зазвичай використовуються методи

машинного навчання, які дозволяють класифікувати текст на позитивний, негативний або нейтральний за значенням слів, які він містить.

Одним із методів для визначення тональності текстового контенту є SVM (Support Vector Machine), що є алгоритмом машинного навчання.

SVM є алгоритмом навчання з учителем, який працює з векторами ознак та маркується з відомою класифікацією (негативний, позитивний). Алгоритм навчання SVM знаходить гіперплощину [9], яка найкраще розділяє дані на два класи, тобто позитивний та негативний (рисунок 1.2).

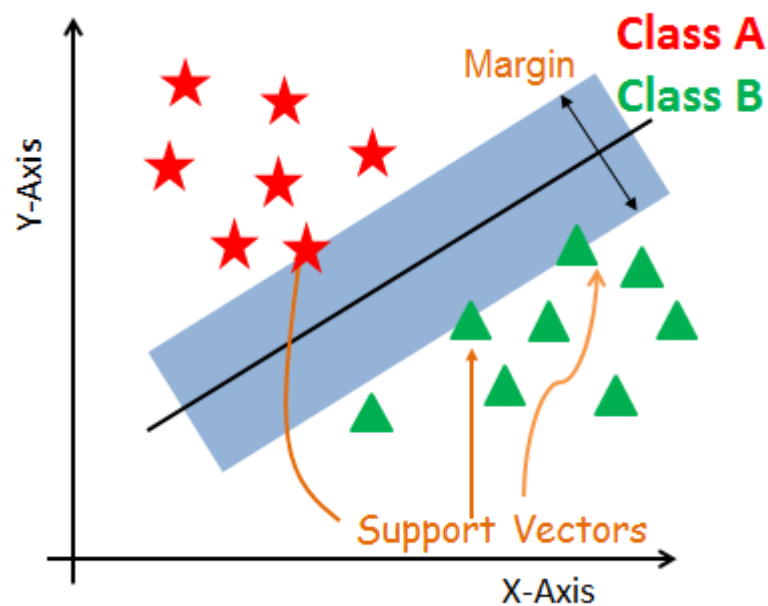


Рисунок 1.2 – Класифікатор SVM [10]

Для використання SVM для визначення тональності текстового контенту, спочатку текст обробляється з використанням алгоритмів обробки природньої мови для вилучення характеристик або ознак. Ці ознаки можуть включати такі параметри, як частота вживання слів, співвідношення позитивних та негативних слів та інші параметри.

Далі, SVM навчається на основі цих ознак, щоб визначити оптимальну гіперплощину для розділення текстів на позитивний та негативний класи. Після цього SVM можна використовувати для класифікації нових текстів.

Для досягнення більш високої точності визначення тональності тексту можна використовувати SVM разом з іншими алгоритмами та методами ШІ.

Ще одним досить популярним напрямом ШІ для вирішення задачі пошуку емоційної забарвленості є нейромережі. Рекурентні нейронні мережі (RNN) широко поширені в задачах обробки тексту, зокрема й для задач аналізу тональності. Особливості рекурентних нейронних мереж полягають у наявності зворотних зв'язків, зв'язок від віддаленого елемента до менш віддаленого. Архітектура нейромережі проілюстрована на рисунку 1.3. Це дозволяє запам'ятовувати та відтворювати послідовності реакцій на один стимул. Значення ваги нейромережі залежить як від поточних, так і від попередніх вхідних даних, завдяки чому вага кожного слова впливає на ваги інших шарів у реченні [11].

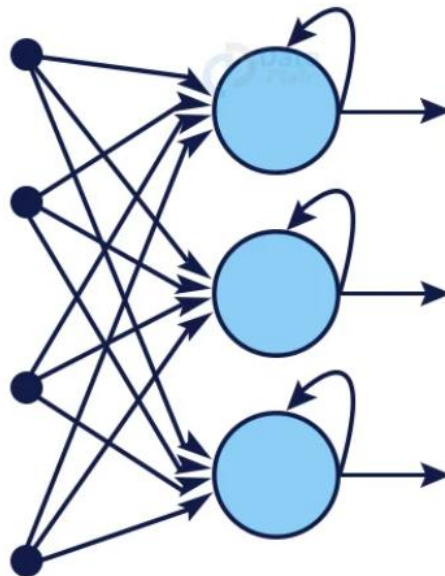


Рисунок 1.3 – Архітектура RNN [11]

Таким чином, у рамках виконання кваліфікаційної роботи бакалавра для виявлення емоційної забарвленості тексту доцільно використати рекурентну нейронну мережу.

1.4 Аналіз існуючих програмних реалізацій та публікацій в області визначення рівня соціальної прийнятності текстового контенту

Напрямок визначення рівня соціальної прийнятності текстового контенту є широко популярним серед науковців по всьому світу.

Нецензурна лексика, в тому числі різні лайливі висловлювання та мова ворожнечі не є новим явищем в мережі інтернет. Є багато причин, чому таке формулювання може вважатися небажаним, включаючи юридичні аспекти, проблеми, пов'язані з легкістю пошуку інформації та запобіганням негативному впливу шкідливого вмісту на людей (особливо підлітків) [12]. До швидкого зростання соціальних медіа та зменшення обсягу створеного користувачами вмісту намагання вручну виявляти та модерувати коментарі чи повідомлення на форумі, що містять така мова була керованою. Однак через поширеність соціальних мереж, модерувати всі коментарі вручну стає економічно недоцільним. Це призвело до активного розвитку методик автоматизованого виявлення нецензурної лексики [13].

Багато опублікованих робіт присвячено виявленню мови ворожнечі і образливої лексики в коментарях англійською мовою. Проте, за дослідженням авторів [12] не існує адекватних методів автоматичного виявлення нецензурної лексики і мови ворожнечі для української мови, оскільки розробка таких методів є складним завданням з кількох причин:

- Немає позначених баз даних або україномовних корпусів коментарів чи дописів у соціальних мережах з образливим вмістом.
- Через особливості словотворення української мови практично неможливо визначити кінцевий список образливих слів.
- Користувачі часто винаходять нові, явно нецензурні слова, з'єднуючи дві спільні основи інфіксом або додаючи префікси або суфікси.
- Використання суржику породжує велику різноманітність лайливих слів.

Соціальні мережі часто містять україномовні коментарі, що написані мовою суржика, яка поєднує українську та інші елементи. Люди можуть намагатися приховати використання нецензурної лексики, використовуючи евфемізми, замінюючи літери зірочками або іншими символами, чи інвертуючи порядок літер у слові.

У проведеному авторами [12] дослідженні увага була зосереджена на неконтрольованих імовірнісних методах, використовуючи початковий словник образливих термінів як вихідні дані.

Щодо програмних реалізацій в області визначення тональності відгуків – їх є також доволі багато. Однією з таких є Semantria – програма для аналізу тональності тексту, що використовує машинне навчання для класифікації тексту на позитивний, негативний або нейтральний [14]. Вона також надає аналіз емоцій, виявлення іменованих сутностей та класифікацію тексту за темою. Інтерфейс демонстрації роботи програми проілюстровано на рисунку 1.4.

The screenshot displays the Semantria web application interface. At the top, there is a navigation bar with the Lexalytics logo and menu items: Platform, Solutions, Tech, More, and NLP Demo. A 'Contact Us' button is located on the right. The main content area is divided into two sections. On the left, under '2. Select a text sample', a text sample is pasted into a box. The text discusses scientific research in Antarctica and mentions a budget deal. On the right, under 'JSON View', there is a 'Document' tab selected. Below it, a 'Summary' section provides a brief overview of the text. At the bottom right, a table lists sentiment phrases and their corresponding scores.

Sentiment Phrase	Sentiment Score
scientific research	0.50
understanding	0.41
productive	0.38
essential	-0.38
safety	-0.55
frigid	-0.63
shut down	-0.75

Рисунок 1.4 – Ілюстрація роботи програми Semantria [14]

Дана програма може аналізувати тексти різного формату, включаючи блоги, соціальні медіа, вебсайти та документи MS Office.

Дане програмне забезпечення має REST API, яке дозволяє взаємодіяти з програмою за допомогою HTTP-запитів. Для використання Semantria потрібно

створити обліковий запис, встановити API-ключ та налаштувати параметри аналізу, такі як мову, країну та тип джерела тексту. Після того, як текст буде надісланий до Semantria, програма розіб'є його на різні сегменти, визначить відносну важливість кожного сегмента та виконає аналіз настроїв, включаючи позитивний, негативний та нейтральний настрій.

Дана програма також пропонує додаткові функції, такі як виявлення ключових слів та визначення тематичного контексту тексту, може бути корисною для підприємств, що прагнуть аналізувати великі обсяги текстової інформації з різних джерел. Проте, використання програми Semantria є платним, тому ця програма може бути недосяжною для невеликих підприємств та фрілансерів.

Ще однією автоматизованою програмою для виявлення настрою текстів є програма SentiStrength, інтерфейс якої проілюстровано на рисунку 1.5.

[Home](#) - [Test](#) - [Non-English](#) - [Download](#) - [Java Version](#) - [Buy](#) - [About](#)



The text 'I love dogs quite a lot but cats I really hate.'
has positive strength **3** and negative strength **-1** with respect to keywords *dogs,frogs*

*Approximate classification rationale: I love[3] dogs quite a lot but [sentence: 3,-1]
[result: max + and - of any sentence][overall result = 1 as pos>-neg] (Detect
Sentiment Around Keywords)*

Рисунок 1.5 – Інтерфейс програми SentiStrength [15]

SentiStrength є програмою для аналізу тональності текстів, що була розроблена фахівцями з університету Кардіффа. Основна особливість SentiStrength полягає в тому, що вона працює з двома шкалами емоцій: позитивною та негативною. Програма може аналізувати текстові дані, які написані на англійській, німецькій, французькій, іспанській, португальській, італійській, нідерландській та російській мовах [16].

SentiStrength застосовує комбінацію підходів для визначення тональності тексту, включаючи в себе аналіз словника, машинне навчання та статистичних методів. Оцінка тональності виконується за шкалою від -5 до +5, де -5 відповідає найбільш негативній тональності, +5 - найбільш позитивній, а 0 - нейтральній.

Її можна використовувати для аналізу текстів з різних джерел, таких як соціальні мережі, вебсторінки, електронні листи, заявки на роботу, звіти тощо. Програма може аналізувати як окремі текстові документи, так і досить великі масиви даних. Може бути використана як у дослідженнях, так і в комерційних проєктах. Проте, дана програма не працює з українською мовою, що є істотним недоліком.

Зважаючи на сучасний стан програмного забезпечення у області автоматизованого виявлення емоційного окрасу для визначення рівня соціальної прийнятності текстового контенту дописів, розробка нових ефективних методів та інформаційних систем у даному напрямку є актуальною.

1.5 Постановка задачі

Метою кваліфікаційної роботи є розробка методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом та відповідного програмного забезпечення, що буде використовувати розроблений метод. Для досягнення поставленої мети в рамках роботи ставляться наступні завдання:

- провести дослідження предметної області для задачі визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом;
- розробити метод визначення рівня соціальної прийнятності текстового контенту дописів;
- обрати архітектуру нейронної мережі для визначення емоційного окрасу текстового допису;

- спроектувати структуру інформаційної системи на базі методу визначення рівня соціальної прийнятності текстового контенту дописів;
- програмно реалізувати інформаційну систему визначення рівня соціальної прийнятності текстового контенту дописів;
- провести тестування створеного програмного засобу.

Висновок до першого розділу

Було проведено дослідження предметної області для задачі визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом. У рамках дослідження предметної області було визначено, який контент є неприйнятним, та обґрунтовано, що автоматизація визначення соціальної неприйнятності текстового контенту є важливою задачею інформаційних технологій, адже належить до соціально-важливих задач.

Розглянуто існуючі підходи до визначення рівня соціальної прийнятності текстового контенту, та вирішено використати комбінований підхід на основі визначення тональності допису та аналізу змісту для контролю кількості слів образливого контенту.

Також було проведено аналіз існуючих програмних реалізацій та публікацій в області визначення рівня соціальної прийнятності текстового контенту. Виділено нерозв'язані задачі та встановлено мету роботи, якою є розробка методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом та відповідного програмного забезпечення, що буде використовувати розроблений метод.

Розділ 2 Метод визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом

2.1 Схема та етапи методу визначення рівня соціальної прийнятності текстового контенту дописів

Метод визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом призначений для перетворення вхідної інформації у вигляді допису для оцінки рівня соціальної прийнятності у вихідні дані у вигляді числової оцінки соціальної прийнятності допису. Узагальнена схема методу проілюстрована на рисунку 2.1.



Рисунок 2.1 – Узагальнена схема методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом

Вхідними даними методу є цифрова текстова інформація у вигляді допису. Дописи характеризуються деякими особливостями [17], основні з них:

– невелика довжина (дописи часто мають обмеження на кількість символів, що спричиняє стислості тексту та вимагає ясного та лаконічного висловлення думок);

– інформальність (дописи у соціальних мережах, зазвичай, мають неформальний характер, і можуть містити скорочення, нестандартну лексику, або суржик);

– використання мультимедіа (дописи можуть включати фото, відео, аудіо, гіфки, емодзі тощо);

– діалогічний характер (соціальні мережі мають розгалужену структуру, що сприяє взаємодії між користувачами та створенню діалогів).

Це все робить обробку дописів специфічною, відмінною від загальних методів для роботи з текстами.

Далі вхідні дані переходять до нейромережевої оцінки тональності допису. Нейромережева оцінка тональності відбувається на основі попередньо натренованої моделі. Схема проведення нейромережевої оцінки тональності допису зображена на рисунку 2.2.

Тональність допису буде визначатись з числового проміжку від 0 до 1, де 0 – повністю негативний допис, 1 – повністю позитивний.

Наступним кроком буде обрахунок статистичної оцінки наявності образливого контенту, яка буде визначатись при порівнянні вмісту повідомлення із словником образливої лексики. Статистична оцінка образливого контенту буде розраховуватись за формулою:

$$OffensiveC = \frac{O_w}{TotalCount},$$

де O_w – кількість образливих слів, що містяться у словнику, $TotalCount$ – загальна кількість слів у дописі.

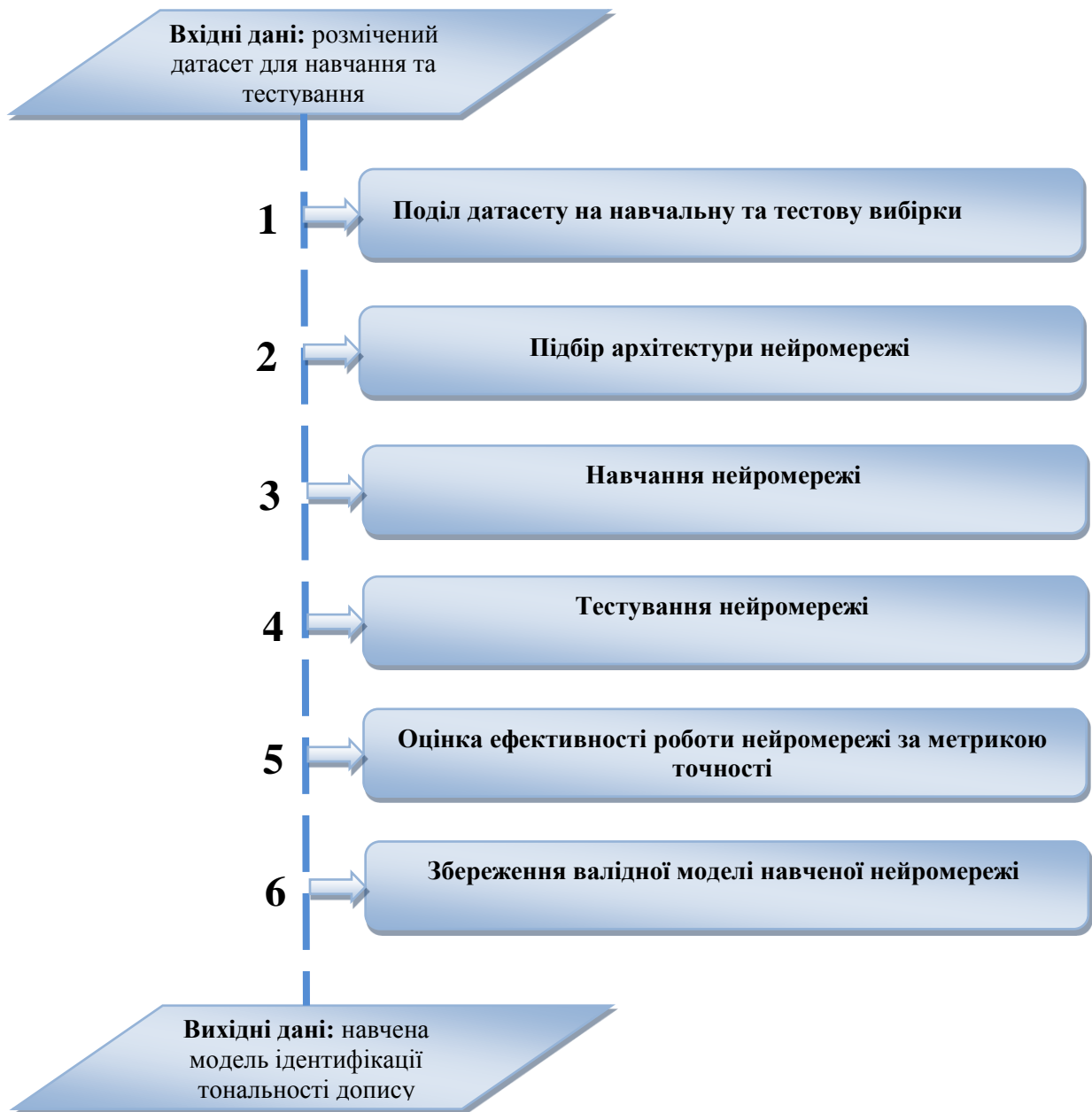


Рисунок 2.2 – Послідовність кроків при проведенні нейромережевої оцінки тональності допису

Числова оцінка рівня соціальної прийнятності текстового контенту дописів буде розраховуватись за формулою:

$$ResultPoint = \frac{OffensiveC + (1 - NNPoint)}{2},$$

де *OffensiveC* – статистична оцінка наявності образливого контенту, *NNPoint* – нейромережева оцінка тональності допису.

Рівні соціальної прийнятності текстового контенту дописів проілюстровано на рисунку 2.3.

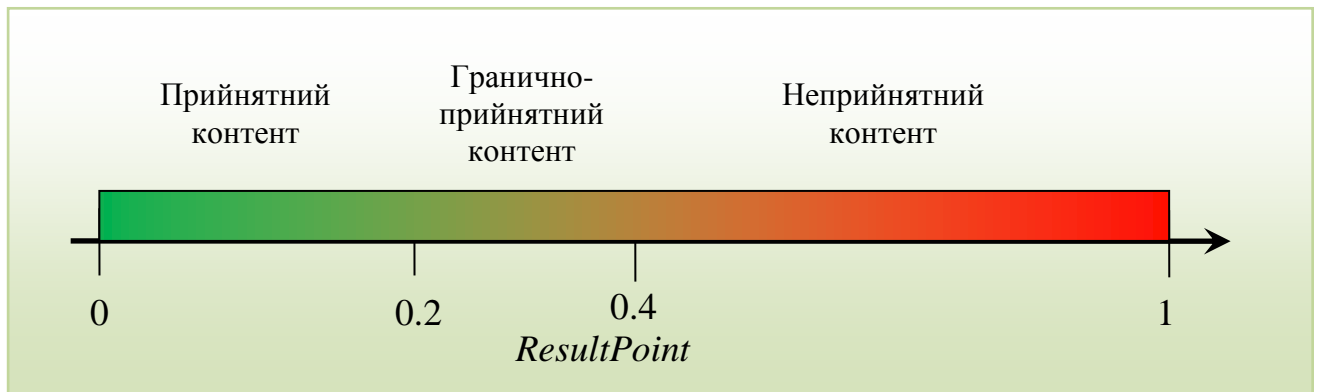


Рисунок 2.3 – Рівні прийнятності допису

Прийнятним вважається допис, *ResultPoint* якого не перевищує 0.2, гранично-прийнятним вважається контент, оцінка якого знаходиться в межах від 0.2 до 0.4. Цілком неприйнятним вважається контент, оцінка якого перевищує 0.4.

Відповідно, вихідними даними методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом є числова оцінка рівня соціальної прийнятності текстового контенту.

Отже, було розроблено метод визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом, що призначений для перетворення вхідної інформації у вигляді допису для оцінки рівня соціальної прийнятності у вихідні дані у вигляді інтегрованої числової оцінки прийнятності допису.

2.2 Структура інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів

На основі розробленого методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом, що призначається для перетворення вхідної інформації у вигляді допису для оцінки рівня соціальної

прийнятності у вихідні дані у вигляді інтегрованої числової оцінки прийнятності допису було визначено структуру відповідної інформаційної системи. У загальному, схема роботи інформаційної системи проілюстровано на рисунку 2.4.



Рисунок 2.4 – Загальна схема етапів роботи інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів

Розроблювана інформаційна система складається із кількох складових, взаємодія яких показана на рисунку 2.5.

Головною підсистемою є підсистема визначення рівня соціальної прийнятності текстового контенту дописів. Дана підсистема взаємодіє з підсистемою взаємодії зі словником образливих слів, підсистемою визначення рівня образливого вмісту та підсистемою нейромережевого визначення емоційного забарвлення допису.

В межах головної підсистеми відбувається визначення рівня соціальної прийнятності допису, що залежить від рівня тональності допису та від рівня визначення образливого вмісту.

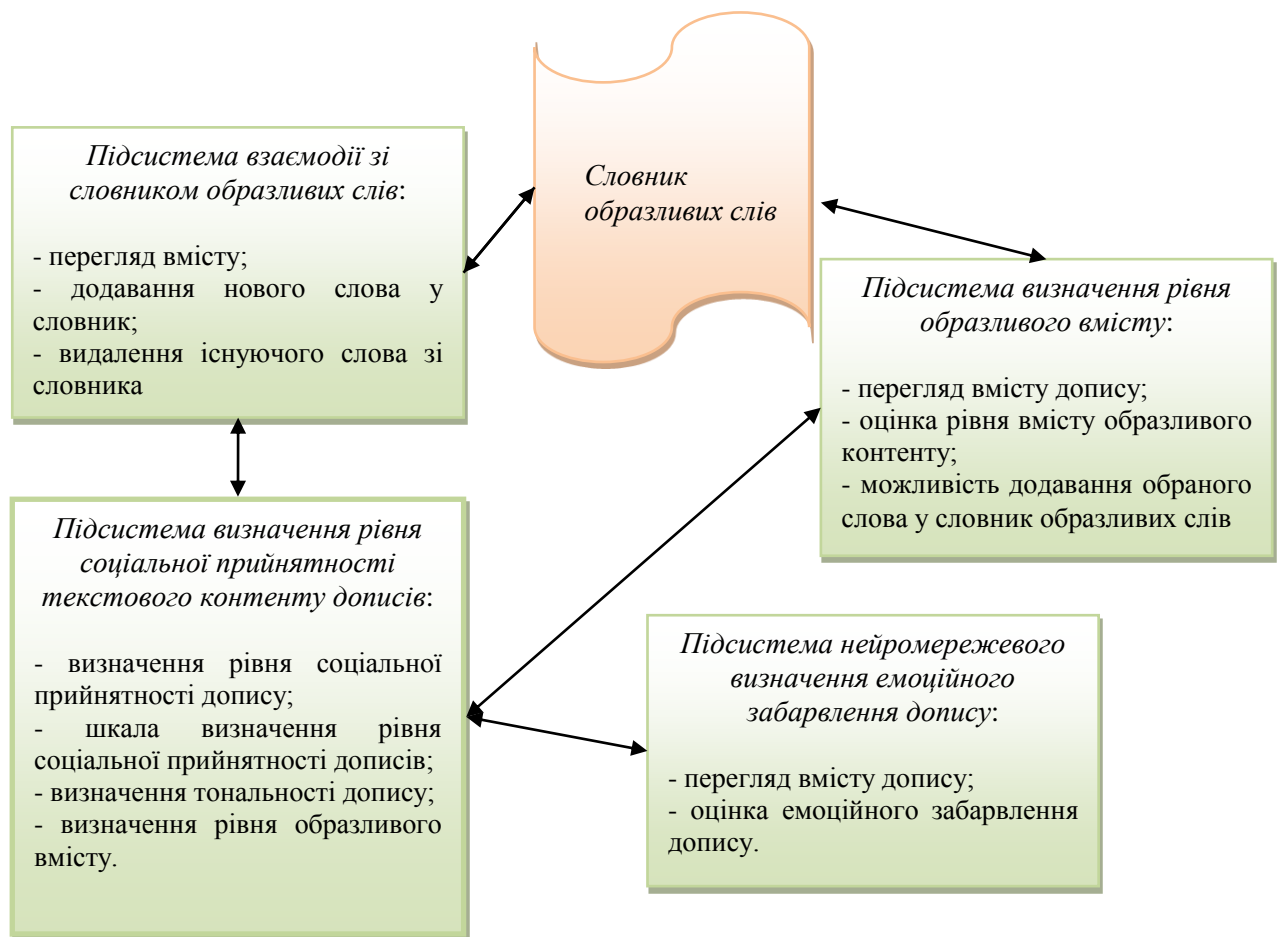


Рисунок 2.5 – Складові інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів

Підсистема нейромережевого визначення емоційного забарвлення допису призначена для оцінки рівня емоційного забарвлення користувацького допису. Оцінка здійснюється за попередньо навченою моделлю нейромережі RNN. Архітектура нейронної мережі проілюстрована на рисунку 2.6.

Для перетворення вхідного тексту на послідовність індексів слів, що є вхідними даними нейромережі, застосовується процес токенізації та індексації. Під час токенізації вхідний текст розбивається на окремі слова, що називаються токенами, прибираються зайві символи та пробіли [18]. Далі, кожному токеному присвоюється відповідний індекс. Після індексації, речення може бути перетворене на послідовність індексів, наприклад, {3, 18, 31}.



Рисунок 2.6 – Архітектура нейронної мережі RNN

Першим шаром після отримання вхідних даних є Embedding шар. Даний перетворює кожне слово у вектор з фіксованою кількістю значень. Це дозволяє нейромережі працювати зі словами, як з числовими векторами.

Наступним шаром буде шар LSTM, що є шаром довготривалої короткочасної пам'яті. LSTM є типом штучної рекурентної нейромережі, що здатна зберігати інформацію про попередні вхідні дані і використовувати її при обробці поточних даних [19].

Останнім шаром є Dense шар. Це повнозв'язний шар, що здійснює класифікацію вхідних даних на два класи – позитивний та негативний сентимент.

Вихідними даними нейромережевої класифікації є ймовірності належності допису до кожної з можливих емоційних категорій (позитиву чи негативу).

Наступною підсистемою інформаційної системи є підсистема визначення рівня образливого вмісту. Дана підсистема призначена для оцінки рівня вмісту образливого контенту та можливості додавання обраного слова у словник образливих слів.

Останньою підсистемою в рамках інформаційної системи визначення рівня соціальної прийнятності допису є підсистема взаємодії зі словником

образливих слів. Її призначенням є перегляд вмісту словника, можливість додавання нового слова у словник та можливість видалення існуючого слова зі словника.

Отже, таким чином представлена структура інформаційної системи на основі розробленого методу для визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом, який призначено для перетворення вхідних даних у вигляді допису для оцінки рівня соціальної прийнятності у вихідні дані у вигляді інтегрованої числової оцінки прийнятності цифрового текстового допису.

2.3 Формування набору даних для визначення рівня соціальної прийнятності текстового контенту дописів

Оскільки розроблювана інформаційна система має в собі нейромережеву складову, для навчання нейромережі необхідно використовувати великі обсяги даних. Тому є доцільним знаходження датасетів для навчання та тестування нейромережі, а також датасету пошуку образливих слів для числової оцінки рівня соціальної прийнятності текстового контенту.

Наприклад, для англійської мови існують відомі датасети, такі як Sentiment Analysis on Movie Reviews (SST-5), Sentiment140, IMDb Movie Review Dataset тощо [20]. Для української мови також є датасети, які можуть бути використані для визначення тональності тексту, проте їх на відміну від англомовних аналогів досить мало.

2.3.1 Набір даних для числової оцінки рівня соціальної прийнятності текстового контенту дописів

Для реалізації методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом було використано набір даних "AbusiveLanguageDataset". Це датасет, який містить україномовні

коментарі з різним рівнем образливого вмісту. Даний датасет складається з 5 тисяч коментарів, з яких більшість є негативними та містять нецензурну лексику. Коментарі були зібрані з популярних українських інтернет-ресурсів, таких як: Українська правда, Табло ID та ТСН [21]. Датасет має підтримку багатьох мов, серед яких українська. Ілюстрація можливих мов датасету на рисунку 2.7.

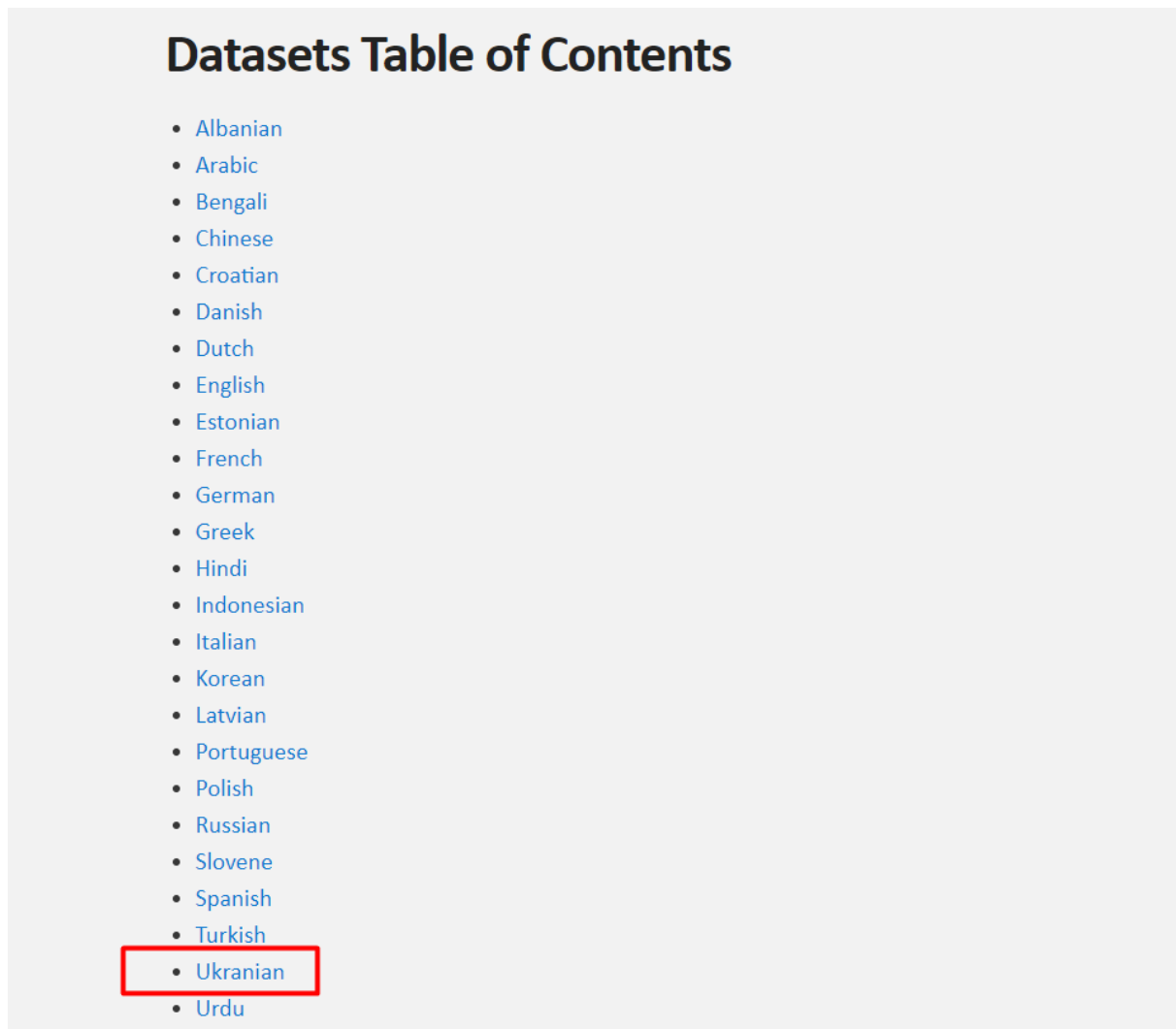


Рисунок 2.7 – Підтримка мов набору даних AbusiveLanguageDataset

Кожен коментар в обраному наборі даних було оцінено людиною, яка визначила, чи містить він образливий вміст, а також визначено рівень образливості (низький, середній або високий). Крім того, для кожного коментаря вказано ідентифікатор джерела, з якого було отримано коментар.

Датасет буде використаний у якості даних для числової оцінки рівня соціальної прийнятності текстового контенту дописів.

2.3.2 Набір навчальних даних нейромережі для оцінки тональності допису

Для навчання та валідації нейромережі теж є потреба використання великого обсягу текстових розмічених даних.

Ukrainian Twitter Corpus [22] – це корпус україномовних твітів, які були зібрані за допомогою Twitter API. Корпус містить більше 6 мільйонів твітів, які були зібрані з грудня 2015 року по травень 2017 року .

Кожен твіт корпусу має мітку позитивного, негативного або нейтрального настрою. Анотування було здійснене за допомогою алгоритмів машинного навчання та оцінювалося на рівні повідомлень. Загалом, пропонуваній корпус містить близько 3 мільйонів позитивних твітів, 1,7 мільйонів негативних та 1,6 мільйонів нейтральних. З них українською мовою 400 тисяч твітів.

Також корпус містить додаткові дані про кількість вподобань, ретвітів та реплік на кожний твіт. Дані є доступними для завантаження на GitHub. Оскільки дані у корпусі зібрані з соціальної мережі Twitter, вони можуть містити неформальну лексику, скорочення та інші особливості мовлення, які характерні для даного джерела. Однак враховуючи тему дослідження, цей корпус доречно використати для навчання RNN, що буде виконувати завдання визначення семантичного окрасу повідомлення.

2.4 Застосування спеціалізованих програмних розширень для реалізації методу

Для реалізації методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом існує ряд доступних бібліотек, що значно спрощують розробку нейромережевих застосувань і роботу з великими обсягами даних.

Однією із таких бібліотек є TensorFlow, що є відкритим програмним забезпеченням для розробки та тренування машинного навчання та глибокого навчання (deep learning). Дана бібліотека розроблена компанією Google та випущена ще у 2015 році [23].

Бібліотека дозволяє розробникам створювати, тренувати та використовувати нейромережі з значною кількістю шарів та доволі складною архітектурою. Бібліотека підтримує широкий спектр моделей машинного навчання, включаючи класифікацію, регресію, кластеризацію, а також і обробку мови.

Однією з ключових особливостей TensorFlow є використання графів обчислень. Графи описують послідовність операцій, які потрібно виконати для отримання результату. Графічна структура дозволяє TensorFlow оптимізувати обчислення та ефективно використовувати ресурси обчислювальної системи.

Дана бібліотека має різні рівні інтерфейсу програмування, включаючи низькорівневий API для ефективної роботи з тензорами та графами обчислень, вищорівневий API для створення та тренування моделей, а також високорівневий API, які спрощують розробку машинного навчання та глибокого навчання.

TensorFlow є однією з найбільш популярних бібліотек машинного навчання, з великою кількістю користувачів та розробників у всьому світі [24].

У межах роботи планується використовувати бібліотеку TensorFlow для визначення моделі нейромережі, компіляції та навчання моделі.

Також TensorFlow має ряд надбудов, серед яких Keras, що є високорівневою надбудовою над TensorFlow, що дозволяє легко і швидко створювати, тренувати та оцінювати різні види нейронних мереж. Вона була розроблена з метою спростити розробку глибоких нейронних мереж, зменшивши кількість коду, необхідного для створення моделей. Стартова сторінка офіційного сайту зображена на рисунку 2.8.

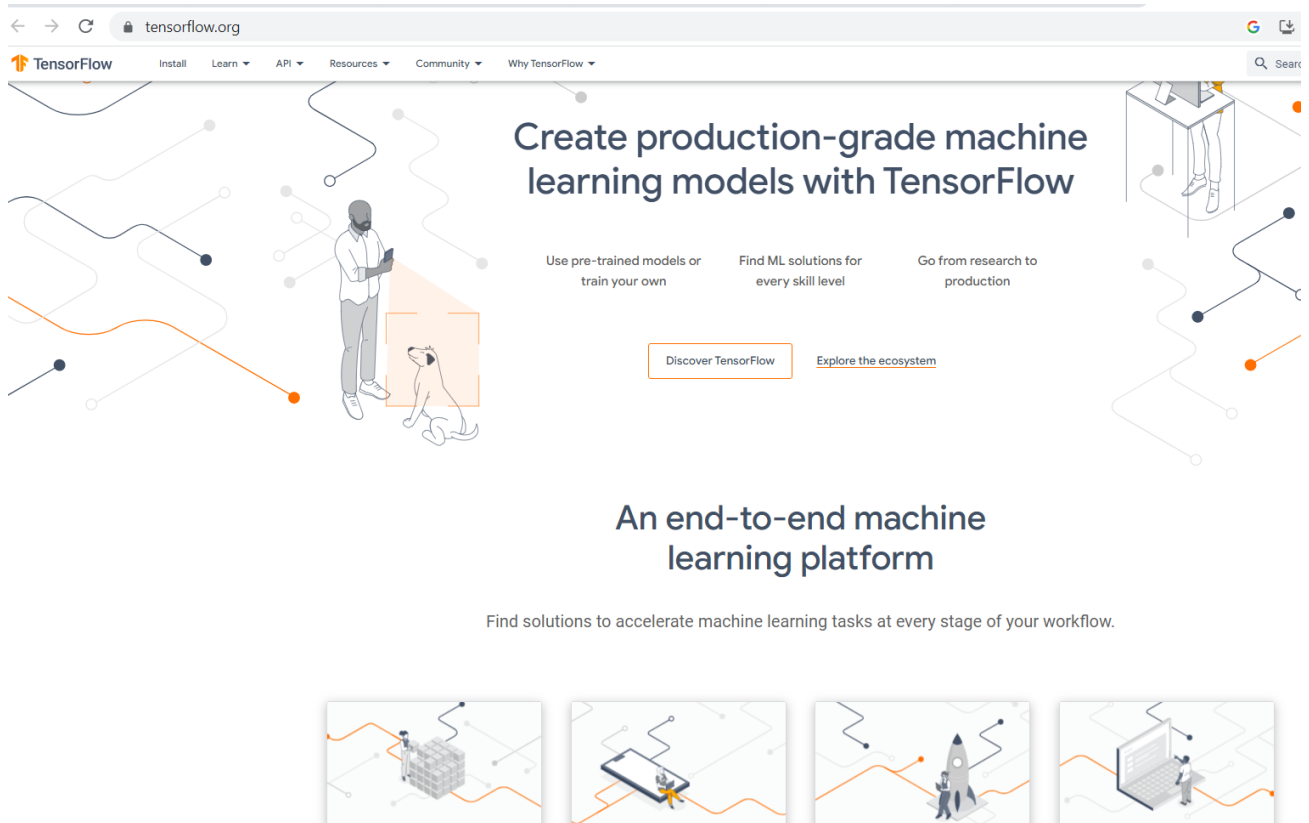


Рисунок 2.8 – Видгляд інформаційної вебсторінки офіційного сайту TensorFlow [23]

Має простий та інтуїтивно зрозумілий інтерфейс, дозволяючи розробникам швидко та легко створювати складні мережі. Вона містить в собі різні типи шарів, які можна використовувати в моделях: від звичайних пов'язаних шарів до більш складних шарів, таких як LSTM або Convolutional [25].

Keras буде використано для визначення послідовності шарів нейромережі.

NumPy (Numerical Python) є бібліотекою мови Python, що дозволяє зручно виконувати наукові обчислення, особливо роботу з масивами і матрицями даних. Вона містить структури даних для ефективного представлення та маніпулювання числовими даними, а також функції для розв'язання різноманітних завдань, пов'язаних з науковими обчисленнями [26].

До основних функцій NumPy належать:

– Багатовимірні масиви, які можуть містити дані різних типів.

- Функції для математичних операцій з масивами, такі як обчислення суми, максимуму, мінімуму, середнього значення тощо.
- Векторизовані обчислення, які дозволяють ефективно виконувати операції над масивами без потреби у використанні циклів.
- Оператори та функції для індексації та зрізання масивів.
- Функції для зчитування та запису даних з файлів у форматі NumPy.

NumPy є однією з основних бібліотек для наукових обчислень у Python, і вона використовується у багатьох інших бібліотеках, таких як Pandas, Matplotlib, SciPy, TensorFlow.

При реалізації методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом бібліотека NumPy буде застосовано для роботи з масивами даних.

Висновок до другого розділу

У межах другого розділу було розроблено метод визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом, який призначений для перетворення вхідних даних у формі допису для оцінки рівня соціальної прийнятності у вихідну інформацію у вигляді числової оцінки етичної прийнятності допису.

На базі запропонованого методу було сформовано архітектуру інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом та визначено функціональне призначення складових.

Знайдено відповідні набори даних для навчання нейромережі та для застосування ідентифікації слів образливого контенту. У якості даних для числової оцінки рівня соціальної прийнятності текстового контенту дописів використано набір даних «AbusiveLanguageDataset», а для навчання та валідації нейромережі було використано «Ukrainian Twitter Corpus».

Описано існуючі бібліотеки, які дозволяють пришвидшити програмну реалізацію для розробленого методу та відповідної інформаційної системи, для роботи з масивами даних при реалізації методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом буде застосовано бібліотеку NumPy.

Розділ 3 Програмна реалізація методу визначення рівня соціальної прийнятності текстового контенту дописів

3.1 Архітектура інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів

На основі розробленого методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом та розробленої структури інформаційної системи розроблено її відповідну архітектуру, зображену на рисунку 3.1.

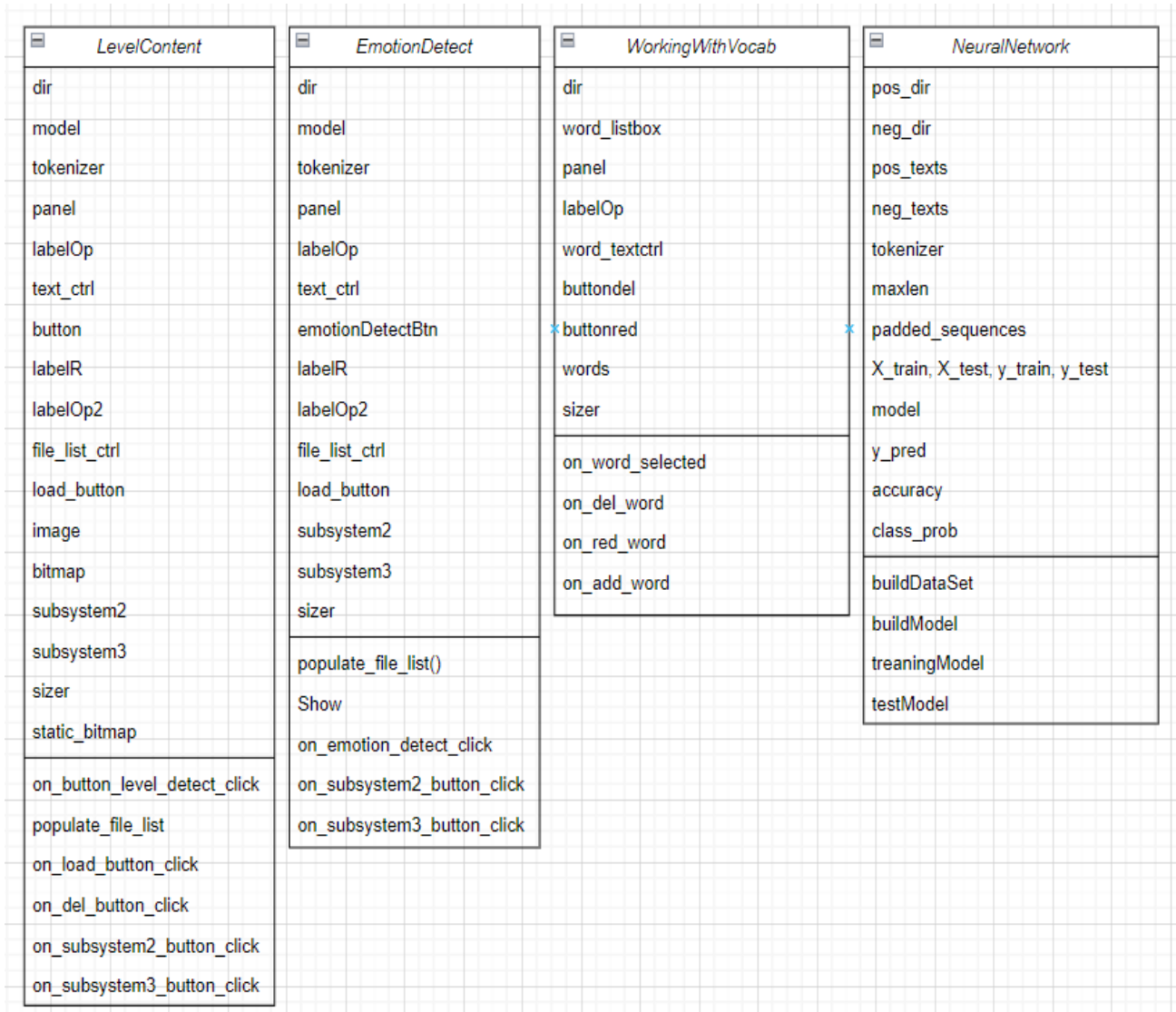


Рисунок 3.1 – Діаграма класів інформаційної системи

Клас *NeuralNetwork* призначений для формування та навчання моделі рекурентної нейромережі. У цьому класі формуються вибірки для навчання та тестування на основі наявних завантажених дописів, відбувається побудова моделі (метод *buildModel*), навчання моделі (метод *trainingModel*) та тестування ефективності застосування моделі (метод *testModel*). Даний клас у результаті формує токенайзер та зберігає його та навчену модель, що буде вхідними даними для класу *LevelContent* та *EmotionDetect*.

Клас *LevelContent* – це головний клас програми, у якому відбувається визначення рівня соціальної прийнятності текстового контенту дописів за нейромережевою оцінкою та оцінкою наявного образливого контенту. Завантажує збережену навчену модель нейромережі та токенайзер, що були збережені як результат роботи класу *NeuralNetwork*. Має інтерфейс користувача, де реалізовано можливість як визначити рівень соціальної прийнятності текстового контенту дописів з датасету, так і завантажити інший текст для аналізу.

Клас *EmotionDetect* призначений для нейромережевого виявлення емоції в тексті. Містить інтерфейс користувача, що дозволяє виявити емоційний окрас не тільки повідомлення введеного користувачем, а і повідомлення з датасету. Метод *populate_file_list* відповідає за відображення назв файлів з текстами з датасету на екран користувача. Для відображення в текстовому полі для аналізу викликається метод *Show*.

Клас *WorkingWithVocab* призначений для взаємодії користувача з аб'юзивним словником. Тут є можливість додавати нові слова у словник, переглядати існуючі і видаляти обрані слова.

Отже, таким чином розроблено архітектуру інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів, що використовує розроблений метод.

3.2 Програмна реалізація складових інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів

На основі створеної архітектури інформаційної системи на базі методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом, що призначається для перетворення вхідної інформації у вигляді допису для оцінки рівня соціальної прийнятності у вихідні дані у вигляді інтегрованої числової оцінки прийнятності допису було створено відповідну програмну реалізацію складових частин.

Для функціонування інформаційної системи, спершу потрібно навчити нейромережу на обраних вище датасетах, та зберегти навчену модель для подальшого використання.

Спершу відбувається налаштування параметрів токенизатора та здійснюється обробка тексту.

Для цього використовується клас *Tokenizer* з бібліотеки *keras.preprocessing.text* для токенизації тексту. Основна мета якого підготувати текстові дані для подальшої обробки в нейромережі для інтегрованої числової оцінки прийнятності допису.

Також виконується ініціалізація об'єкту *Tokenizer* з параметром *num_words*, що вказує максимальну кількість слів, які повинні бути враховані при токенизації. У рамках кваліфікаційної роботи бакалавра мова йде про обробку невеликих текстових даних, тому даний параметр встановлено на 1000, що означає, що будуть використані 1000 найпоширеніших слів з тексту. Навчання токенизатора на вхідних текстах, здійснюється викликом стандартного методу *fit_on_texts(texts)*. Це дає токенизатору інформацію про слова, що зустрічаються в текстах і їхні частоти. Далі послідовності генеруються за допомогою стандартного методу *texts_to_sequences(texts)*, що перетворює текст у послідовність цілих чисел (індексів слів), де кожне число відповідає індексу слова у словнику токенизатора.

Після налаштування токенизатора задається максимальна довжина послідовності, яка для задачі обчислення інтегрованої числової оцінки прийнятності допису сягатиме 100. Далі за допомогою функції *pad_sequences* з бібліотеки *keras.preprocessing.sequence* відбувається заповнення послідовностей цілих чисел нулями або для обрізки їх до фіксованої довжини.

Для побудови моделі нейромережі використано клас *Sequential* з бібліотеки *keras.models* з використанням шарів *Embedding*, *LSTM* та *Dense*. Це потрібно для побудови і компіляції моделі нейронної мережі для бінарної класифікації для виявлення нейприйняттого контенту за допомогою рекурентної нейронної мережі з *LSTM* шаром.

Перший доданий шар є шаром *Embedding*, що використовується для перетворення вхідних цілих чисел у вектори фіксованої довжини. Параметр *num_words* вказує на кількість унікальних слів у корпусі тексту, а число 32 є розміром векторного простору.

Другим шаром є шар *LSTM* (рекурентна нейронна мережа), що використовується для обробки послідовностей даних. Параметр зі значенням 32 вказує на кількість нейронів у шарі. Параметри *dropout* і *recurrent_dropout* визначають ймовірність випадкового відключення нейронів для регуляризації моделі та запобігання перенавчанню.

Третім доданим шаром є шар *Dense*, що є повнозв'язним шаром з одним нейроном і функцією активації *sigmoid*, оскільки модель вирішує задачу бінарної класифікації, де вихідний нейрон відповідає ймовірності належності або не належності до одного класу.

Наступним етапом після побудови моделі відбувається її навчання. Навчання моделі нейромережі відбувається за допомогою функції *fit* та отримує передбачення для тестової вибірки з метою виявлення нейприйняттого контенту.

Ілюстрація старту навчання та проходження епох зображена на рисунку 3.2. Дана модель запущена на 15 епох.

```

=====
Total params: 328,353
Trainable params: 328,353
Non-trainable params: 0
-----
None
Epoch 1/15
108/132 [=====>.....] - ETA: 1s - loss: 0.5376 - accuracy: 0.7752

```

Рисунок 3.2 – Запуск старту навчання

Далі здійснюється оцінка якості навченої моделі за метрикою точності. Обчислюється точність моделі на тестовій вибірці, використовуючи функцію *accuracy_score* стандартної бібліотеки *sklearn.metrics* і виводить результат.

Після завершення процесу навчання нейромережі буде виведено на екран користувача тривалість виконання кожної епохи та точність класифікації (рисунок 3.3).

```

Epoch 14/15
132/132 [=====] - 8s 63ms/step - loss: 0.0878 - accuracy: 0.9511 - val_loss: 0.5921 - val_accuracy: 0.8493
Epoch 15/15
132/132 [=====] - 8s 63ms/step - loss: 0.0865 - accuracy: 0.9504 - val_loss: 0.6027 - val_accuracy: 0.8474
33/33 [=====] - 1s 7ms/step
Accuracy: 0.8473933649289099

```

Рисунок 3.3 – Завершення процесу навчання

Для передбачення рівня позитивної чи негативної емоції використовується метод *predict* моделі, що належить до бібліотеки *keras* (зокрема, класу *Sequential* моделі з бібліотеки *keras.models*). Метод *predict* є методом моделі, який застосовує навчену модель до вхідних даних та генерує передбачення для нового тексту за допомогою навченої моделі і виведення ймовірності наявності негативного контенту.

Для використання нейромережі було створено підсистему визначення тональності допису з відповідним графічним інтерфейсом користувача (рисунок 3.4).

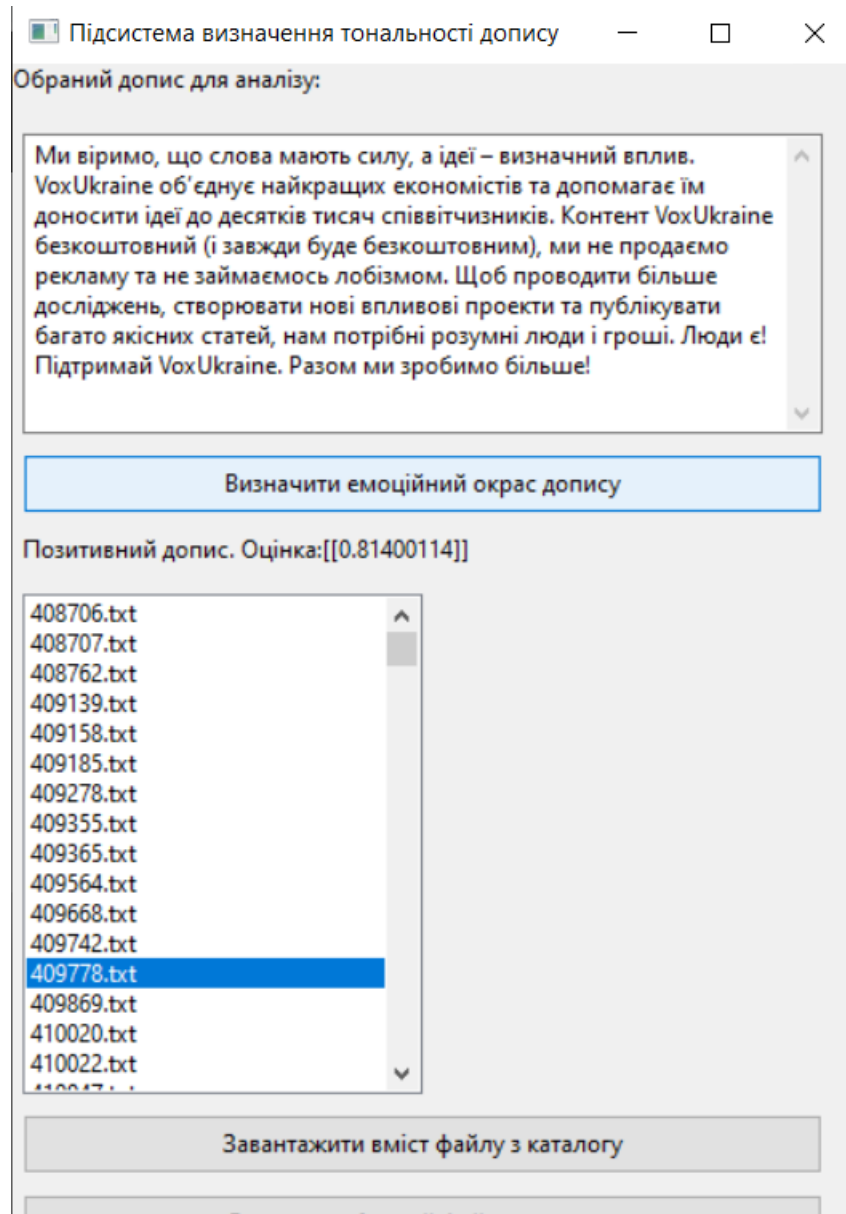


Рисунок 3.4 – Інтерфейс підсистеми визначення тональності допису

Для оцінки рівня прийнятності дописів підсистеми «Визначення рівня соціальної прийнятності допису» використано наступний алгоритм дій.

Спершу відбувається отримання значення з текстового поля (*self.text_ctrl*) за допомогою методу *GetValue()*. Значення текстового поля зберігається у змінній *text*.

Наступним кроком здійснюється токенизація тексту шляхом використання методу *texts_to_sequences* токенизатора (*tokenizer*) для перетворення тексту на послідовності токенів. Послідовності зберігаються у змінній *sequences*.

Після чого здійснюється заповнення послідовності до максимальної довжини (*maxlen*) за допомогою методу *pad_sequences*. Результат заповненої послідовності зберігається у змінній *padded_sequences*. Заповнена послідовність (*padded_sequences*) передається на передбачення моделі (*model.predict*), що повертає передбачення. Результат передбачення зберігається у змінній *prediction*.

Далі здійснюється зчитування вмісту з файлу "*bad_words.txt*" і збереження його у змінну *self.file_content* з метою подальшого розбиття вхідного тексту та вмісту файлу на окремі слова та обчислення кількості спільних слів між ними. Результат зберігається у змінну *levelOfNegativeCont*.

Далі відбувається обчислення значення *diferentialPoint* на основі оцінки прийнятності тексту та рівня негативного вмісту та відображення повідомлення з кількістю спільних слів з образливим вмістом.

Залежно від значення передбачення (*prediction*), встановлюється мітка для відображення результату визначення тональності тексту (*self.labelR*). Якщо передбачення більше 0.5, встановлюється мітка "*Позитивний допис*". У протилежному випадку, встановлюється мітка "*Негативний допис*". Додатково, у мітці відображається значення передбачення (*prediction*) та оцінка рівня образливого контенту (*levelOfNegativeCont*).

В залежності від значення *diferentialPoint*, встановлюється мітка (*SetLabel*) для відображення рівня прийнятності тексту (*self.label*). Якщо *diferentialPoint* менше 0.2, встановлюється мітка "*Допис є прийнятним*". Якщо *diferentialPoint* знаходиться в діапазоні від 0.2 до 0.4 включно, встановлюється мітка "*Допис є гранично прийнятним*". В іншому випадку, встановлюється мітка "*Допис є неприйнятним*".

Залежно від значення *diferentialPoint*, встановлюється мітка *self.label* з відповідним повідомленням про оцінку прийнятності тексту.

Дана функція обраховує рівень прийнятності допису в залежності від отриманої нейромережевої оцінки та від статистичної оцінки обрахунку кількості неприйнятних слів. Результат виконання методу у рамках підсистеми визначення рівня соціальної прийнятності допису наведено на рисунку 3.5

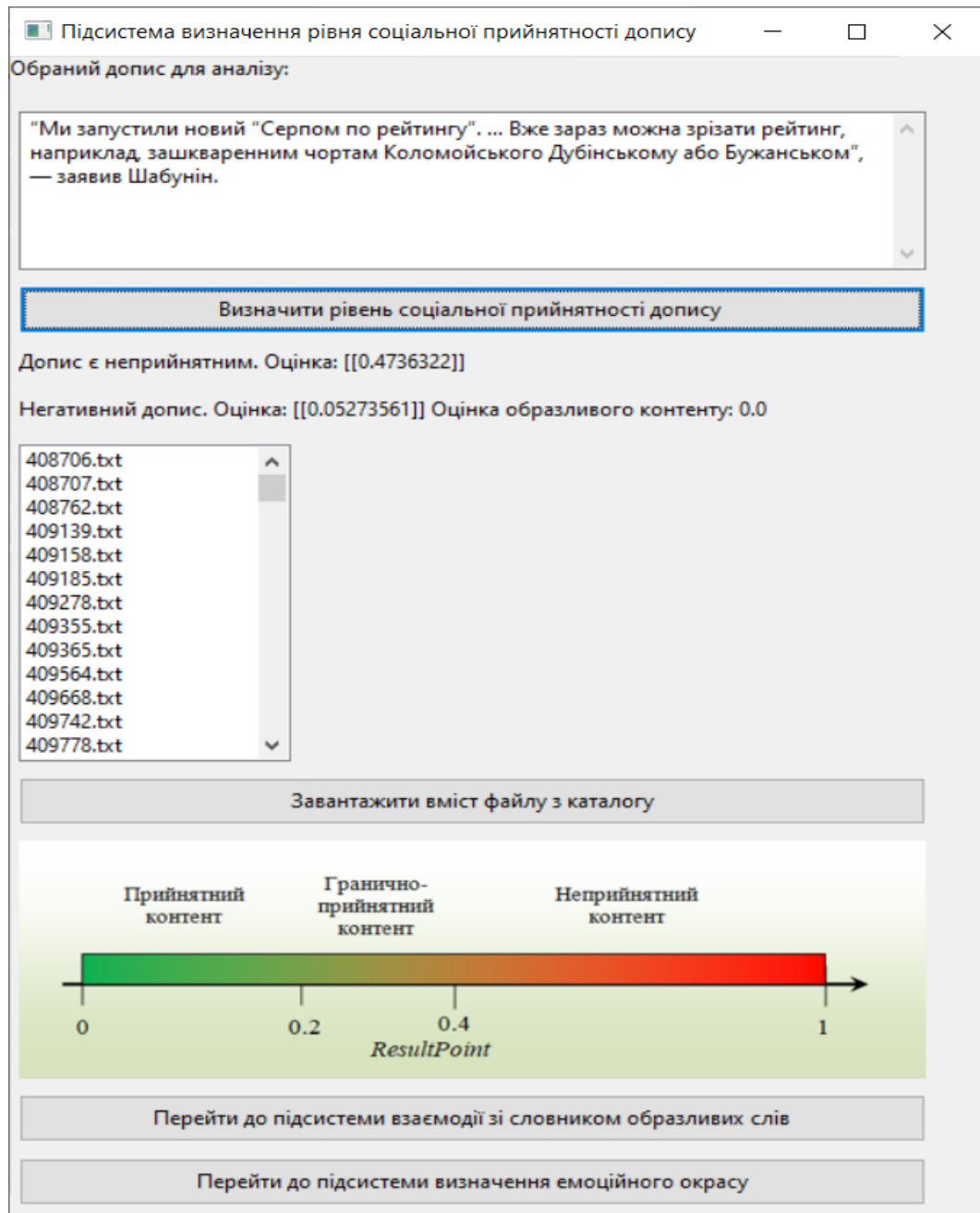


Рисунок 3.5 – Підсистема визначення рівня соціальної прийнятності допису

Отже, таким чином було реалізовано складові частини інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів, що складалася із підсистеми неймережевого визначення емоційного забарвлення допису, підсистеми визначення рівня соціальної прийнятності текстового контенту дописів, підсистеми визначення рівня образливого вмісту та підсистеми взаємодії зі словником образливих слів.

3.3 Тестування інформаційної системи

Для перевірки коректності роботи створеної інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів за семантичними ознаками було створено ряд тестових випадків.

Першим тестовим випадком буде перевірка відображення обраного образливого слова зі словника образливих слів.

Таблиця 3.1 – Тест-кейс VB0001

Тест-кейс ID: VB0001	Приоритет: 1	Створено: 1.04.2023, Загуровська Я.С.
Назва: Перевірка відображення обраного образливого слова зі словника образливих слів		
Кроки		Очікуваний результат
1. Запустити програму		Відкривається головна форма програми
2. Натиснути кнопку «Перейти до підсистеми взаємодії зі словником образливих слів»		Відкривається підсистема роботи зі словником образливих слів
3. Натиснути на слово з переліку образливих слів словника		
4. Перевірити показ обраного слова у полі «Обране слово»		Обране слово показано у полі «Образливе слово»
Результат виконання тест-кейсу: перевірку пройдено успішно.		

Після запуску застосування та виконання кроків, що вказані в таблиці 3.1, користувач буде бачити результат проілюстрований на рисунку 3.6.

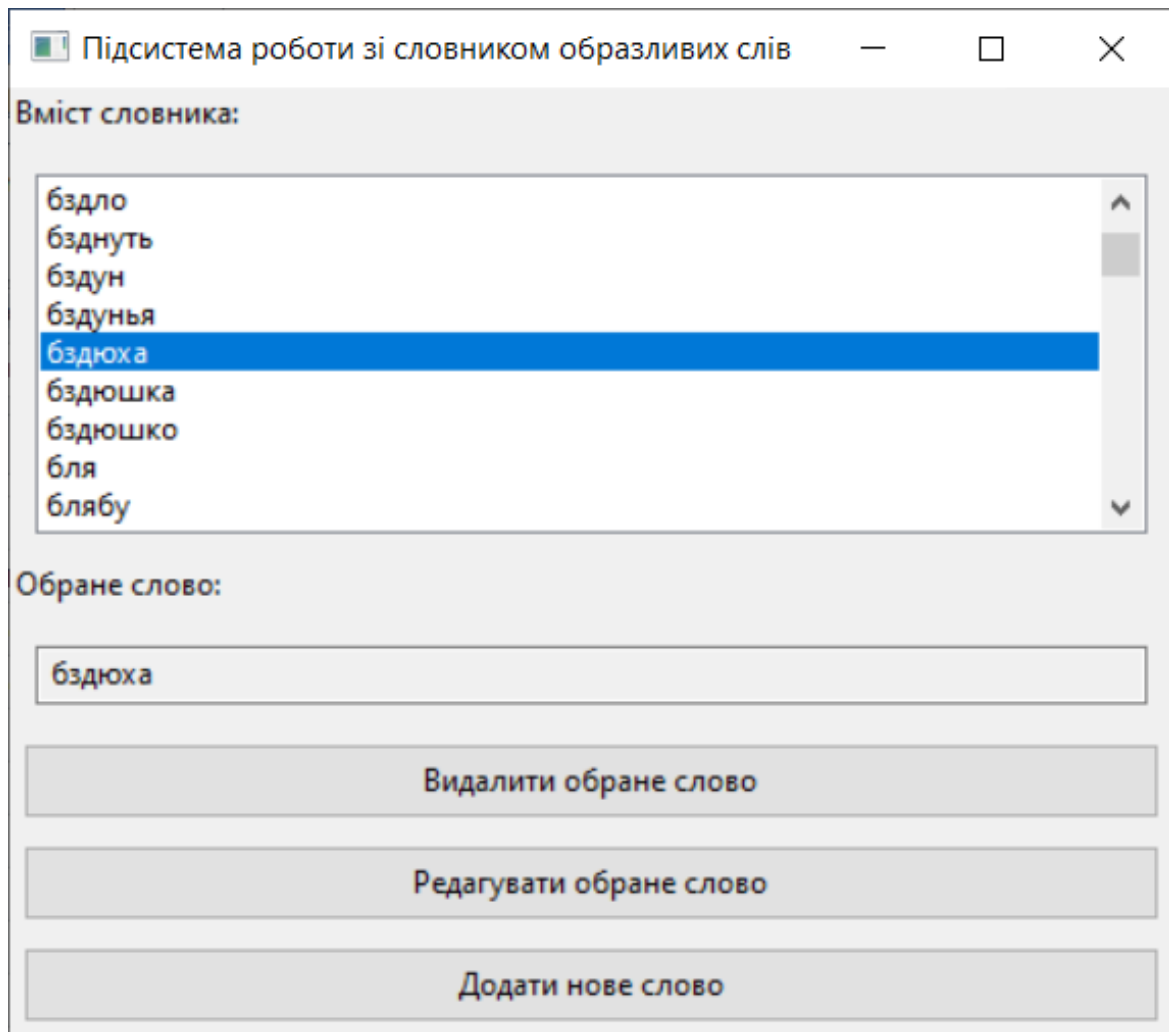


Рисунок 3.6 – Результат коректного виконання тесту

Другим тестовим випадком буде перевірка коректності нейромережевого визначення емоційного окрасу допису. Кроки пропонованого тесту відображені у таблиці 3.2.

Таблиця 3.2 – Тест-кейс VB0002

Тест-кейс ID: VB0002	Пріоритет: 1	Створено: 1.04.2023, Загуровська Я.С.
Назва: Перевірка коректності нейромережевого визначення емоційного окрасу допису		
Кроки		Очікуваний результат
<ol style="list-style-type: none"> 1. Запустити програму 2. Натиснути кнопку «Перейти до підсистеми визначення емоційного окрасу допису» 3. Ввести у текстове поле «Обраний допис для аналізу» текст: «Другий день тече вода зі стелі, але ці дурнуваті козли нічого не роблять! Скільки можна це терпіти? Ідіоти !» 4. Натиснути на кнопку «Визначити емоційний окрас допису» 5. Перевірити результат 		<p>Відкривається головна форма програми</p> <p>Відкривається підсистема визначення емоційного окрасу допису</p> <p>Очікується негативний допис</p>
Результат виконання тест-кейсу: перевірку пройдено успішно.		

Після запуску застосування та виконання кроків, що вказані в таблиці 3.2, користувач побачить очікуваний результат. Підтвердження успішного виконання тестового випадку VB0002 наведено на рисунку 3.7.

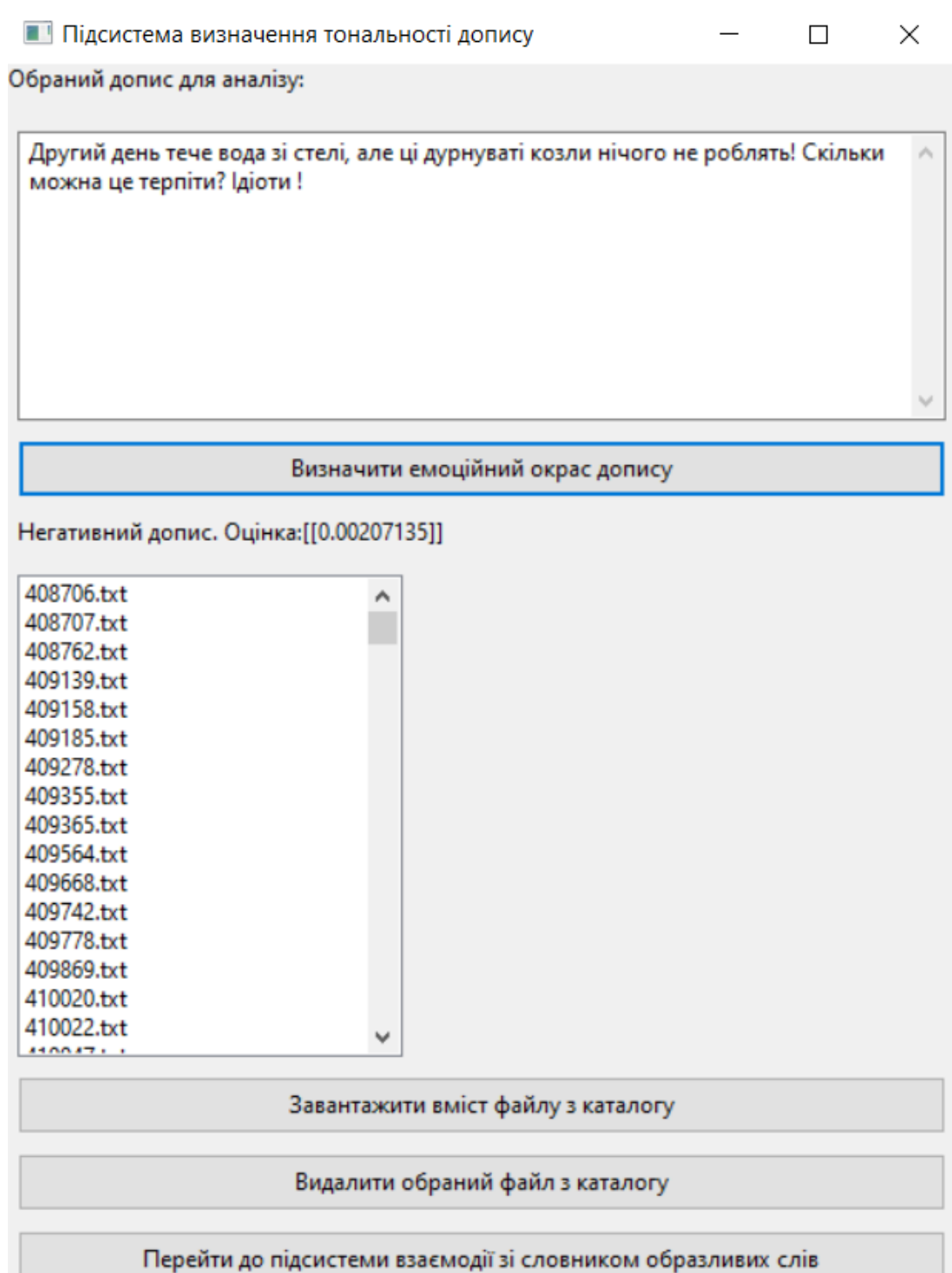


Рисунок 3.7 – Результат проходження тесту VB0002

Наступним буде протестовано коректність визначення рівня соціальної прийнятності текстового контенту допису. Кроки тестового випадку описано у таблиці 3.3.

Таблиця 3.3 – Тест-кейс VB0003

Тест-кейс ID: VB0003	Пріоритет: 1	Створено: 1.04.2023, Загуровська Я.С.
Назва: Перевірка коректності визначення рівня соціальної прийнятності текстового контенту допису		
Кроки		Очікуваний результат
<ol style="list-style-type: none"> 1. Запустити програму 2. Ввести у текстове поле «Обраний допис для аналізу» текст: «Обраний допис для аналізу» текст: «Другий день тече вода зі стелі, але ці дурнувати козли нічого не роблять! Скільки можна це терпіти? Ідіоти !» 3. Натиснути на кнопку «Визначити рівень соціальної прийнятності допису» 4. Перевірити результат 		<p>Відкривається головна форма програми</p> <p>Допис визначено неприйнятним</p>
Результат виконання тест-кейсу: перевірку пройдено успішно.		

Після запуску застосування та виконання кроків, що вказані в таблиці 3.2, користувач побачить очікуваний результат. Підтвердження успішного виконання тестового випадку VB0002 наведено на рисунку 3.8.

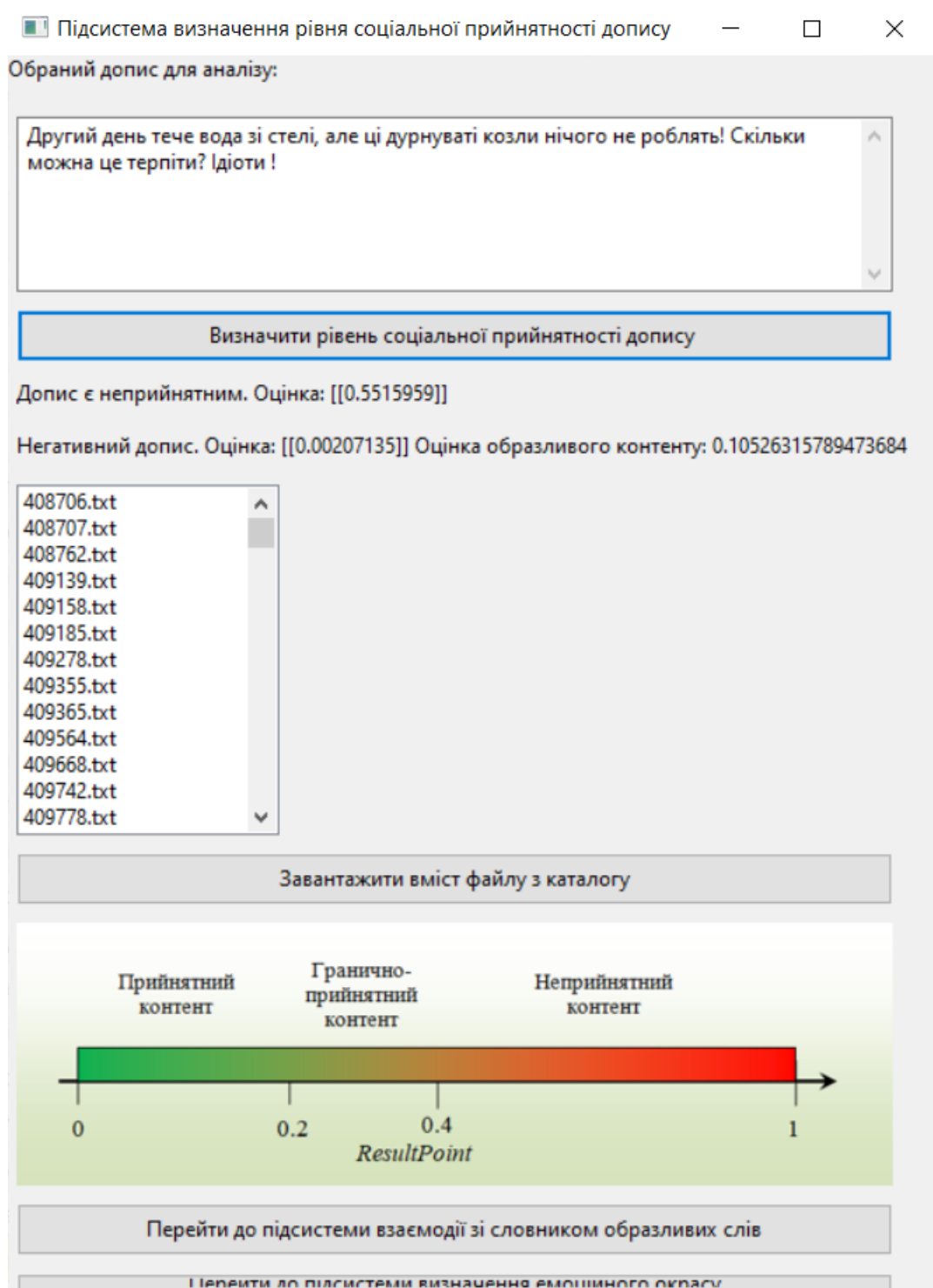


Рисунок 3.8 – Результат проходження тесту VB0003

Отже, засобами тест-кейсів було виконано перевірку коректності роботи створеної інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів за семантичними ознаками. При тестуванні системи некоректно працюючих функцій не виявлено.

3.4 Інструкція користувача для програмної реалізації методу визначення рівня соціальної прийнятності текстового контенту дописів

Для зручності використання створеної та протестованої програмної реалізації було розроблено відповідну інструкцію користувача. Головною є підсистема визначення рівня соціальної прийнятності допису. Вона є головним вікном, яке запускається одразу після старту програми. Інтерфейс зображено на рисунку 3.9.

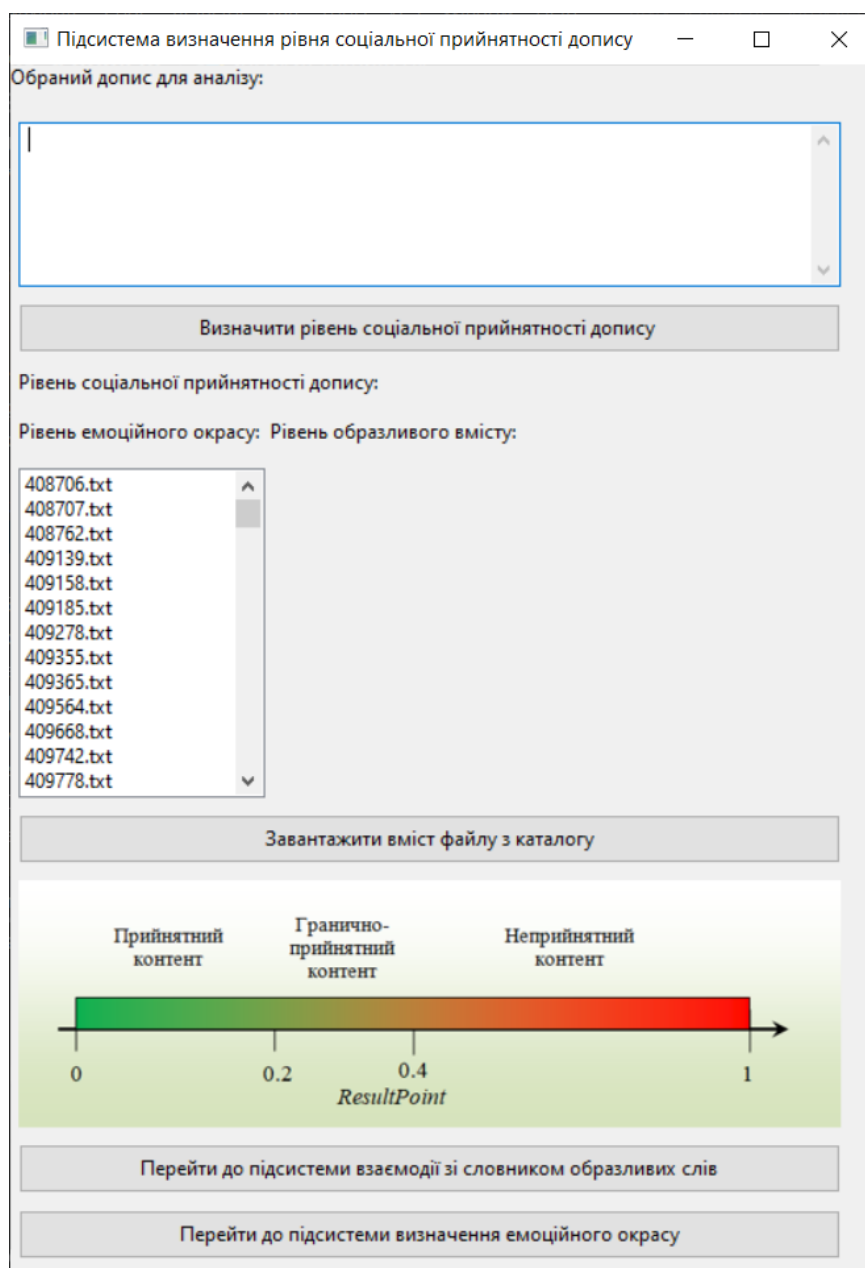


Рисунок 3.9 – Підсистема визначення рівня соціальної прийнятності допису

Тут можна визначити як рівень соціальної прийнятності власного або скопійованого допису, так і одного з дописів датасету. Для визначення рівня соціальної прийнятності власного допису потрібно у вікно «Обраний допис для аналізу» увести текст, який є потреба проаналізувати. Після уведення тексту потрібно натиснути кнопку «Визначити рівень соціальної прийнятності допису» (рисунок 3.10).

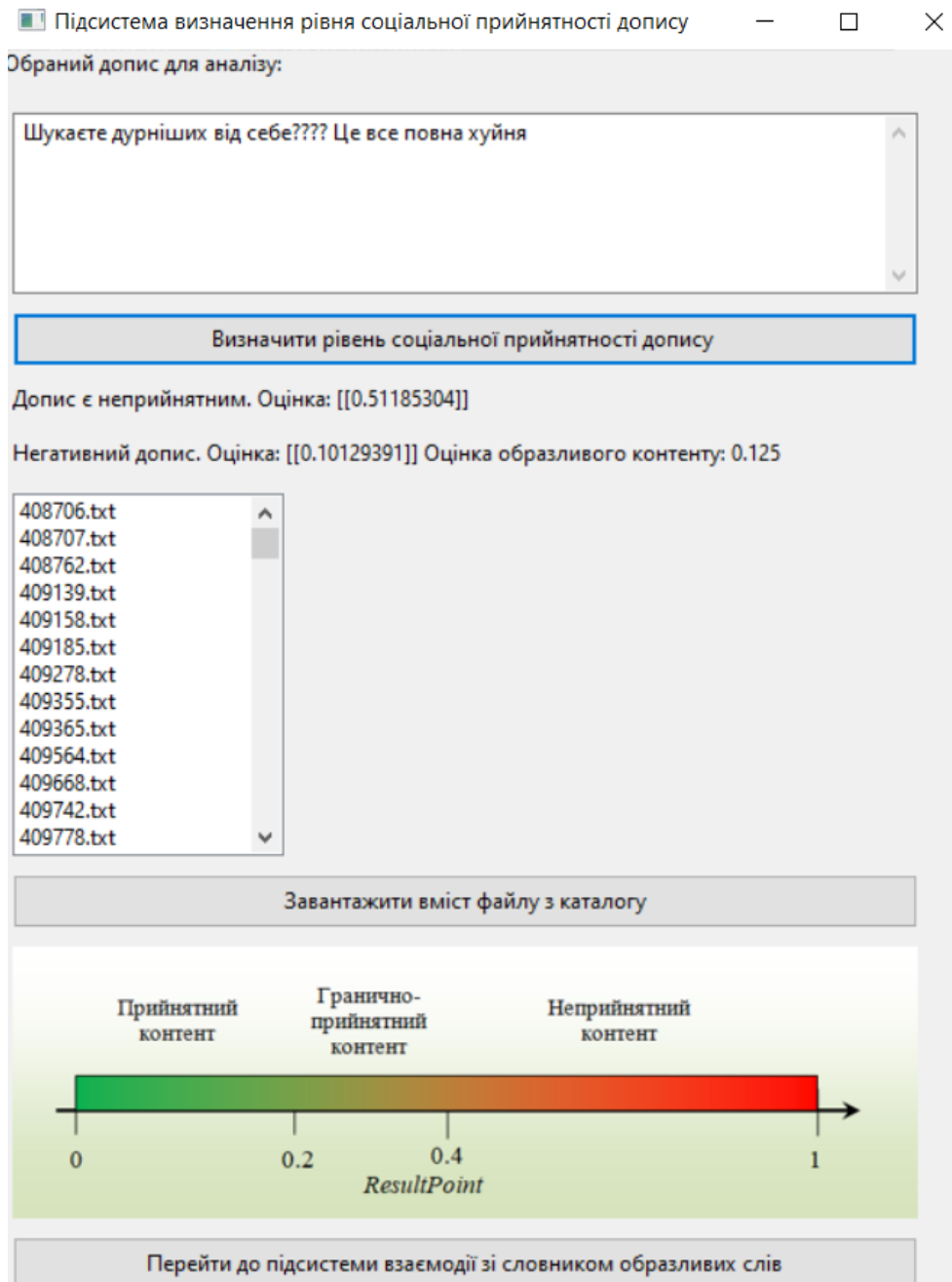


Рисунок 3.10 – Приклад визначення рівня соціальної прийнятності власного допису

Також можна визначити рівень соціальної прийнятності допису з датасету. Для цього потрібно обрати з переліку текстів датасету текст, який користувач хоче проаналізувати, та натиснути кнопку «Завантажити вміст файлу з каталогу». Для відображення результатів оцінки допису потрібно ще натиснути кнопку «Визначити рівень соціальної прийнятності допису» (рисунок 3.11).

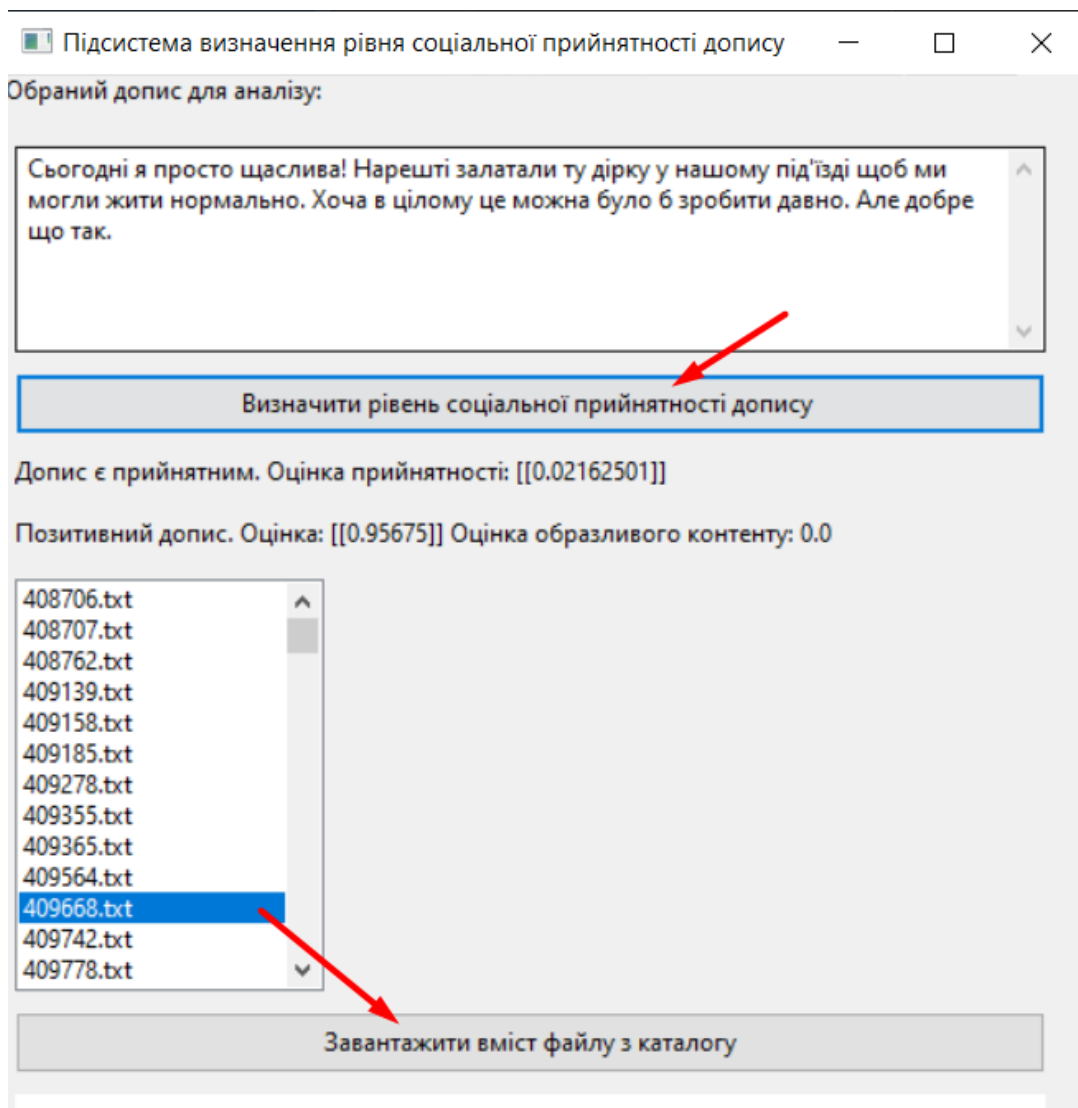


Рисунок 3.11 – Визначення рівню прийнятності допису з датасету

З цієї підсистеми можна також перейти на підсистему визначення емоційного окрасу повідомлення. Для цього потрібно натиснути на відповідну кнопку внизу екрану «Перейти до підсистеми визначення емоційного окрасу» (рисунок 3.12).

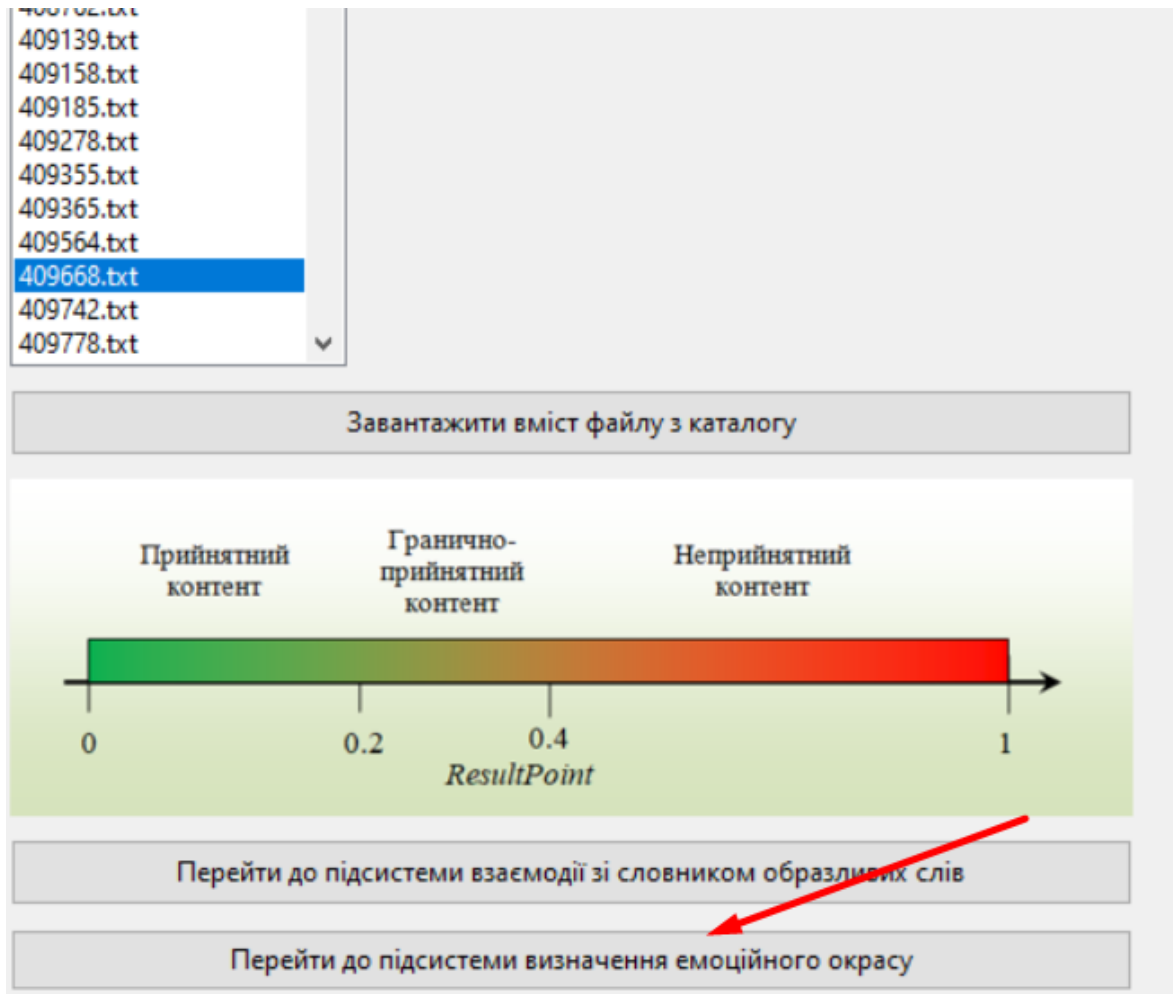


Рисунок 3.12 – Кнопка переходу на підсистему визначення емоційного окрасу допису

Після переходу на підсистему визначення емоційного окрасу допису користувач буде бачити інтерфейс, проілюстрований на рисунку 3.13. Функціонал підсистеми буде схожим на вищеписану, проте має ще додаткові можливості видалення обраного файлу з датасету.

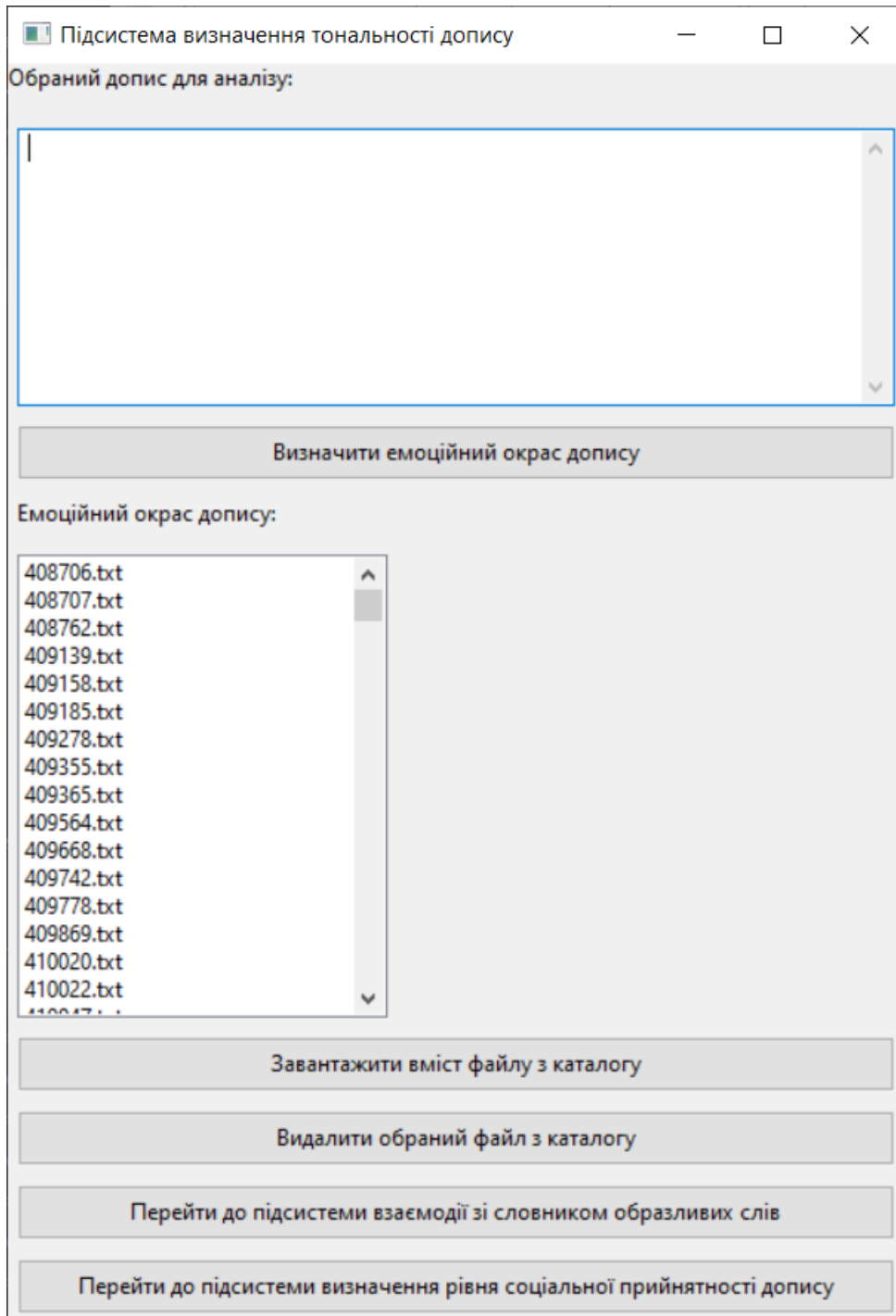


Рисунок 3.13 – Стартовий вигляд підсистеми визначення емоційного окрасу допису

Для визначення емоційного окрасу власного допису, необхідно ввести (або скопіювати і вставити) текст у поле «Обраний допис для аналізу». І натиснути кнопку «Визначити емоційний окрас допису». Після чого користувач побачить оцінку рівня емоційного окрасу (рисунок 3.14).

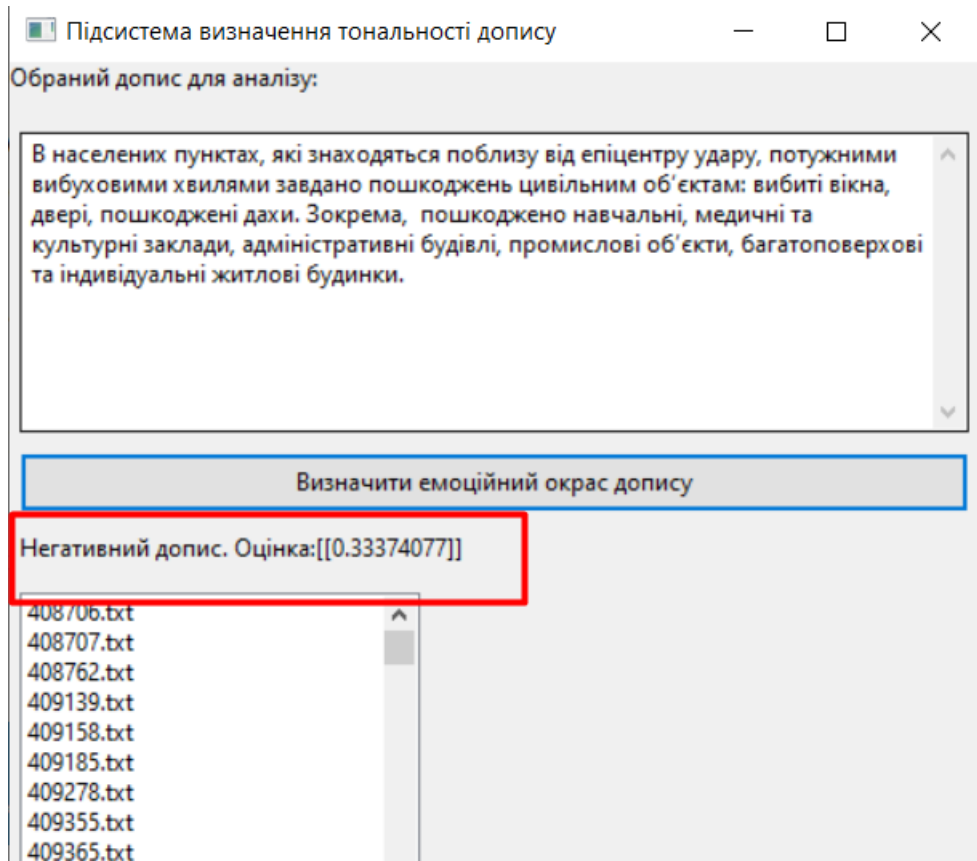


Рисунок 3.14 – Приклад визначення емоційного окрасу допису

Також у рамках даної підсистеми є можливість видалити дописи з датасету. Для цього користувачу потрібно натиснути на допис, який необхідно видалити, та натиснути кнопку «Видалити обраний файл з каталогу».

Також є можливість із даної підсистеми перейти на підсистему визначення рівню прийнятності допису, та перейти на підсистему роботи із словником образливих слів. Відповідні кнопки проілюстровано на рисунку 3.15.

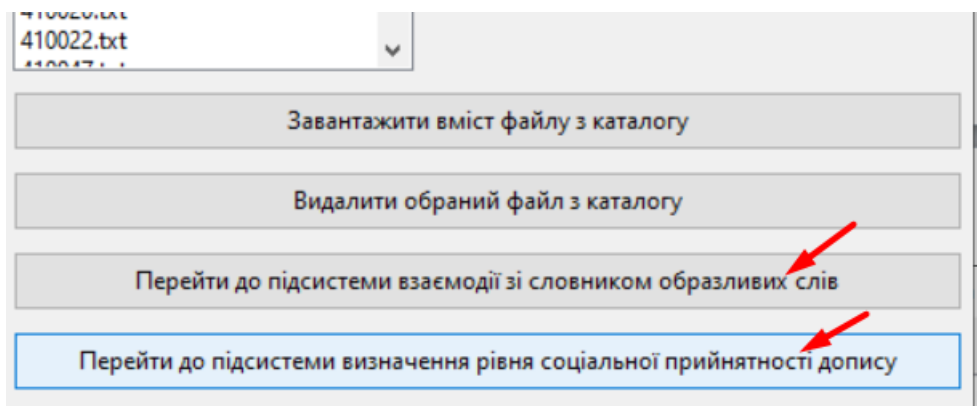


Рисунок 3.15 – Кнопки переходів на інші підсистеми

Якщо користувач натисне на кнопку «Перехід до підсистеми взаємодії зі словником образливих слів», то відбудеться перехід на вказану підсистему (рисунок 3.16).

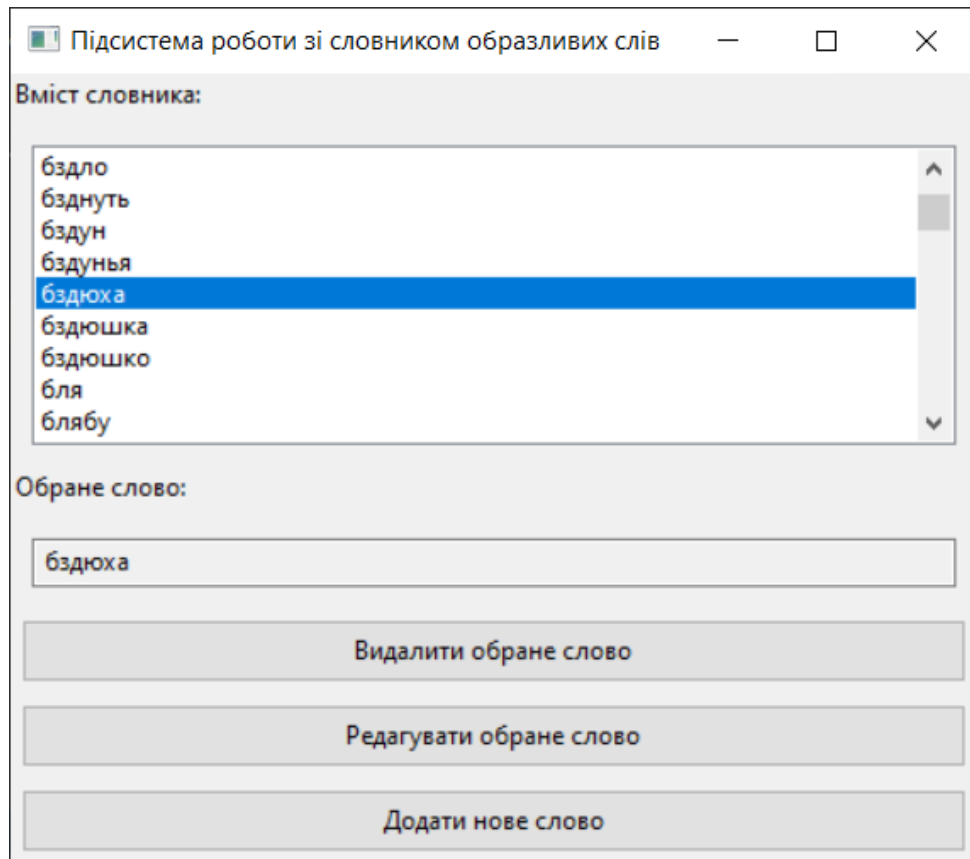


Рисунок 3.16 – Підсистема роботи зі словником образливих слів

У рамках підсистеми є можливість перегляду обраного слова зі словником, видалення обраного слова і додавання нового слова у словник.

Для видалення обраного слова зі словника необхідно натиснути кнопку «Видалити обране слово» (рисунок 3.17).

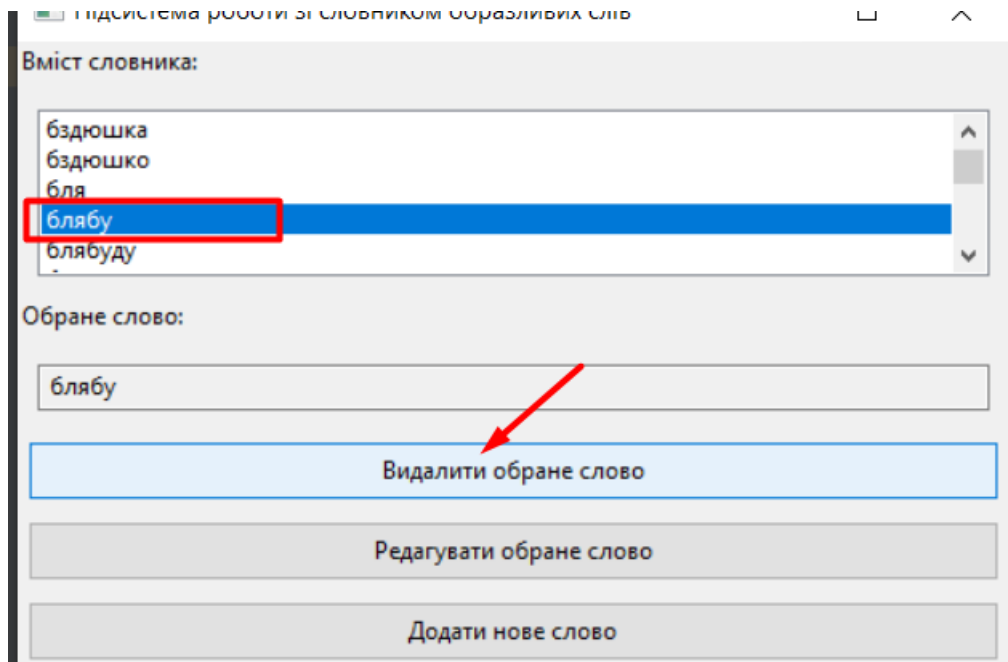


Рисунок 3.17 – Вибір слова для видалення

Результат видалення обраного слова показано на рисунку 3.18.

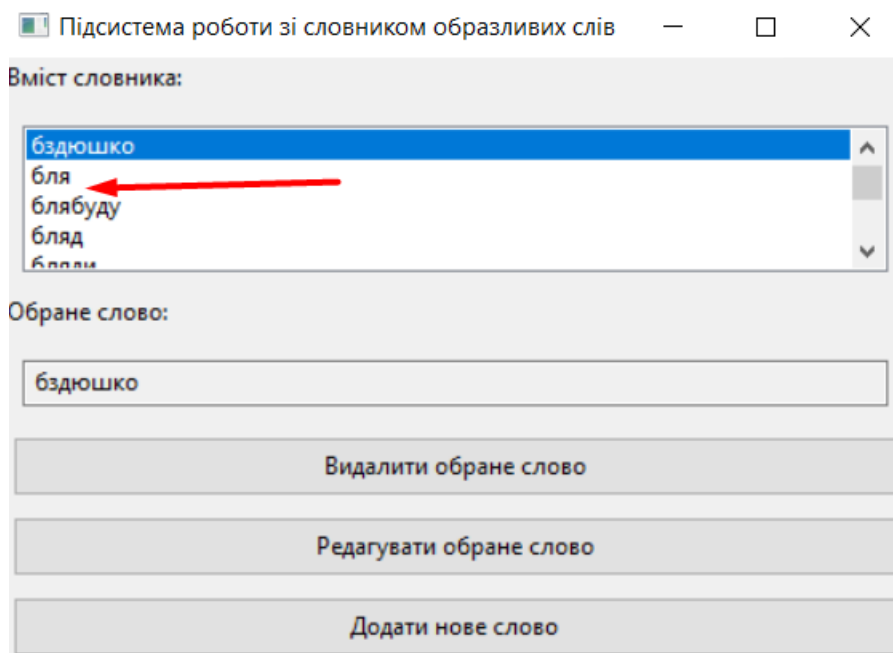


Рисунок 3.18 – Результат видалення слова

Для додавання нового слова в поле «Обране слово» потрібно додати слово, після чого натиснути на кнопку «Додати нове слово» (рисунок 3.19). Відповідне слово буде додано до словника образливих слів.

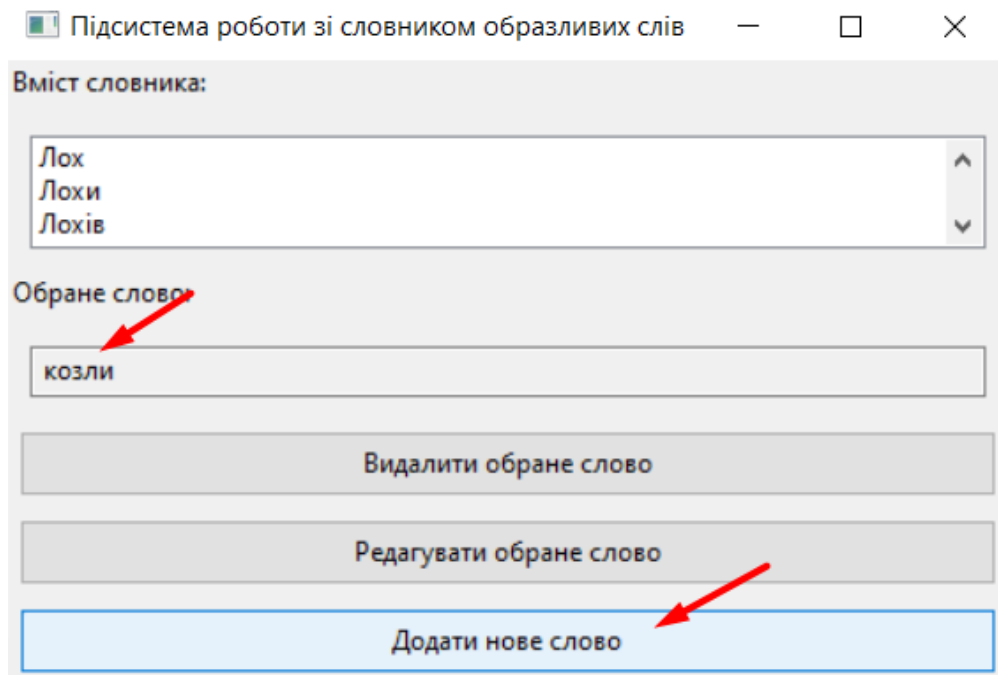


Рисунок 3.19 – Процес додавання нового слова у словник

Отже, таким чином було створено інструкцію до розробленої інформаційної системи на основі розробленого методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом.

3.5 Дослідження ефективності впливу параметрів навчання на точність класифікації

Оскільки RNN – нейромережа з великою кількістю параметрів, є потреба дослідити їх вплив на якість роботи навченої моделі. У ході роботи було здійснено ряд експериментів з навчання нейромережі, з параметрами, наведеними в таблиці 3.4.

З таблиці 3.4 видно, що найкращий результат спостерігається при кількості епох навчання рівним 10. Також результати проведеного експерименту проілюстровано на рисунку 3.20.

Таблиця 3.4 – Вплив кількості епох на точність класифікації

Кількість епох:	7	8	9	10	11	15	20
Точність:	0.847393365	0.852132701	0.850236967	0.854028436	0.850236967	0.846445498	0.840758294
Оцінка тестового позитивного зразка:	0.9527581	0.9727791	0.99589443	0.99385715	0.9938995	0.99821	0.9983682



Рисунок 3.20 – Діаграма впливу кількості епох на точність класифікації

Також, як видно з рисунка 3.20 – збільшення кількості епох не призводить до покращення процесу бінарної класифікації для задачі виявлення негативного окрасу з метою подальшої ідентифікації неприйняттого контенту.

Висновок до третього розділу

У рамках третього розділу було розроблено архітектуру інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів, що використовує розроблений метод, що складається із таких основних класів:

NeuralNetwork, LevelContent, EmotionDetect, WorkingWithVocab, які спроможні реалізувати зазначений функціонал.

Було реалізовано відповідні складові частини інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів, що складалася із підсистеми нейромережевого визначення емоційного забарвлення допису, підсистеми визначення рівня соціальної прийнятності текстового контенту дописів, підсистеми визначення рівня образливого вмісту та підсистеми взаємодії зі словником образливих слів. Для функціонування інформаційної системи, було навчено нейромережу на обраних вище датасетах, та збережено навчену модель для подальшого використання.

Для перевірки коректності функціонування складових було розроблено ряд тест-кейсів для створеної інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів за семантичними ознаками. При тестуванні системи некоректно працюючих функцій не виявлено і весь зазначений функціонал працює коректно.

Для зручності користування розробленою інформаційною системою було створено інструкцію користувача.

Висновки

Метою кваліфікаційної роботи було встановлено розробку методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом та відповідного програмного забезпечення, що буде використовувати розроблений метод. Для досягнення наведеної мети в рамках роботи були поставлені та виконані наступні завдання:

- проведено дослідження предметної області для задачі визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом;

- розроблено метод визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом призначеного для перетворення вхідної інформації у вигляді допису для оцінки рівня соціальної прийнятності у вихідні дані у вигляді числової оцінки прийнятності допису;

- обрано архітектуру нейронної мережі для визначення емоційного окрасу текстового допису у вигляді тришарової рекурентної нейромережі, яка дозволяє здійснювати визначення емоційного окрасу з точністю бінарної класифікації 0.86;

- спроектовано структуру інформаційної системи на базі методу визначення рівня соціальної прийнятності текстового контенту дописів, що складалася із підсистеми нейромережевого визначення емоційного забарвлення допису, підсистеми визначення рівня соціальної прийнятності текстового контенту дописів, підсистеми визначення рівня образливого вмісту та підсистеми взаємодії зі словником образливих слів;

- програмно реалізовано інформаційну систему визначення рівня соціальної прийнятності текстового контенту дописів;

- проведено тестування створеного програмного засобу, де було наочно показано що весь заявлений функціонал працює коректно.

При реалізації методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом для роботи з масивами даних було застосовано бібліотеку NumPy.

Також у роботі було визначено відповідні набори даних для навчання неймережі та для застосування ідентифікації слів образливого контенту. У якості даних для числової оцінки рівня соціальної прийнятності текстового контенту дописів використано набір даних «AbusiveLanguageDataset», а для навчання та валідації неймережі було використано «Ukrainian Twitter Corpus».

Перелік посилань

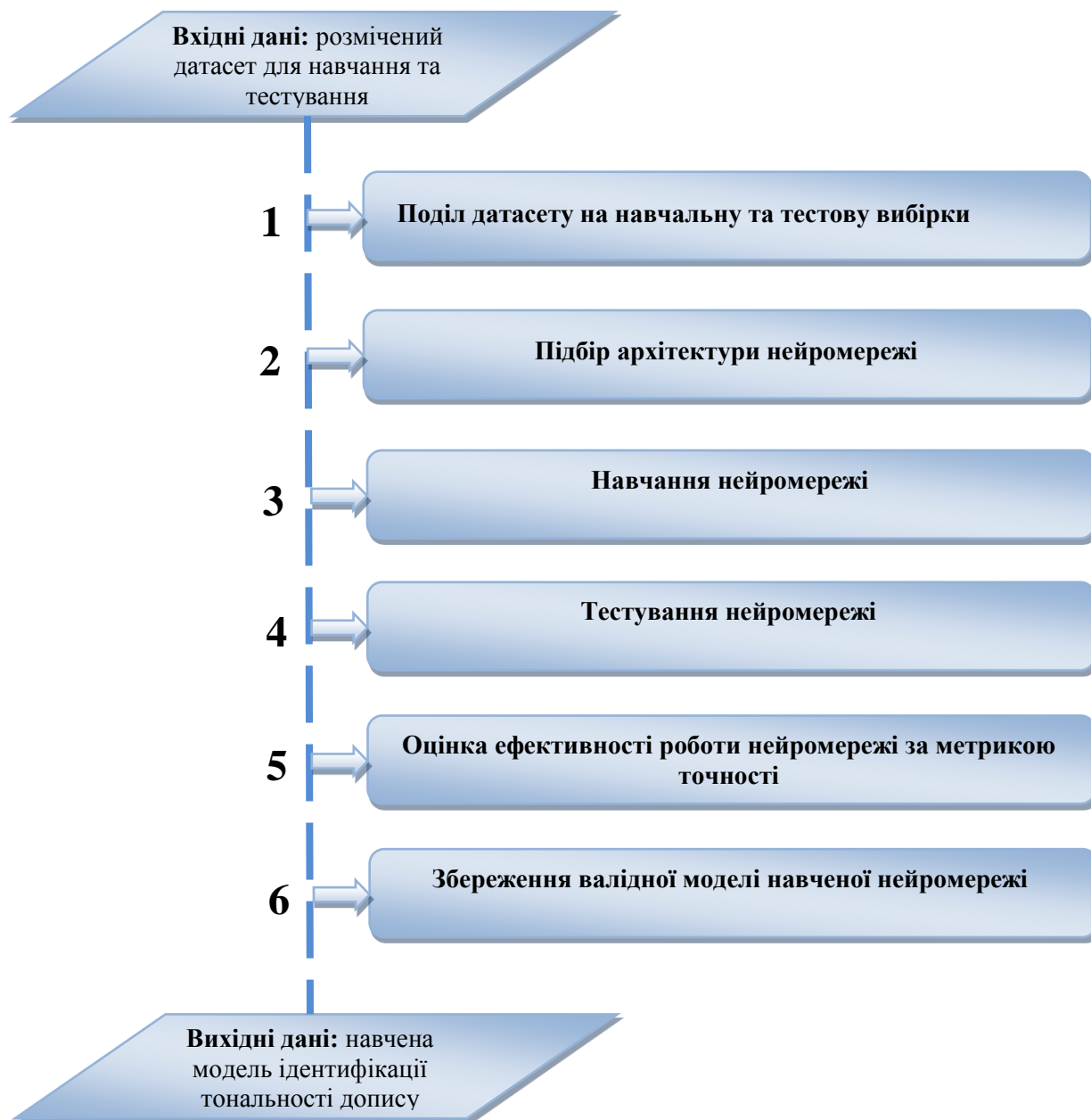
1. Internet Social Media & User-Generated Content. URL: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
2. Блокування в соцмережах: хто і як це має регулювати. URL: <https://detector.media/infospace/article/172358/2019-11-12-blokuvannya-v-sotsmerezhakh-khto-i-yak-tse-maie-regulyuvaty/>
3. Violence and incitement. URL: <https://transparency.fb.com/en-gb/policies/community-standards/violence-incitement/>
4. Bullying and harassment. URL: <https://transparency.fb.com/en-gb/policies/community-standards/bullying-harassment/>
5. Центр політики Google ADS. Неприйнятний контент. URL: <https://support.google.com/adspolicy/answer/6015406?hl=uk>
6. Аналіз емоційного окрасу тексту в соціальних мережах. URL: <https://ekmair.ukma.edu.ua/server/api/core/bitstreams/673d4b8d-e93f-42ec-b260-a503e13a4eb0/content>
7. Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS" by Goutam Chakraborty, Murali Pagolu, and Satish Garla. URL: <http://support.sas.com/publishing/pubcat/chaps/65646.pdf>
8. Visual text analytics in context of digital humanities. URL: <https://kops.uni-konstanz.de/server/api/core/bitstreams/6a0e30d7-c917-4593-98b6-758f49ca836c/content>
9. Learn How to Use Support Vector Machines (SVM) for Data Science. URL: <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/>
10. Support Vector Machines with Scikit-learn Tutorial. URL: <https://www.datacamp.com/tutorial/svm-classification-scikit-learn-python>
11. Sentiment Analysis. URL: <https://www.helenkapatsa.ru/analiz-tonalnosti-tieksta/>

12. Detection of Abusive Speech for Mixed Sociolects of Russian and Ukrainian Languages. URL: <https://nlp.fi.muni.cz/raslan/2018/paper04-Andrusyak.pdf>
13. Chen, Y., Zhou, Y., Zhu, S., And Xu, H. Detecting offensive language in social media to protect adolescent online safety. URL: <https://ieeexplore.ieee.org/document/6406271>
14. Semantria API. Lexalytics' Text and Sentiment Analysis, in the cloud. URL: <https://www.lexalytics.com/semantria/>
15. Sentistrength For Mac. URL: <https://sohafilms.weebly.com/blog/sentistrength-for-mac>
16. SentiStrength. URL: <http://sentistrength.wlv.ac.uk/>
17. How do text characteristics impact user engagement in social media posts: Modeling content readability, length, and hashtags number in Facebook. URL: <https://www.sciencedirect.com/science/article/pii/S2667096822000118>
18. СЛОВНИК NLP. URL: <https://medium.com/stinopys/%D1%81%D0%BB%D0%BE%D0%B2%D0%BD%D0%B8%D0%BA-nlp-b0fab1027551#9cdc>
19. Understanding RNN and LSTM. URL: <https://aditi-mittal.medium.com/understanding-rnn-and-lstm-f7cdf6dfc14e>
20. 10 Popular Datasets For Sentiment Analysis. URL: <https://analyticsindiamag.com/10-popular-datasets-for-sentiment-analysis/>
21. Hate Speech Dataset Catalogue. URL: <https://hatespeechdata.com/>
22. ukr-twi-corpus. URL: <https://github.com/saganoren/ukr-twi-corpus>
23. TensorFlow. URL: <https://en.wikipedia.org/wiki/TensorFlow>
24. Create production-grade machine learning models with TensorFlow. URL: <https://www.tensorflow.org/>
25. Simple. Flexible. Powerful. URL: <https://keras.io/>
26. Meet the new NumPy docs team leads. URL: [Meet the new NumPy docs team leads](#)

ДОДАТКИ

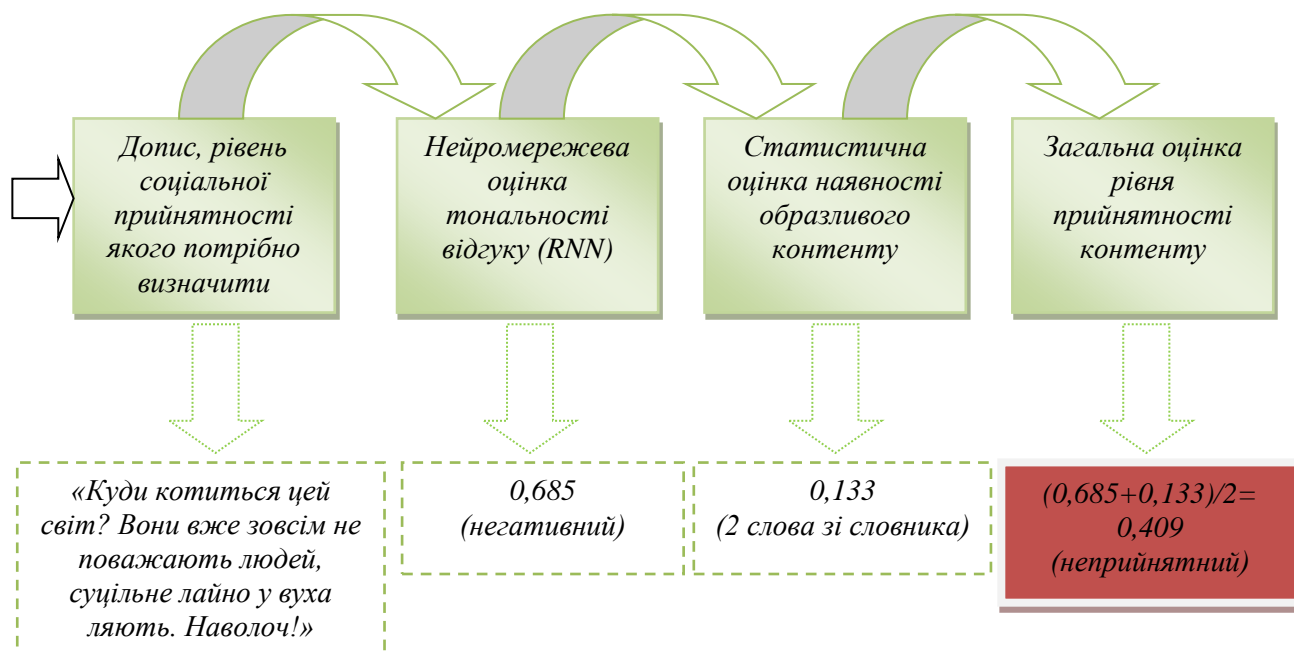
Додаток А

Послідовність кроків при проведенні нейромережевої оцінки тональності допису



Додаток Б

Схема етапів роботи інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів



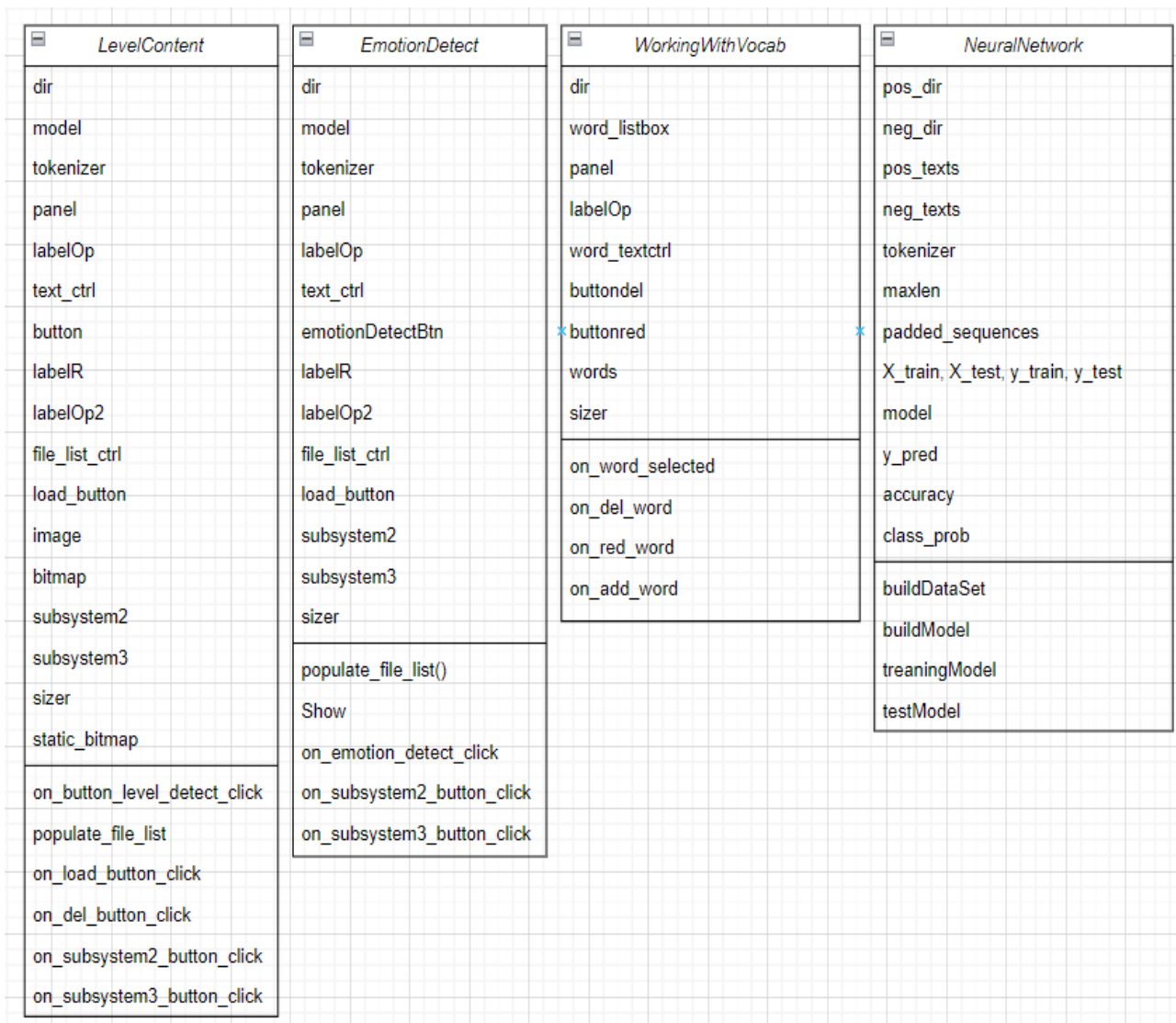
Додаток В

Архітектура нейронної мережі RNN



Додаток Г

Діаграма класів



Додаток Д

Презентаційний матеріал

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

МЕТОД ВИЗНАЧЕННЯ РІВНЯ СОЦІАЛЬНОЇ ПРИЙНЯТНОСТІ ТЕКСТОВОГО КОНТЕНТУ ДОПИСІВ ЗА ЇХ СЕМАНТИЧНИМ АНАЛІЗОМ



Виконала:
студентка 3 курсу, групи КНС-20-1
Загуровська Яна Сергіївна



Керівник:
к.т.н., доцент кафедри КН
Багрій Руслан Олександрович

Актуальність

Кількість користувачів соціальних мереж у світі постійно зростає, що свідчить про значний обсяг контенту, що був опублікований і публікується щодня в цих мережах. У зв'язку з цим зростає і важливість забезпечення якості цього контенту. Зокрема, важливим аспектом є соціальна прийнятність контенту, оскільки несанкціонована, образлива або неправдива інформація може викликати шкоду та негативні наслідки для людей, груп чи суспільства загалом.

Об'єкт дослідження – цифровий текстовий контент дописів, який публікується в соціальних мережах.

Предмет дослідження – моделі, методи та засоби машинного навчання, які дозволяють визначити рівень соціальної прийнятності цифрового текстового контенту на основі його семантичного аналізу.

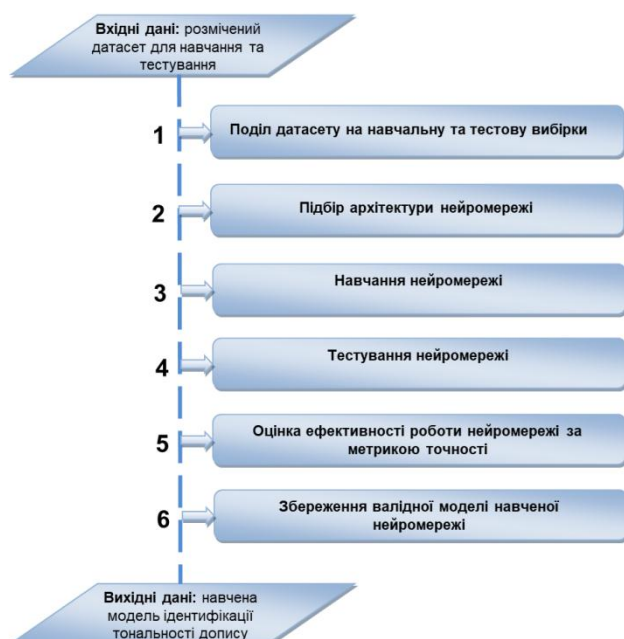
Мета і задачі роботи

Метою кваліфікаційної роботи бакалавра розробка методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом та відповідного програмного забезпечення, що буде використовувати розроблений метод, для чого слід вирішити задачі:

1. провести дослідження предметної області для задачі визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом;
2. розробити метод визначення рівня соціальної прийнятності текстового контенту дописів;
3. обрати архітектуру нейронної мережі для визначення емоційного окрасу текстового допису;
4. спроектувати структуру інформаційної системи на базі методу визначення рівня соціальної прийнятності текстового контенту дописів;
5. програмно реалізувати інформаційну систему визначення рівня соціальної прийнятності текстового контенту дописів;
6. провести тестування створеного програмного засобу.



Узагальнена схема методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом



Послідовність кроків при проведенні нейромережевої оцінки тональності допису

Числова оцінка рівня соціальної прийнятності текстового контенту дописів

Числова оцінка рівня соціальної прийнятності текстового контенту дописів буде розраховуватись за формулою:

$$ResultPoint = \frac{OffensiveC + (1 - NNPoint)}{2},$$

де *OffensiveC* – статистична оцінка наявності образливого контенту, *NNPoint* – нейромережева оцінка тональності допису.

Числова оцінка рівня соціальної прийнятності текстового контенту дописів

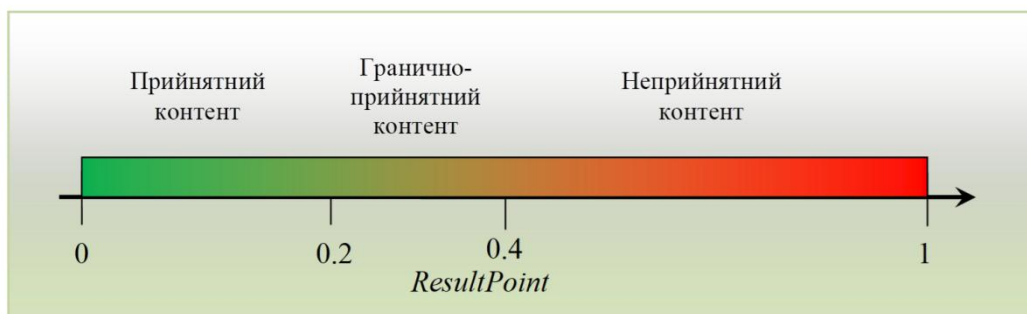
Статистична оцінка образливого контенту буде розраховуватись за формулою:

$$OffensiveC = \frac{O_w}{TotalCount},$$

де O_w – кількість образливих слів, що містяться у словнику, $TotalCount$ – загальна кількість слів у дописі.



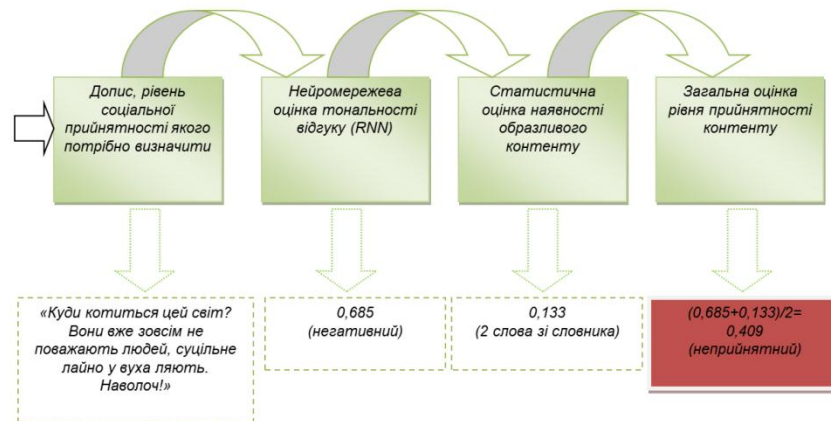
Рівні соціальної прийнятності текстового контенту дописів



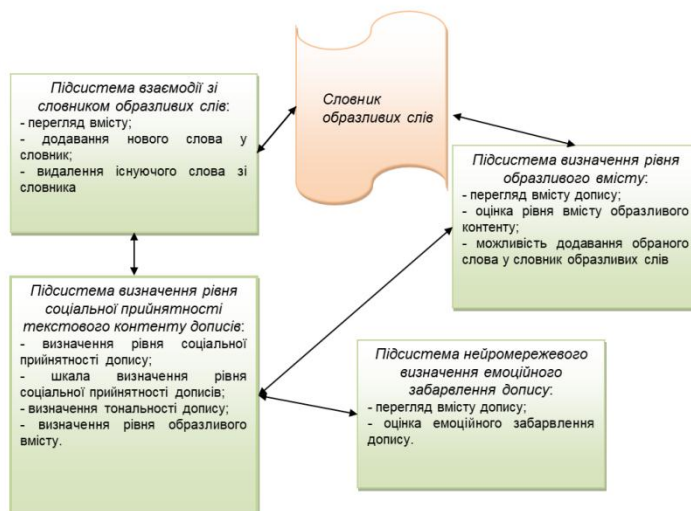
Прийнятним вважається допис, *ResultPoint* якого не перевищує 0.2, гранично-прийнятним вважається контент, оцінка якого знаходиться в межах від 0.2 до 0.4. Ціком неприйнятним вважається контент, оцінка якого перевищує 0.4.



Узагальнена схема методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом



← → ↻ ↺



Складові інформаційної системи визначення рівня соціальної прийнятності текстового контенту дописів

Архітектура нейронної мережі RNN



Засоби розробки

Мова програмування: **Python**

Інтерфейси користувача: бібліотека «**wx**»

Для побудови нейромережевої моделі, навчання та оцінки її продуктивності:

- «**TensorFlow**» та «**keras**» – для побудови та навчання RNN.
- «**sklearn.metrics**» – оцінка продуктивності за метриками

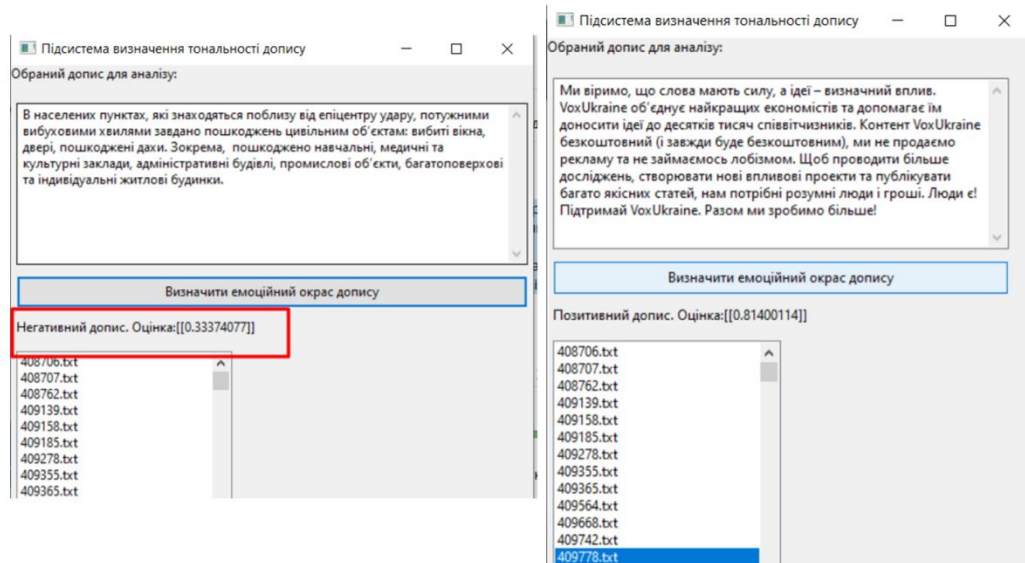
Набір даних для числової оцінки рівня соціальної прийнятності текстового контенту дописів:

- «**AbusiveLanguageDataset**» - датасет, який містить україномовні коментарі з різним рівнем образливого вмісту. Було використаний у якості даних для числової оцінки рівня соціальної прийнятності текстового контенту дописів.
- «**Ukrainian Twitter Corpus**» - для навчання та тестування нейромережі для виявлення емоційного забарвлення тексту, містить позитивні та негативні україномовні коментарі.

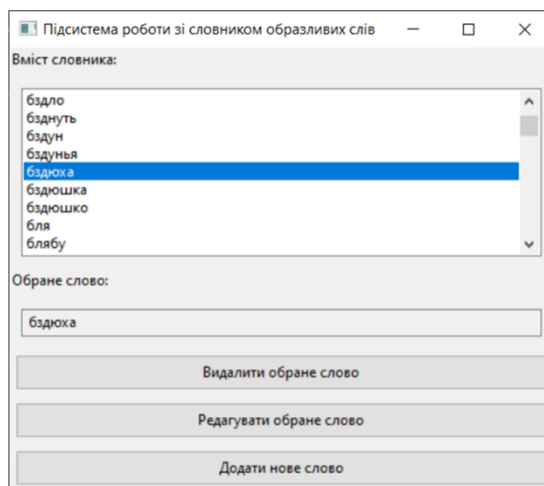


Програмна реалізація методу визначення рівня соціальної прийнятності текстового контенту дописів

Підсистема визначення тональності допису



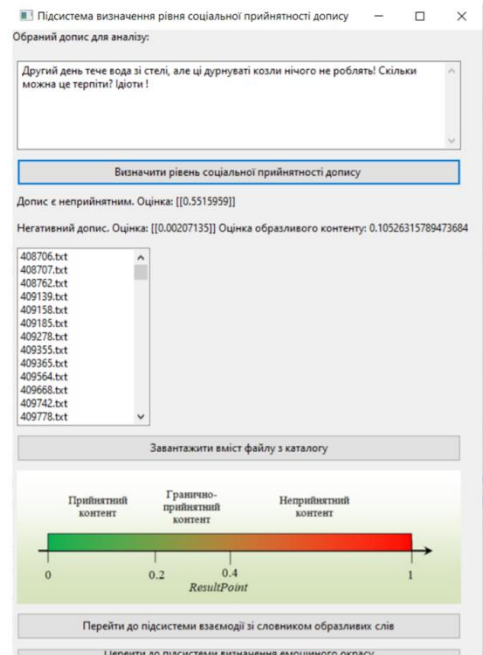
Програмна реалізація методу визначення рівня соціальної прийнятності текстового контенту дописів



Підсистема роботи зі словником образливих слів

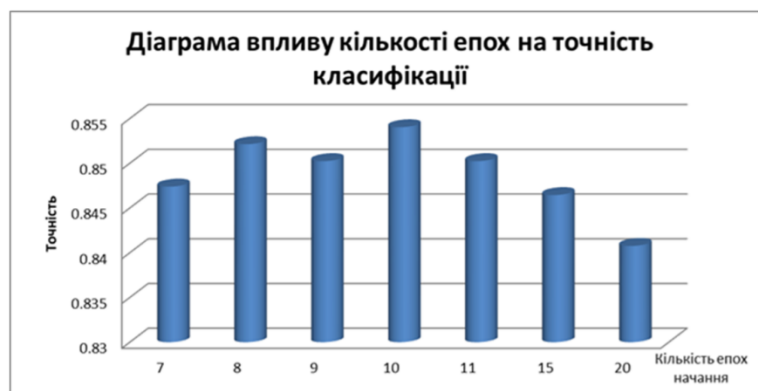
Програмна реалізація методу визначення рівня соціальної прийнятності текстового контенту дописів

*Підсистема
визначення рівня
соціальної
прийнятності
текстового
контенту допису*



Дослідження ефективності - вплив кількості епох на точність класифікації

Кількість епох:	7	8	9	10	11	15	20
Точність:	0.8473933	0.8521327	0.8502369	0.8540284	0.8502369	0.8464454	0.8407582
Оцінка тестового позитивного зразка:	65	01	67	36	67	98	94
Оцінка тестового позитивного зразка:	0.9527581	0.9727791	0.9958944	0.9938571	0.9938995	0.99821	0.9983682



Висновки

Метою кваліфікаційної роботи було встановлено розробку методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом та відповідного програмного забезпечення, що буде використовувати розроблений метод. Для досягнення наведеної мети в рамках роботи були поставлені та виконані наступні завдання:

- проведено дослідження предметної області для задачі визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом;
- розроблено метод визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом призначеного для перетворення вхідної інформації у вигляді допису для оцінки рівня соціальної прийнятності у вихідні дані у вигляді числової оцінки прийнятності допису;
- обрано архітектуру нейронної мережі для визначення емоційного окрасу текстового допису у вигляді тришарової рекурентної нейромережі, яка дозволяє здійснювати визначення емоційного окрасу з точністю бінарної класифікації 0.86;

Висновки

- спроектовано структуру інформаційної системи на базі методу визначення рівня соціальної прийнятності текстового контенту дописів, що складалася із підсистеми нейромережевого визначення емоційного забарвлення допису, підсистеми визначення рівня соціальної прийнятності текстового контенту дописів, підсистеми визначення рівня образливого вмісту та підсистеми взаємодії зі словником образливих слів;

- програмно реалізовано інформаційну систему визначення рівня соціальної прийнятності текстового контенту дописів;

- проведено тестування створеного програмного засобу, де було наочно показано що весь заявлений функціонал працює коректно.

При реалізації методу визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом для роботи з масивами даних було застосовано бібліотеку NumPy.

Також у роботі було визначено відповідні набори даних для навчання нейромережі та для застосування ідентифікації слів образливого контенту. У якості даних для числової оцінки рівня соціальної прийнятності текстового контенту дописів використано набір даних «AbusiveLanguageDataset», а для навчання та валідації нейромережі було використано «Ukrainian Twitter Corpus».

Ім'я користувача:
Кафедра КН

ID перевірки:
1015410618

Дата перевірки:
04.06.2023 11:48:08 EEST

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
04.06.2023 11:49:01 EEST

ID користувача:
100005671

Назва документа: КНс-20-1 Загуровська

Кількість сторінок: 59 Кількість слів: 8408 Кількість символів: 65989 Розмір файлу: 2.27 MB ID файлу: 1015073907

Виявлено модифікації тексту (можуть впливати на відсоток схожості)

5.17% Схожість

Найбільша схожість: 1.27% з джерелом з Бібліотеки (ID файлу: 1011420928)

4.82% Джерела з Інтернету 374 Сторінка 61

2.14% Джерела з Бібліотеки 120 Сторінка 62

0% Цитат

Вилучення цитат вимкнене

Вилучення списку бібліографічних посилань вимкнене

0% Вилучень

Немає вилучених джерел

Модифікації

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

Підозріле форматування 16 сторінок

Anti-Plagiarism v-15.257**Максимальне співпадіння з одним документом 41.0%**

Словники перевірки: en_US, ru_RU, ua_UA. Помилки в документах: 10%

ID: 114638 Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА Додано в БД: 2023-06-04 Автора: Я.С. Загуровська Керівники: Р.О. Багрій Консультанти: Опоненти:	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	55560	812	23636 (43%)	325 (40%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми
114078	Назва: ЗВІТ з професійної практики Додано в БД: 2023-05-26 Автора: Загуровська Я.С. Керівники: Скрипник Т.К. Консультанти: Опоненти:	22952 (41.0%)	348 (43.0%)

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

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

Назва: Метод визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом

Автор: студентка групи КНс-20-1 Загуровській Яні Сергіївні

Спеціальність: 122 – Комп'ютерні науки

Освітня програма: освітньо-професійна

Науковий керівник: к.т.н., доцент кафедри КН Багрій Р.О.

Після аналізу звіту подібності зроблено такий висновок:

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

Підтвердження:

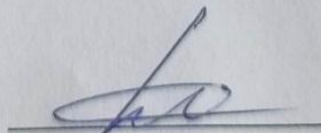
Запозичення, виявлені в роботі Загуровської Я.С., не є плагіатом, оскільки: запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; серед запозичень знаходяться загальновідомі терміни, скорочення та матеріали статей.

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

- за системою Anti-Plagiarism: 41%, з яких 41% є посиланням на власний звіт з професійної практики, що є допустимими запозиченнями які відносяться до описаних вище;

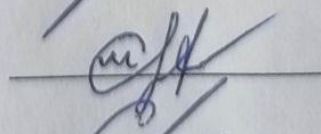
- за системою Unichек: 5.17 %.

Керівник роботи



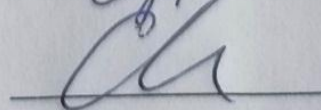
Руслан БАГРІЙ

Гарант ОП



Олександр МАЗУРЕЦЬ

Завідувач кафедри КН



Олександр БАРМАК



ВІДГУК НАУКОВОГО КЕРІВНИКА
на кваліфікаційну роботу бакалавра

студента гр. КНс-20-1 Загуровська Яна Сергіївна

за темою: Метод визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом

1. Актуальність теми

В останні роки зростає потреба у розвитку автоматизованих методів виявлення соціально неприйнятної контенту, оскільки Інтернет стає все більш популярним серед спільнот і вимагає ефективних механізмів фільтрації та модерації контенту.

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

Робота бакалавра повністю відповідає предметній області стандарту спеціальності 122 Комп'ютерні науки, оскільки вона поєднує елементи обробки природної мови, машинного навчання та інформаційних систем для розробки автоматизованого методу визначення рівня соціальної прийнятності текстового контенту.

3. Професійні та особистісні якості бакалавра

У роботі бакалавра видно прояв професійних якостей, таких як глибоке розуміння семантичного аналізу та програмування, а також здатність до творчого мислення та ініціативи. Особистісні якості, такі як систематичність, самодисципліна та наполеглива праця, були проявлені під час виконання кваліфікаційної роботи.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Загуровська Яна Сергіївна продемонстрував високий рівень самостійності під час виконання кваліфікаційної роботи. Вона визначила мету та завдання роботи, розробив метод визначення рівня соціальної прийнятності, реалізував програмний засіб і провів експериментальну оцінку, все це вимагало самостійної роботи та прийняття рішень.

5. Ступінь оволодіння методами дослідження

Під час виконання кваліфікаційної роботи бакалавр продемонстрував високий рівень оволодіння методами дослідження. Він використав методологію семантичного аналізу, вивчив існуючі підходи та впровадив їх у власну розробку, а також провів експерименти для оцінки ефективності розробленого методу.

6. Повнота та якість розкриття теми роботи

Кваліфікаційна робота бакалавра повно і якісно розкриває тему. Усі аспекти методу визначення рівня соціальної прийнятності контенту, включаючи семантичний аналіз, обробку природної мови та програмну реалізацію, були детально описані та проілюстровані прикладами.

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

Кваліфікаційна робота бакалавра характеризується логічним та послідовним викладенням матеріалу. Кожен розділ має чітку структуру і аргументованість, а текстовий матеріал написаний літературно грамотно з урахуванням вимог наукового стилю.

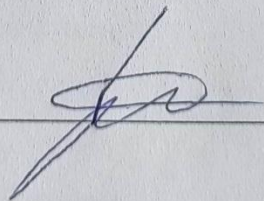
8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин

Розроблений програмний засіб для автоматизованого визначення рівня соціальної прийнятності текстового контенту може мати практичне застосування в сферах, де важлива модерація та фільтрація веб-контенту, таких як соціальні мережі, форуми, коментарі до статей та інші платформи. Він може стати корисним інструментом для виявлення та прибирання неприйняттого або образливого контенту, покращуючи якість взаємодії користувачів з цими платформами.

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

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

Керівник _____



доцент кафедри КН Руслан БАГРІЙ



РЕЦЕНЗІЯ

на кваліфікаційну роботу бакалавра

студента *гр. КНС-20-1 Загуровська Яна Сергіївна*

за темою: Метод визначення рівня соціальної прийнятності текстового контенту дописів за їх семантичним аналізом

1. Актуальність обраної теми

Обрана тема є актуальною, оскільки в останні роки зростає важливість виявлення та фільтрації соціально неприйняттого контенту в онлайн-середовищі. Метод визначення рівня соціальної прийнятності текстового контенту за його семантичним аналізом може допомогти автоматизувати цей процес та поліпшити безпеку та якість взаємодії користувачів.

2. Повнота розкриття мети та завдань роботи

В роботі бакалавра повністю розкрито мету і завдання. Визначено, що метою є розробка методу визначення рівня соціальної прийнятності текстового контенту та створення відповідного програмного забезпечення. Завдання роботи були чітко сформульовані і досягнуті у процесі виконання роботи.

3. Зміст кожного розділу роботи

Кожен розділ роботи містить необхідну інформацію для розуміння та впровадження розробленого методу. Вони охоплюють теоретичний огляд семантичного аналізу тексту, опис методу визначення рівня соціальної прийнятності, розробку програмного засобу та його експериментальну оцінку. Зміст розділів логічно побудований та допомагає зрозуміти процес розробки та його результати.

4. Оцінка розробленої інформаційної системи, її практична цінність

Розроблена інформаційна система для визначення рівня соціальної прийнятності текстового контенту за його семантичним аналізом є значним досягненням. Вона може мати велику практичну цінність, оскільки допомагає автоматизувати процес виявлення соціально неприйняттого контенту та забезпечити безпечну та відповідальну комунікацію в онлайн-середовищі.

5. Якість оформлення кваліфікаційної роботи бакалавра

Кваліфікаційна робота бакалавра має високу якість оформлення. Її структура є логічною та послідовною, а кожен розділ містить необхідні розділи, такі як вступ, огляд літератури, опис методології, експериментальні результати та висновки. Також, текстовий матеріал презентовано зрозуміло та літературно грамотно.

6. Недоліки кваліфікаційної роботи бакалавра

Хоча робота бакалавра в цілому є добре структурованою та виконаною, можна розглянути можливість більш докладного опису експериментальної частини та результатів. Також варто розглянути можливість порівняння розробленого методу з існуючими підходами та навести порівняльні результати ефективності та точності розробленого методу. Але наведене не впливає на одержані результати 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 « Відмінно ».

Рецензент

Григорук І.М. доц. каф. ІІТЗ

