Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет комп'ютерних наук
(повна назва)

Кафедра програмної інженерії
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження методів аналізу емоційного забарвлення коментарів. Підготовка даних.
(тема)

Виконав:
студент 2 курсу, групи ІПЗм-22-6

Бугай Д. Ю.
(прізвище, ініціали)

Спеціальність 121 – Інженерія програмного забезпечення
(код і повна назва спеціальності)

Тип програми освітньо-наукова

Керівник доц. Валенда Н. А.
(посада, прізвище, ініціали)

Допускається до захисту
Зав. кафедри _____ Дудар З. В.
(підпис) (прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет комп'ютерних наук
Кафедра програмної інженерії
Рівень вищої освіти другий (магістерський)
Спеціальність 121 – Інженерія програмного забезпечення
Тип програми освітньо-наукова програма
Освітня програма Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)
«___» _____ 2024 р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Бугаю Дмитру Юрійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи *«Дослідження методів аналізу емоційного забарвлення коментарів. Підготовка даних»*

затверджена наказом по університету від 29.03.2024р. № 250 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 08.06.2024р.

3. Вихідні дані до роботи: *новини, коментарі під новинами, методи обробки великих обсягів інформації, Selenium WebDriver, мова програмування C#, середовище розробки Visual Studio 2022 Community edition*

4. Перелік питань, що потрібно опрацювати в роботі: *мета роботи, аналіз предметної галузі і постановка задачі, методи вирішення проблеми, дослідження реалізації методів збору даних, опис створеного проєкту з деталями реалізації модулів, аналіз отриманих результатів та пропозиції щодо покращення системи.*

КАЛЕНДАРНИЙ ПЛАН

Номер	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Видача завдання	29.04.2024	виконано
2	Аналіз предметної галузі	01.05.2024	виконано
3	Постановка задачі	02.05.2024	виконано
4	Експериментальні дослідження	02.05 – 20.05.24	виконано
5	Аналіз результатів експериментальних досліджень та розробка рекомендацій	20.05 – 22.05.24	виконано
6	Написання та оформлення статті та тез доповіді	20.05 – 23.05.24	виконано
7	Підготовка пояснювальної записки	01.05 – 26.05.24	виконано
8	Підготовка презентації та доповіді	26.05 – 2.05.24	виконано
9	Нормоконтроль	3.06 – 08.06.24	виконано
10	Рецензування	08.06 – 13.06.24	виконано
11	Занесення диплома в електронний архів	15.06.2024	виконано
12	Попередній захист	15.06.2024	виконано
13	Допуск до захисту у зав. кафедри	16.06.2024	виконано

Дата видачі завдання «29» березня 2024 р.

Студент _____ Бугай Д. Ю.
(підпис) (прізвище, ініціали)

Керівник роботи _____ доц. Валенда Н. А.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Робота містить: 60 с., 32 рис., 6 табл., 11 джер.

ВЕБ-СКРЕПІНГ, ЗБІР ДАНИХ, ЕМОЦІЙНЕ ЗАБАРВЛЕННЯ, МЕТОДИ АНАЛІЗУ, НАБІР ДАНИХ, ПІДГОТОВКА ДАНИХ, ТОН ТЕКСТУ, SELENIUM WEB DRIVER, CSV.

Об’єктом дослідження є методи збору коментарів під новинами у соціальних мережах Facebook і Twitter та їх підготовки для подальшого аналізу емоційного забарвлення.

Метою роботи є створення програмної системи для збору, попередньої обробки та підготовки даних для подальшого аналізу емоційного забарвлення.

Методи розробки базуються на таких технологіях, як Selenium Web Driver, C#, .NET 8.

У результаті роботи було досліджено існуючі методи збору та підготовки даних, розроблено програмну систему, яка зчитує коментарі, обробляє їх та зберігає у форматі CSV для подальшого аналізу емоційного забарвлення, проведено демонстрацію роботи системи та проаналізовано отримані результати.

WEB SCRAPING, DATA COLLECTION, EMOTIONAL COLORING, ANALYSIS METHODS, DATA SET, DATA PREPARATION, TEXT TONE, SELENIUM WEB DRIVER, CSV.

The object of the study is the methods of collecting comments under news on Facebook and Twitter and preparing them for further analysis of emotional coloring.

The aim of the study is to create a software system for collecting, pre-processing, and preparing data for further analysis of emotional coloring.

The development methods are based on such technologies as Selenium Web Driver, C#, .NET 8.

As a result of the work, the existing methods of data collection and preparation

were investigated, a software system was developed that reads comments, processes them, and saves them in CSV format for further analysis of emotional coloring, the system was demonstrated, and the results were analyzed.

Заява щодо самостійного виконання кваліфікаційної роботи та можливості її публікації в електронному архіві відкритого доступу ElArKhNURE.

Я, Бугай Дмитро Юрійович, студент групи ІПЗм-22-6, здобувач вищої освіти на другому (магістерському) рівні кафедри «Програмна інженерія», заявляю: моя кваліфікаційна робота на тему «Дослідження методів аналізу емоційного забарвлення коментарів. Алгоритми», що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу ElArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

ЗМІСТ

ВСТУП

У світі, переповненому інформацією, аналіз суспільних настроїв та виявлення емоційних реакцій в середовищі онлайн-спільнот набуває все більшого значення. Такі платформи, як Twitter та Facebook, надають користувачам можливість висловлювати свої думки, реагувати на новини та взаємодіяти один з одним. Таким чином, вивчення емоційної тональності коментарів під публікаціями новин є актуальним напрямком дослідження.

Мета цієї кваліфікаційної роботи – створення програмної системи для збору та попередньої обробки коментарів-реакцій на новини, для подальшого дослідження їх емоційного забарвлення, де основна увага буде зосереджена на виявленні позитивних, негативних та нейтральних настроїв у висловлюваннях користувачів.

Для збору та підготовки даних було обрано Selenium WebDriver – потужний драйвер для автоматизації взаємодії з веб-ресурсами. Цей вибір був зумовлений його новизною та рідкісністю в подібних дослідженнях. Якщо розглядати активну взаємодію з веб-сторінками та збір даних, то він обіцяє бути ефективним інструментом для досягнення цілей дослідження.

Використовуючи Selenium WebDriver, важливо оцінити альтернативні технології та існуючі системи для збору коментарів до новинних статей. Ця оцінка визначить сильні та слабкі сторони обраного підходу, що в кінцевому підсумку обґрунтує його впровадження.

У даній роботі будуть розглянуті можливості емоційного аналізу, визначені потенційні виклики та досліджені перспективи використання отриманих результатів для об'єктивного аналізу реакції громадськості в мережі на сучасні події.

Це дослідження може бути корисним у плані систематичного статистичного збору даних, як позитивних так і негативних на новини в країні. Таким чином можна вести облік дат, які визвали ті чи інші емоційні реакції у суспільстві, за рахунок чого складати карту емоцій населення за місяці, квартали або роки та розуміти тенденції настроїв людей у країні більшою мірою, а саме з точки зору

статистики. Ідея роботи полягає в збереженні даних емоційної реакції людей на новини, для розуміння тенденцій настроїв, а також у відкриттю шляхів для планування, як покращити емоційне самопочуття населення, маючи на руках реальні дані про виражені емоції протягом певного періоду часу, для того, щоб підтримувати емоційну складову суспільства в так скажімо здоровому стані. Про контроль емоціями в статті не йдеться, так як збір реакції людей на новини не передбачає жодного контролювання поведінкою чи реакцією суспільства на обставини.

Ведення статистичних даних, у яких буде відображена реакція суспільства на новими, по суті, не несе повного опису емоційного стану суспільства і не може повністю цілковито передати, але це може слугувати якоюсь певною частиною цілого комплексу заходів, для того щоб мати більше уяви про емоційний настрій людей у різних ситуаціях, що на практиці може слугувати своєрідним індикатором невдоволеності життям або ж навпаки – скажімо, радості.

Кінцевим результатом даного дослідження має бути готова програма для зчитування коментарів людей в соцмережах, а саме їх реакція на новини, що буде зберігатися в певному форматі даних, який міститиме список коментарів. Кількість зчитаних даних можна встановити вручну, яка буде обмежена кількістю наявних коментарів. Потрібно взяти до уваги, що коментар може містити нетекстові символи або зображення, що зменшує кінцеву кількість зібраних даних. Та й уцілому, розробку системи, яка активно буде відслідковувати кожний новий коментар протягом тривалого часу, немає сенсу розглядати, так як це буде недоцільно затратно. До того ж, кількість нових дописів та новин кожного дня все зростає, що робить обсяги даних неймовірно великими. У будь-якому випадку, систему можна налаштувати, за допомогою задання кількісті коментарів, необхідних для аналізу. У результаті це буде файл, формату CSV, який буде готовим для подальшого аналізу іншою системою, для розпізнавання емоцій у тексті. Найімовірніше ця система буде основана на словниках або комбінованому варіанті з машинним навчанням, так як усього необхідного технічного забезпечення, часу та ресурсів недостатньо, для того, щоб натренувати

високоякісну модель, для розпізнавання емоційного забарвлення в тексті, яка буде визначати контекст новини та контекст повідомлення і на основі цього робити висновки та класифікувати до певної категорії. Слід зазначити, що негативні коментарі у свою чергу, так само, як і позитивні, більшою мірою мають певний набір слів та словосполучень, що дає змогу класифікувати текст за словниками емоцій і мати досить правдиві результати емоційного забарвлення. Варто додати також, що емоційний аналіз тексту за допомогою словників або в комбінованому варіанті з машинним навчанням, буде слугувати своєрідним MVP, щоб перевірити роботу та продуктивність даної системи на практиці, так як більшість програмних продуктів та систем починається з мінімально необхідної кількості функцій для роботи, запуску та для тестування першими клієнтами.

Дана робота буде включати в себе частину про збір, підготовку та збереження даних, тобто отримання коментарів під постами новин, для подальшої їх передачі для емоційного аналізу. Метою даного дослідження є розробка системи на основі Selenium, який буде слугувати механізмом, який дозволяє програмно керувати браузером та імітувати дії користувача для збирання необхідних даних з веб-сторінок. Також буде проведено аналіз необхідності бази даних, яка буде найкраще підходити під дану систему, для збереження видобутих результатів. Програмна система буде мати змогу сформувати CSV файл з даними коментарів для подальшого емоційного аналізу. Обрана мова програмування - C#.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Аналіз предметної області

Тема "Дослідження методів аналізу емоційного забарвлення коментарів. Підготовка даних" висвітлює загальний інтерес до розуміння емоційного виміру мовлення в онлайн-середовищі. Безперечно, у сучасному світі кількість спілкування в Інтернеті через соціальні мережі, новинні портали, блоги та інші платформи є колосальною та зростає кожен день. Це створює потребу у вивченні та розумінні того, які емоції висловлюють користувачі у своїх коментарях.

Розуміння емоційного тону тексту є важливим, коли йдеться про розуміння настроїв, ідей та реакції широкої громадськості. Вплив соціальних медіа-платформ на формування громадської думки є особливо важливим. Розпізнаючи емоційний підтекст у реакціях користувачів, стає можливим виявити гарячі теми та оцінити рівень підтримки або опозиції до певних концепцій чи подій [1].

Впровадження аналізу емоційного забарвлення охоплює такі практичні сфери, як маркетингові стратегії, управління репутацією та політичний аналіз. Застосовуючи цей метод, бізнес може оцінити реакцію користувачів на свої пропозиції, що сприяє вдосконаленню маркетингових стратегій та ефективному управлінню репутацією.

Важливість аналізу емоційного забарвлення текстів також визначається його роллю в забезпеченні психологічного благополуччя інтернет-користувачів і кібербезпеки. Виявлення негативних емоцій може слугувати інструментом запобігання кібербулінгу та іншим загрозам психічному здоров'ю. Основні тенденції розвитку технологій та збільшення обчислювальних можливостей відкривають нові можливості для вдосконалення методів аналізу емоційного забарвлення тексту. Такий аналіз стає ключовим елементом для розуміння динаміки онлайн-комунікації та використання отриманої інформації в різних сферах життя. Все це підтверджує велику актуальність теми в сучасному інформаційному суспільстві.

1.1.1 Сфера дослідження

У сфері сучасних досліджень взаємодії онлайн-спільнот важливим завданням є встановлення та вивчення емоційного підтексту, присутнього в коментарях соціальних мереж таких як Twitter або Facebook. Вибір цього напряму дослідження зумовлений вивченням багатогранних елементів, що охоплюють лінгвістичні характеристики, технологічні перешкоди та соціокультурні обставини.

Основна увага приділяється англомовним соціальним медіа-платформам з огляду на їхнє широке використання та вплив на формування глобальної громадської думки. Заглиблення в емоційний підтекст коментарів на цих платформах дозволяє не лише зрозуміти реакцію аудиторії на різноманітні події, але й відкриває перспективи для розробки ефективних методик емоційного аналізу текстів.

1.1.2 Методологія дослідження

Соціальні мережі, такі як Facebook і Twitter, використовують алгоритми аналізу емоцій для автоматичної категоризації та розпізнавання емоційного контенту, що допомагає виявляти тенденції та реакції груп користувачів.

Розвиток методів глибокого навчання відкриває нові можливості для підвищення точності аналізу. Використання передових алгоритмів і нейронних мереж спрощує автоматизацію виявлення емоційної тональності в текстах. У сфері медіа-моніторингу аналіз емоційного забарвлення використовується для відстеження та аналізу реакції громадськості на події, бренди та новини, що стає важливим для управління репутацією та прийняття стратегічних рішень.

Заглиблюючись у цей розділ, стає зрозуміло, що розглядається важливий етап наукового дослідження, а саме методи веб-скрепінгу, які будуть використані в цій роботі для попередньої обробки даних та збору даних про емоційне забарвлення коментарів під новинами для подальшого аналізу. Оскільки об'єктом нашого аналізу є висвітлення емоційної тональності великої кількості коментарів, то необхідно обрати ефективні та відповідні інструменти

для веб-скрепінгу. Опис можливих методів веб-скрепінгу:

- використання HTTP – запитує бібліотеки для взаємодії з веб-ресурсами. Простий, але не забезпечує повного моделювання взаємодії користувача та втручання в динамічний контент;
- використання браузерних бібліотек, таких як Selenium, для емуляції дій користувача в браузері. Забезпечує доступ до динамічних елементів сторінки та взаємодію зі сторінками, що використовують JavaScript;
- використання API веб-ресурсу для ефективного отримання даних без необхідності обробки HTML-коду сторінок;
- використання бібліотек веб-скрепінгу, таких як Beautiful Soup або Scrapy, для аналізу HTML-коду та вилучення даних. Scrapy надає інструменти для структурування веб-колекції;
- використання браузеру в "headless" режимі для отримання даних без відображення графічного інтерфейсу користувача;
- використання бібліотек JavaScript, таких як Splash, для взаємодії з веб-ресурсами, що використовують складні технології, такі як JavaScript.

Вибір того чи іншого методу веб-скрепінгу визначається конкретними вимогами та специфікою досліджуваного веб-ресурсу [2]. Обрані методи взаємодії забезпечують необхідну гнучкість та оперативність для отримання інформації, важливої для подальшого аналізу емоційного контексту коментарів під новинами.

Оскільки коментарі є динамічною інформацією, яка постійно зростає в часі, особливо на піку популярності новини, для програмного рішення необхідно розробити динамічну систему, яка може швидко оновлювати та доповнювати дані новою інформацією.

1.1.3 Аналіз існуючих рішень

Існує кілька ефективних методів і технологій веб-скрепінгу для збору даних з веб-ресурсів. Веб-скрепінг – це процес автоматизованого отримання даних з

інтернет-ресурсів [3]. Було створено порівняльну таблицю 1.1 для Selenium WebDriver та існуючих готових рішень для досягнення цієї мети.

Таблиця 1.1 – Порівняння технологій збору даних

Характеристики	Selenium WebDriver	Beautiful Soup	Scrapy	Requests
Тип інструменту	Автоматизація веб-браузера	Парсер HTML та XML	Фреймворк для веб скрапінгу	HTTP бібліотека у Python
Імітація дій користувача	Так	Ні	Ні	Ні
Виконання JavaScript	Так	Ні	Так	Ні
Швидкість збору даних	Залежить від веб-сторінок	Залежить від обсягу даних	Залежить від структури сайту	Висока, для статичних даних
Зручність у використанні	Зручний, але важливо оптимізувати	Простий та легкий у використанні	Розвинутий, але може бути складним	Простий і легкий для початківців
Адаптованість до сучасних технологій	Так	Ні	Так	Ні
Гнучкість та розширюваність	Так	Обмежена	Так	Ні

У рамках дослідження, присвяченого аналізу емоційного забарвлення коментарів під новинами, вибір інструменту збору даних є надзвичайно важливим кроком, оскільки він впливає на якість та повноту отриманих результатів. Для цього дослідження було обрано Selenium WebDriver, і ось обґрунтування цього вибору:

- у контексті аналізу коментарів швидкість і релевантність є ключовими факторами, Selenium дозволяє імітувати дії користувачів, що важливо для отримання даних в режимі реального часу і взаємодії з динамічним контентом;
- багато сучасних веб-сайтів використовують JavaScript для додавання та

відображення коментарів, Selenium вміє ефективно виконувати JavaScript-код, що робить його чудовим інструментом для роботи з динамічним контентом;

- Selenium має широкий спектр функціональних можливостей і зручний інтерфейс, що дозволяє розробникам легко налаштовувати і розширювати його під конкретні вимоги дослідження, що важливо для адаптації інструменту до специфічних вимог і умов дослідження;
- Selenium адаптований до сучасних технологій та веб-платформ, що забезпечує стабільність у взаємодії з різноманітними веб-сайтами, включаючи соціальні мережі та новинні портали;
- Selenium забезпечує гнучкість у виборі веб-браузера, підтримує різні мови програмування та розширюваність за рахунок плагінів і додаткових бібліотек.

Обираючи Selenium, ставиться завдання забезпечити ефективний та автоматизований збір даних для подальшого отримання об'єктивних результатів дослідження емоційного забарвлення коментарів під новинами.

1.2 Постановка задачі

1.2.1 Мета дослідження

Метою даного дослідження є аналіз емоційного забарвлення коментарів під новинами з метою визначення їх позитивного, негативного чи нейтрального характеру. Основна частина цієї роботи безпосередньо пов'язана з попередньою обробкою та підготовкою даних – коментарів під новинами в соціальних мережах.

1.2.2 Цілі дослідження

У даній роботі будуть розглядатися наступні цілі:

- використання Selenium для автоматизації збору коментарів під новинами з вибраних англомовних соціальних мереж;
- обробка та стандартизація зібраних коментарів для подальшого аналізу;

- консолідація в єдиний формат CSV для полегшення передачі даних на етап емоційної обробки;
- дослідження використання словників, машинного навчання або інших методів для визначення емоційного змісту кожного коментаря;
- перевірка адаптивності розробленої програмної системи до реальних умов;
- аналіз отриманих результатів;
- визначення можливих покращень.
- визначення перспектив розвитку даного дослідження;

Результати та висновки, отримані в ході дослідження, можуть мати практичне значення для різних застосувань, таких як аналіз відгуків клієнтів, маркетингові дослідження, а також безпосередньо в політології.

1.2.3 Очікувані результати

Очікуваний результат та мета цього дослідження – зробити вагомий внесок у покращення аналізу емоційної тональності коментарів у мережевих інформаційних середовищах. Завдяки використанню Selenium WebDriver для збору даних з англомовних соціальних мереж, цей підхід, як очікується, буде ефективним та інноваційним, розширюючи сферу досліджень у цій галузі. Збір даних та правильна їх підготовка буде слугувати важливим фундаментом для подальшого аналізу.

Беручи до уваги, що сучасні технології, штучний інтелект та машинне навчання розвиваються з колосальною швидкістю, використовуючи ці передові інструменти та алгоритми обробки природної мови, аналіз емоційного забарвлення коментарів може досягти виняткової точності. Переконливі дані, які будуть отримані в результаті такого порівняльного аналізу разом у поєднанні з відомими рішеннями, продемонструють надзвичайну ефективність і потенціал цього інноваційного підходу, що гарантовано підтверджує важливість даного дослідження для розвитку галузі вивчення емоційних реакцій населення в мережі Інтернет.

2 МЕТОДИ ВИРІШЕННЯ ПРОБЛЕМИ

2.1 Теорія баз даних

В епоху великих даних важливість ефективного управління та аналізу даних набуває першорядного значення, особливо у сфері наукових досліджень. Це дослідження заглиблюється у складний світ коментарів у соціальних мережах з метою аналізу їхнього емоційного підтексту. Виконання такого завдання вимагає ретельної організації та зберігання величезних обсягів даних, і саме тут у гру вступає теорія баз даних.

Теорія баз даних, фундаментальний розділ інформатики, надає набір принципів, методів і структур для зберігання, обробки та управління даними. В її основі лежить концепція структурованих даних, де інформація чітко організована в таблиці, що дозволяє ефективно її шукати та аналізувати. Визначаючи взаємозв'язки між елементами даних та впроваджуючи заходи цілісності та безпеки, теорія баз даних забезпечує точність та надійність збереженої інформації.

У контексті цього дослідження теорія баз даних відіграє ключову роль у забезпеченні структурованих і легкодоступних даних. Система зберігання даних, розроблена на основі принципів баз даних, дозволяє швидко та ефективно аналізувати емоційні реакції, пов'язані з коментарями. Це полегшує вилучення цінної інформації, наприклад, про загальний настрій щодо певної теми або про емоційні тригери, які найбільше резонують з аудиторією.

Більше того, теорія баз даних виходить за рамки простої організації даних. Вона слугує основою для створення аналітичних звітів та візуалізацій, які необхідні для представлення результатів дослідження у переконливий та зрозумілий спосіб. Використовуючи можливості систем управління базами даних, дослідники можуть створювати інтерактивні інформаційні панелі, діаграми та графіки, які ілюструють емоційне забарвлення коментарів, дозволяючи зацікавленим сторонам глибше зрозуміти дані.

Таким чином, теорія баз даних – це не просто інструмент для організації даних у цьому дослідженні; це незамінна основа, яка лежить в основі всього дослідницького процесу. Вона забезпечує цілісність, доступність і зручність

використання даних, сприяє ефективному аналізу та дозволяє створювати змістовні звіти і візуалізації. Використовуючи можливості теорії баз даних, дослідники можуть розкрити весь потенціал даних соціальних мереж, отримуючи цінну інформацію про емоційне сприйняття та настрої онлайн-спільнот.

2.2 Продуктивні обчислення та перетворення даних

Ефективність бази даних можна оцінити за допомогою більш складних математичних розрахунків і формул. Одна з таких формул включає поняття інформаційної ентропії, яка вимірює невизначеність, пов'язану з певним набором даних. Рівняння (формула 2.1) показує обчислення інформаційної ентропії [4].

$$H(x) = -\sum p(x)\, log_2\big(p(x)\big) \tag{2.1}$$

де: H(x) – інформаційна ентропія набору даних x;

p(x) – ймовірність появи кожного значення даних x.

Ефективність бази даних можна визначити як відношення інформаційної ентропії набору даних до кількості операцій, необхідних для обробки даних. Рівняння (формула 2.2) показує обчислення ефективності бази даних.

$$E = \frac{H(X)}{O} \tag{2.2}$$

де: E – ефективність бази даних

H(X) – інформаційна ентропія набору даних X

O – кількість операцій, необхідних для обробки даних

Ця формула дозволяє більш тонко оцінити ефективність бази даних, оскільки враховує невизначеність, пов'язану з даними. На практиці її можна використовувати для порівняння різних проектів або конфігурацій баз даних, а також для виявлення можливостей для вдосконалення.

Розглянемо конкретний приклад, щоб проілюструвати розрахунок ефективності бази даних за допомогою інформаційної ентропії. Припустимо, що у

нас є база даних з 10 000 коментарів, і кожен коментар можна віднести до однієї з п'яти емоційних категорій: позитивний, негативний, нейтральний, змішаний або нерелевантний. Ймовірність появи кожної категорії наступна:

- позитивний: 0.2;
- негативний: 0.1;
- нейтральний: 0.5;
- змішаний: 0.1;
- ірелевантний: 0.1.

Рівняння (формула 2.3) показує приклад обчислення інформаційної ентропії цього набору.

$$H(X) = -(0.2\,log_2(0.2) + 0.1\,log_2(0.1) + 0.5\,log_2(0.5) + 0.1\,log_2(0.1) + 0.1\,log_2(0.1)) = 1.92\ bit \quad (2.3)$$

Припустимо далі, що база даних вимагає 5 операцій для обробки кожного коментаря, таких як отримання коментаря зі сховища, аналіз його емоційного забарвлення та оновлення бази даних. Рівняння (формула 2.4) показує приклад обчислення ефективності бази даних.

$$E = \frac{H(X)}{O} = \frac{1.92\ bit}{5\ \text{операцій}} = 0.384\ bit\ (\text{за одну операцію}) \quad (2.4)$$

Цей результат вказує на те, що база даних може обробити кожен коментар із середньою ефективністю 0,384 біт за операцію. Даний приклад обчислення може бути використаний для оцінки продуктивності бази даних у майбутньому та визначення напрямків для покращення.

Вибір бази даних для зберігання зібраних та оброблених коментарів, залежить від конкретних вимог і міркувань. Для цього випадку використання, коли пріоритетними є економічність і швидкість, а дані не потребують складних обчислень у майбутньому, може бути корисним порівняння декількох найпоширеніших БД, які використовуються найчастіше, наведене в таблиці 2.1.

Таблиця 2.1 – Порівняння найпоширеніших СУБД

СУБД	Вартість	Швидкість	Придатність
MySQL	З відкритим вихідним кодом і безкоштовна у використанні	Висока, особливо для зчитування	Для ефективного зберігання та пошуку великих обсягів коментарів
PostgreSQL	З відкритим вихідним кодом і безкоштовна у використанні	Порівнянна з MySQL, з додатковими функціями для складних запитів	Для простих і складних потреб зберігання та аналізу даних
MongoDB	Безкоштовна версія для спільноти та платна корпоративна	Висока масштабованість та швидкість запису	Для зберігання та обробки неструктурованих даних
Redis	З відкритим вихідним кодом і безкоштовна у використанні	Надвисока швидкість зберігання та пошуку даних у пам'яті	Для кешування часто використовуваних даних або зберігання тимчасових даних
SQLite	З відкритим вихідним кодом і безкоштовна у використанні	Підходить для малих та середніх наборів даних	Як вбудована БД для локального зберігання даних

У контексті використання інструменту Selenium для веб-скрепінгу коментарів, вибір СУБД пріоритетно буде спиратися на швидкість зберігання даних. Виходячи з наведеного вище порівняння, MySQL і PostgreSQL є потенційними варіантами, бо пропонують поєднання економічної ефективності, швидкої продуктивності та масштабованості. Обидві СУБД мають обширну документацію, велику спільноту користувачів та широкий спектр доступних інструментів і бібліотек, що полегшить їх інтеграцію з Selenium WebDriver та іншими інструментами веб-скрейпінгу.

2.3 Доцільність використання баз даних

У контексті проекту мінімального життєздатного продукту (MVP) використання файлу у форматі CSV для зберігання даних має кілька переваг порівняно з традиційною базою даних. Переваги та їх опис наведено в таблиці 2.2.

Таблиця 2.2 – Переваги зберігання даних у форматі CSV

Переваги	Опис
Простота	Прості текстові файли з легкою структурою, що дозволяє швидко налаштовувати і працювати з ними без необхідності складної конфігурації або адміністрування бази даних.
Переносимість	Легко переміщуються між різними системами і платформами, що особливо корисно в невеликих проєктах, де основна увага приділяється тестуванню і перевірці основних функцій і можливостей продукту.
Сумісність з іншими системами	Підтримуються широким спектром програм, включаючи електронні таблиці, текстові редактори та інструменти для аналізу даних.
Масштабованість	Можуть забезпечити адекватне зберігання даних для MVP-проекту. У міру розвитку продукту й виникнення потреби в більш надійному управлінні даними, міграція з CSV-файлів до бази даних становиться необхідністю.
Економічна ефективність	Пропонують недорогу і легку альтернативу, у порівнянні зі створенням і підтримкою бази даних.

Важливо зазначити, що файли CSV мають певні обмеження, такі як обмежена кількість типів даних, відсутність підтримки складних запитів і потенційні проблеми з масштабуванням для великих наборів даних. З розвитком проєкту та зростанням обсягу інформації перехід до бази даних може стати необхідним. Однак для MVP-проєкту, де основна увага приділяється демонстрації основної функціональності, ці обмеження можуть бути переважені перевагами простоти, портативності, сумісності, масштабованості та економічної ефективності.

2.4 Методи підготовки даних для розпізнавання тексту

Методи підготовки даних для аналізу емоційних коментарів є ключовим кроком у роботі з текстовою інформацією. Токенізація, видалення надлишкової інформації, перетворення тексту в нижній регістр, лематизація та стеммінг є основними елементами обробки текстових даних [5]. Тому нижче наведено більш детальний опис кожного з методів підготовки даних:

Токенізація – важливий етап підготовки даних, спрямований на розбиття тексту на токени (слова або фрази). У контексті аналізу емоційних коментарів

кожен токен представляє окремий елемент тексту. Цей процес полегшує подальший аналіз текстової інформації та виділення значущих одиниць для аналізу емоційного забарвлення.

Видалення зайвої інформації – для забезпечення точності аналізу емоцій важливо видалити зайву інформацію, таку як розділові знаки, числові дані та інші символи, які не несуть значного смислового навантаження. Це покращує якість текстового матеріалу і дозволяє уникнути спотворень під час аналізу.

Стандартизація регістру – важливий крок в обробці тексту. Перетворення всіх символів у малі літери дозволяє уніфікувати дані, уникнути дублікатів через різні регістри та полегшити подальший аналіз.

Лематизація та стеммінг спрямовані на зменшення розмірності даних шляхом перетворення слів до їхніх базових або кореневих форм. Це зменшує варіативність слів зі збереженням семантичної інформації, що полегшує подальший аналіз і забезпечує більшу швидкість обробки.

Видалення стоп-слів – крок, спрямований на фільтрацію загальних і неважливих слів, які не вносять істотного внеску в контекст емоційного забарвлення. Це підвищує точність і робить ключові слова більш помітними для аналізу.

Видалення емодзі та спеціальних символів є важливим кроком, оскільки ці елементи можуть впливати на аналіз і класифікацію емоцій. Цей процес допомагає уникнути шуму в даних і підвищити точність аналізу емоцій.

Таким чином, з'являється перелік операцій, які потрібно буде виконати над даними для їх попередньої обробки. Хоча система аналізу емоційного тонусу може самостійно фільтрувати та обробляти дані, виникає проблема зберігання великих обсягів даних у базі даних. У свою чергу, це вимагає оптимізації зберігання, що призводить до вищезгаданих методів підготовки.

Усі ці етапи підготовки даних можуть бути виконані на стороні як системи, яка зчитує коментарі під новинами, так і на стороні системи, яка вже аналізує емоційне забарвлення зчитаних даних. Це дає більшої гнучкості у використанні та робить механізм підготовки даних гнучкішим до налаштувань. Також це можна

використати для перерозподілу навантаження на будь-яку із систем, та вказати яка саме із них буде трансформувати набір даних для подальшого аналізу. Розглядаючи сторону першої системи для зчитування коментарів, було б необхідним реалізувати базову систему очистки тексту, яка б форматувала непотрібні символи, емодзі та знаки пунктуації. А лематизацію, стеммінг, видалення стоп-слів тощо, покласти на плечі другої системи для визначення тональності. Таким чином бізнес-логіка кожного із інструментів буде унікальною та виконуватиме тільки свої безпосередні функції.

2.5 WebDriver та його застосування для веб-скрепінгу

Selenium WebDriver – це інструмент автоматизації веб-браузера, який можна використовувати для збору даних з веб-сайтів. Цей драйвер в першу чергу використовується як інструмент для автоматизованого тестування веб-додатків і виконання інших рутинних завдань, пов'язаних з роботою в Інтернеті. Цей інструмент дозволяє програмно взаємодіяти з веб-сторінками, виконуючи дії, подібні до дій користувача Інтернету, переходи за посиланнями, відкриття нових вікон, отримання вмісту сторінки та багато інших [6].

Цей драйвер є потужним інструментом для автоматизації веб-застосунків. Він надає широкий інтерфейс для взаємодії з веб-браузерами і автоматизації ряду завдань на веб-сторінках. Нижче наведено основні елементи Selenium драйверу.

WebDriver – основний інтерфейс для взаємодії з різними браузерами. Він дозволяє виконувати різні дії, такі як відкриття сторінок, введення тексту, натискання на елементи і багато іншого.

WebElement – цей інтерфейс представляє окремий елемент на веб-сторінці. За допомогою WebElement можна отримати інформацію про елемент, взаємодіяти з ним і виконувати такі дії, як введення тексту або клацання.

By – клас By визначає методи пошуку елементів на сторінці за різними критеріями, такими як ID, клас, тег тощо.

ExpectedConditions – цей клас визначає різні умови, які можуть бути використані в поєднанні з методами очікування для забезпечення стабільного

виконання дій до завантаження сторінки або настання певної події.

Використання цих компонентів у поєднанні з мовою програмування C#, дозволяє створювати скрипти для навігації, взаємодії та зчитування даних з веб-сторінок. Бібліотека Selenium складається з багатьох методів, які надають широкі можливості для автоматизації взаємодії з веб-браузерами. Ключовими для роботи з веб-скрепінгом та збором даних є WebDriver та WebElement, які надають необхідний функціонал для взаємодії з HTML-документами:

- get(url) – завантажити вказану веб-сторінку;
- find_element(by, value), щоб знайти перший елемент за вказаним критерієм;
- find_elements(by, value), для пошуку всіх елементів за вказаним критерієм;
- click(), щоб клацнути вибраний елемент;
- send_keys(keys), для введення тексту або комбінацій клавіш;
- text, щоб отримати текстовий вміст елемента;
- get_attribute(name), щоб отримати значення атрибуту елемента;
- is_displayed(), щоб перевірити видимість елемента;
- clear(), щоб очистити вміст елемента (для полів введення).

Ці методи є одними з фундаментальних для взаємодії з веб-сторінками та їх елементами.

3 ТЕОРІЯ ЕМОЦІЙНОГО АНАЛІЗУ КОМЕНТАРІВ

3.1 Основні підходи до емоційного аналізу текстів

Основні підходи до емоційного аналізу тексту включають використання методів машинного навчання та обробки природної мови для розпізнавання та класифікації емоційного забарвлення текстового матеріалу. Узагалі, одним із найсучасніших методів вирішення різних задач обробки природних мов є використання попередньо тренованих моделей та нейронних мереж з архітектурою «трансформер», найвідоміший приклад якого є BERT (Bidirectional Encoder Representations from Transformers) від Google. Цей інструмент дозволяє оброблювати велику кількість даних, на основі якісної моделі, яку можна використовувати для різноманітних задач з мінімальними змінами [7].

BERT від Google, незважаючи на свою ефективність у вирішенні завдань обробки природної мови, має низку обмежень. По-перше, модель вимагає значних обчислювальних ресурсів, що робить її навчання та використання дорогим процесом. По-друге, архітектура BERT є складною для розуміння та налаштування, що вимагає від розробників глибоких знань у галузі машинного навчання. По-третє, процес навчання та налаштування BERT є часозатратним, що може уповільнити впровадження моделі у реальні проекти, тому у цьому дослідженні буде розглянуто кілька ключових загальних підходів, спрямованих на емоційний аналіз коментарів під новинами, не вдаючись в конкретно реалізовані моделі.

Методи машинного навчання: використання алгоритмів машинного навчання, таких як класифікатори (наприклад, машини опорних векторів, наївний Байєс) для навчання моделей на маркованих емоційних даних. Моделі можна навчити розпізнавати позитивне, негативне та нейтральне забарвлення текстів.

Аналіз лексики та лексичних особливостей: врахування лексики та лексичних особливостей текстів для визначення емоційної тональності. Використання словників емоційних слів та їх вагових коефіцієнтів для оцінки емоційної інтенсивності.

Глибоке навчання: використання глибоких нейронних мереж для емоційного аналізу, що дозволяє враховувати складні взаємозв'язки та контекстність мови.

Глибинне навчання може допомогти автоматизувати виявлення відмінностей у використанні мови для різних емоцій.

Безперервне навчання: використання методів для постійного навчання моделей, що дозволяє їм адаптуватися до контекстів, які можуть часто змінюватися, та еволюції мови в Інтернеті.

У наступній таблиці 3.1 узагальнено різні методи в контексті їхніх переваг та обмежень.

Таблиця 3.1 – Переваги та недоліки підходів до емоційного аналізу текстів

Метод	Переваги	Недоліки
Методи машинного навчання	Висока точність і гнучкість	Потреба в маркованих даних. Залежність від якості даних
Аналіз словника та лексичних особливостей	Простота, додатковий контекст	Не врахування контексту. Залежність від словників
Глибоке навчання	Автоматичне виявлення особливостей, адаптивність	Вимоги до обчислювальних ресурсів. Потреба у великомасштабних даних
Безперервне навчання	Адаптивність, оптимізація ефективності	Потреба в реальних даних. Контрольованість процесу

У наступній таблиці систематизовано інформацію про різні методи в контексті їхніх переваг. Можна розглянути аналіз словникових та лексичних особливостей як потенційний підхід до визначення емоційного забарвлення коментарів. Відмінності в емоційному забарвленні тексту можна виявити за допомогою аналізу елементів лексики та лексичних особливостей, а використання словникових ресурсів та врахування мовної специфіки сприяє точності визначення тональності тексту.

Обрання такого підходу дозволяє отримати швидкі результати, водночас забезпечуючи можливість адаптації до особливостей аналізованих даних. Важливо враховувати обмеження методу, такі як залежність від якості використовуваних словників і необхідність ретельного врахування мовної специфіки в коментарях.

Цей підхід, зважаючи на його переваги та обмеження, може виявитися

ефективним у контексті визначення емоційного забарвлення коментарів під новинами за обраною темою дослідження. Незважаючи на це, можна також розглядати комбінацію машинного навчання зі словниками, що дозволить подвоїти сильні сторони обох, та мінімізувати недоліки [8].

3.2. Обґрунтування можливостей використання гібридного підходу

Аналіз лексичних і словникових особливостей лежить в основі словникового методу, який є добре відомим підходом в аналізі емоцій. Цей метод може виявити тонкі варіації емоційного тону, шляхом ретельного вивчення підбору та контекстуального використання слів і фраз. Він також відносно простий у застосуванні, що робить його практичним для цього дослідницького проекту.

Однак, незважаючи на свою потужність, ці методи, мають певні обмеження, пов'язані з обмеженим обсягом та якістю використовуваних словників. Щоб подолати ці обмеження і досягти більш надійної та точної ідентифікації емоційного забарвлення, в цьому дослідженні розглядається можливість інтеграції методів машинного навчання. Алгоритми машинного навчання можуть вивчати складні патерни з великих наборів даних, що дозволяє їм адаптуватися до конкретної мови та контексту аналізованих коментарів. Така інтеграція ефективно усуває недоліки підходів, що ґрунтуються виключно на словниках.

Застосування гібридного підходу має кілька переваг. По-перше, він поєднує сильні сторони обох складових методів, використовуючи інтерпретованість і мовно-специфічне розуміння глосаріїв з гнучкістю та адаптивністю машинного навчання. По-друге, він має потенціал для швидшого отримання результатів порівняно з методами, що ґрунтуються виключно на MachineLearning, оскільки може використовувати наявні лінгвістичні знання. Нарешті, він може бути адаптований до унікальних нюансів аналізованих даних, забезпечуючи вищу точність у визначенні емоційного забарвлення.

Однак застосування цього гібридного підходу супроводжується певними викликами. Ефективність методу безпосередньо залежить від якості та повноти використовуваних словників. Крім того, необхідно ретельно враховувати тонкощі

мови, такі як сарказм, іронія та культурний контекст, щоб уникнути неправильних інтерпретацій, тощо.

Незважаючи на ці виклики, гібридний підхід є перспективним напрямком для аналізу настроїв, особливо в контексті коментарів до новин. Майбутні дослідження можуть розглянути шляхи оптимальної інтеграції методів ML і словникового підходу. Це може включати розробку спеціалізованих словників, пристосованих до конкретних тем, і вдосконалення алгоритмів для кращого врахування складнощів мови [9].

3.3 Сценарії для аналізу емоційної тональності текстів

Що стосується теми цього дослідження, то для того, щоб точно та ефективно визначити емоційне забарвлення коментарів, розглядаються два основні підходи: використання емоційних словників та впровадження методів машинного навчання. Кожен з цих підходів має свої переваги та обмеження, і вибір між ними залежить від конкретних завдань і параметрів дослідження. Можливий варіант також часткової комбінації даних підходів, щоб використати переваги обох та мінімізувати недоліки. Наступна таблиця 3.2 наводить порівняльний аналіз емоційних словників та алгоритмів машинного навчання для емоційного аналізу коментарів

Таблиця 3.2 – Порівняльний аналіз емоційних словників та алгоритмів машинного навчання для емоційного аналізу коментарів.

Параметер	Словники емоцій	Машинне навчання
Точність	Залежить від якості словника та анотацій	Залежить від обсягу та якості навчальних даних, здатна досягати високого рівня точності
Контекст	Не враховує контекстні міркування	Здатна враховувати контекст і виражати складні взаємозалежності між словами
Вартість розробки	Як правило, дешевше у розробці	Потребує значних обчислювальних ресурсів і навчальних даних
Можливість прогнозування	Менш схильні адаптуватися до нових даних	Здатні адаптуватися до нових контекстів та оновлень даних

Це порівняння представляє всебічний огляд обох методологій з метою визначення найбільш придатної для конкретного завдання емоційного аналізу коментарів. Емоційні словники містять лексику слів, що представляють спектр емоцій, від позитивних до негативних.

У цій роботі буде можемо припустити, що обраний спосіб для другої системи буде містити використання словників для визначення емоційної тональності коментарів, завдяки їхнім перевагам. Використання готових емоційних словників дозволяє швидко визначати емоційне забарвлення текстів без необхідності тривалого машинного навчання. Емоційні словники детально визначають різні відтінки емоцій, що дозволяє точно визначити емоційний тон тексту. Готові словники можна легко інтегрувати в дослідження, не докладаючи особливих зусиль для їх використання. Їх використання не потребує великих витрат коштів і часу порівняно з іншими методами аналізу емоційного забарвлення. Готові словники з емоціями та текстовими прикладами можуть бути легко знайдені в мережі Інтернет, що економить час і ресурси на їхню розробку. Звісно такі словники можуть бути суто для навчання та тренування невеликих систем розпізнавання, так як здебільшого найякісніші дані у вільному доступі не перебувають. Також не потрібно забувати про збалансованість словників, а саме акцент ставиться на рівній кількості прикладів тексту для відповідних емоцій. Отже, ця методологія дозволяє швидко та ефективно аналізувати емоційне забарвлення в дослідженнях, що дає сприятливий результат. Вона забезпечує об'єктивний підхід до вивчення психологічного стану аудиторії через призму її емоційної реакції на новини.

4 ДОСЛІДЖЕННЯ РЕАЛІЗАЦІЇ МЕТОДІВ ЗБОРУ ДАНИХ. ПІДГОТОВКИ ДО АНАЛІЗУ

4.1 Дослідження реалізації збору даних

Для того, щоб розпочати розробку програми з використанням Selenium WebDriver та C# з метою вилучення коментарів з відомих новинних статей у соціальних мережах для подальшого аналізу настроїв користувачів, необхідно виконати кілька важливих кроків.

По-перше, необхідно створити робоче середовище, встановивши Visual Studio, останню версію .NET та пакет Selenium WebDriver NuGet пакет. Після успішного налаштування середовища можна створити новий проєкт на мові програмування C# та імпортувати бібліотеку Selenium WebDriver.

Згодом необхідно визначити веб-сторінки, що містять потрібні коментарі. Це можна зробити шляхом ручного переходу до новинних постів на цільових платформах соціальних мереж і вивчення HTML-структури веб-сторінок.

Після того, як елементи HTML, що містять коментарі, ідентифіковані, Selenium WebDriver може бути використаний для імітації користувацьких дій, таких як прокрутка веб-сторінки, для поступового вилучення потрібних даних.

Дотримуючись цих кроків, Selenium WebDriver буде ефективним інструментом для зчитування коментарів під новинами у соціальних мережах (Facebook та Twitter). Зчитані дані оброблюватимуться під CSV формат, для подальшого їх аналізу емоційного забарвлення, яке було висловлене користувачами, тим самим надаючи важливу інформацію про громадську думку і суспільні настрої [10].

4.2 Опис створеного проєкту

Розроблений проєкт для веб-скрепінгу використовує багаторівневу архітектуру, що складається з двох основних модулів: SeleniumScraper та WebScrapingApp. Такий дизайн сприяє чіткому розподілу завдань, сприяючи модульності та зручності обслуговування.

SeleniumScraper втілює в собі основний движок скрапінгу, який використовує

Selenium WebDriver для автоматизації взаємодії з браузером та вилучення даних з цільових веб-сайтів. Він структурований на окремі компоненти: скрипти для автоматизації браузерів, методи вилучення даних, утиліти для загальних завдань і конфігураційні файли. Така модульна структура дозволяє розробляти і тестувати кожен компонент окремо, що підвищує гнучкість і адаптивність проекту.

WebScrapingApp слугує в якості контролюючого шару. Він організовує процес скрапінгу, ініціюючи завдання, керуючи їх плануванням та координуючись з SeleniumScraper для делегування завдань та отримання результатів. Механізми обробки помилок забезпечують надійність, а можливості ведення журналів дають уявлення про продуктивність програми. Дизайн WebScrapingApp дозволяє паралельне або послідовне виконання декількох завдань, що сприяє масштабованості.

Взаємодія між цими модулями визначається чітким розподілом обов'язків. SeleniumScraper обробляє низькорівневі деталі автоматизації браузера та вилучення даних, тоді як WebScrapingApp зосереджується на високорівневому управлінні робочим процесом, плануванні завдань та агрегації результатів. Таке розділення сприяє гнучкості та дозволяє вносити незалежні зміни або вдосконалення в окремі модулі, не впливаючи на загальну архітектуру. Багаторівнева архітектура разом із модульним дизайном і можливостями масштабування сприяє створенню надійного, зручного в обслуговуванні та адаптованого рішення для веб-скрепінгу.

Розглянутий проєкт зі зчитування даних з веб-сторінок – це структурований і модульний проєкт, що складається з двох окремих компонентів: SeleniumScraper та WebScrapingApp. Такий дизайн забезпечує чіткий розподіл завдань, сприяючи модульності, легкості в обслуговуванні та масштабованості.

4.2.1 Деталі реалізації модулю SeleniumScraper

SeleniumScraper слугує ключовим інструментом для вилучення даних. Він використовує можливості Selenium WebDriver для автоматизації взаємодії з веб-браузерами та вилучення релевантної інформації з цільових веб-сайтів.

SeleniumScraper складається з окремих спеціалізованих компонентів:

- скрипти автоматизації браузерів: Ці скрипти імітують дії користувачів у веб-браузерах, включаючи такі завдання, як натискання на елементи, заповнення форм і прокрутка сторінок;
- методи вилучення даних: Ці методи ретельно розроблені для вилучення конкретних даних з HTML-контенту веб-сторінок. Вони використовують різні методи, включаючи регулярні вирази, селектори CSS і XPath;
- службові функції: Утиліти надають загальні функціональні можливості, які можна легко повторно використовувати в різних скриптах і компонентах. Приклади таких функцій включають обробку помилок, розбір дат і перетворення форматів даних;
- конфігураційні файли: Файли конфігурації містять налаштування та параметри, які є ключовими для роботи SeleniumScraper. Вони можуть включати налаштування браузера, таймаути та конфігурації проксі-серверів.

Модульна парадигма проектування SeleniumScraper полегшує незалежну розробку і тестування кожного складового компонента. Такий підхід сприяє швидкій ітерації, розширює можливості повторного використання коду та спрощує завдання технічного обслуговування.

4.2.2 Деталі реалізації модулю WebScrapingApp

WebScrapingApp, як загальний рівень управління, точно організовує процес вилучення даних. Він ініціює завдання, керує їх плануванням і делегує їх SeleniumScraper. Крім того, WebScrapingApp займається обробкою помилок і надає можливості ведення журналів, які є ключовими для моніторингу продуктивності. Ключові особливості WebScrapingApp наступні:

- управління завданнями: WebScrapingApp вміло керує чергою завдань вилучення даних, розставляючи пріоритети і плануючи їх виконання на основі різних параметрів, таких як складність завдання, своєчасність і доступність ресурсів;

- делегування завдань: WebScrapingApp делегує завдання вилучення даних SeleniumScraper і старанно відстежує їх прогрес. Після завершення завдання WebScrapingApp отримує результати і консолідує їх;
- обробка помилок: Надійні механізми обробки помилок гарантують, що процес вилучення даних продовжиться навіть у разі виникнення непередбачуваних помилок. Повідомлення про помилки та трасування стеку ретельно фіксуються та реєструються для ретельного аналізу;
- логування та моніторинг: WebScrapingApp пропонує комплексні можливості ведення журналу для ретельного відстеження продуктивності та поведінки процесу вилучення даних. Ця інформація є безцінною для налагодження проблем, оптимізації продуктивності та захисту загального стану системи.

Взаємодія між SeleniumScraper і WebScrapingApp чітко визначена і дотримується суворого розподілу обов'язків. SeleniumScraper вправно обробляє складні деталі автоматизації браузера та вилучення даних, в той час як WebScrapingApp зосереджується на високорівневому управлінні робочим процесом, плануванні завдань та агрегації результатів. Таке розділення сприяє гнучкості та дозволяє вносити незалежні зміни або вдосконалення в окремі модулі без шкоди для загальної архітектурної цілісності.

Багаторівнева архітектура, модульний дизайн і можливості масштабування об'єднуються для створення надійного, зручного в обслуговуванні та легко адаптованого рішення для веб-скрепінгу. SeleniumScraper та WebScrappingApp створюють міцну основу для побудови масштабованих конвеєрів веб-скрепінгу та вилучення цінних даних з безкрайніх просторів Інтернету.

4.2.3 Демонстрація Use-case сценаріїв

На рисунку 4.1 зображено побудову Use-case діаграми для програмної системи "Дослідження методів аналізу емоційного забарвлення коментарів. Підготовка даних":

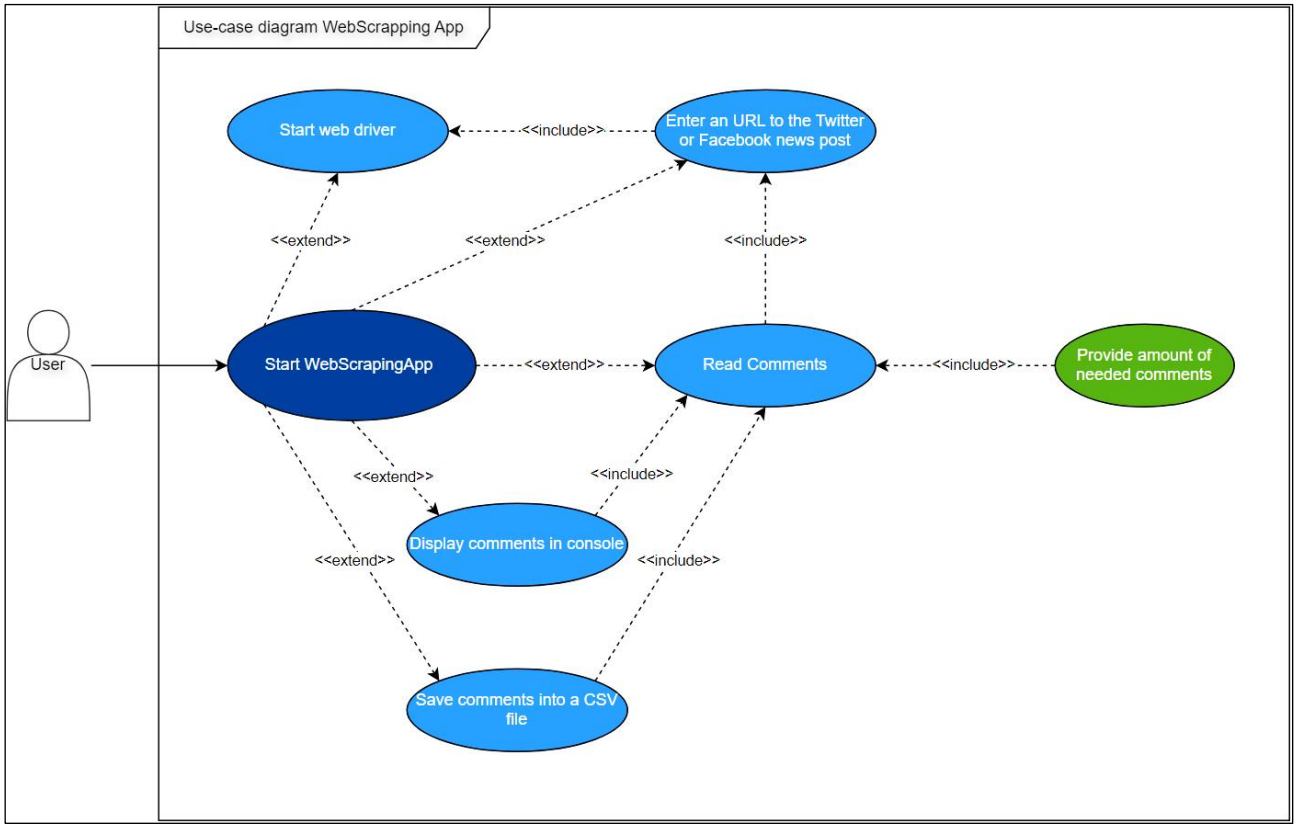


Рисунок 4.1 – Use-case діаграма програмної системи

Проєкт веб-скрепінгу, як показано на діаграмі варіантів використання, окреслює орієнтований на користувача процес вилучення коментарів з платформ соціальних мереж. Основний актор, "Користувач", ініціює операцію вилучення шляхом запуску вікна консолі WebScrapingApp, яка слугує центральним інтерфейсом для управління процесом вилучення. WebScrapingApp, в свою чергу, сприяє автоматизації взаємодії з веб-браузером за допомогою веб-драйвера, який, ймовірно, використовує фреймворк Selenium.

Користувач відіграє вирішальну роль у визначенні цілі вилучення, вводячи в інтерфейс програми URL-адресу новинного повідомлення в Twitter або Facebook. Згодом користувач вказує бажану кількість коментарів, які потрібно витягти, натиснувши кнопку "Прочитати коментарі".

Після цього WebScrapingApp бере на себе контроль над робочим процесом вилучення. Вона взаємодіє з веб-драйвером для переходу на вказану URL-адресу і програмно натискає на відповідні елементи, щоб відкрити розділ з коментарями. Потім програма аналізує HTML-структуру сторінки, щоб витягти вказану кількість

коментарів. Витягнуті коментарі потім представляються користувачеві у вікні консолі, що дозволяє негайно перевірити і проаналізувати їх. Нарешті, WebScrapingApp зберігає зібрані коментарі у файлі CSV (Comma-Separated Values), що полегшує їх зберігання в автономному режимі і подальшу обробку.

Успішне виконання цього процесу скрепінгу залежить від виконання певних умов. Користувач повинен встановити WebScrapingApp на своїй системі і мати дійсні облікові дані для доступу до цільового облікового запису Twitter або Facebook. Крім того, користувач повинен призначити CSV-файл як сховище для вилучених коментарів. Після успішного завершення процесу вилучення, витягнуті коментарі не тільки відображаються в консолі, але й зберігаються у вказаному CSV-файлі, забезпечуючи їх доступність для подальшого використання та аналізу.

4.3 Демонстрація реалізованого програмної системи

У цьому розділі буде представлено детальну ілюстрацію практичного застосування реалізованої програмної системи WebScrapingApp та буде проілюстровано, як WebScrapingApp успішно застосовується в реальних сценаріях для зчитування коментарів під постами новин, на прикладі соціальних мереж Twitter та Facebook.

4.3.1 Запуск застосунку та відкриття консолі

Першим кроком у процесі є запуск програми і спостереження за консоллю. Для початку необхідно запустити застосунок WebScrapingApp. Після запуску з'явиться інтерфейс у вигляді консолі, яка містить список команд, для керування програмою (див.рис.4.2).

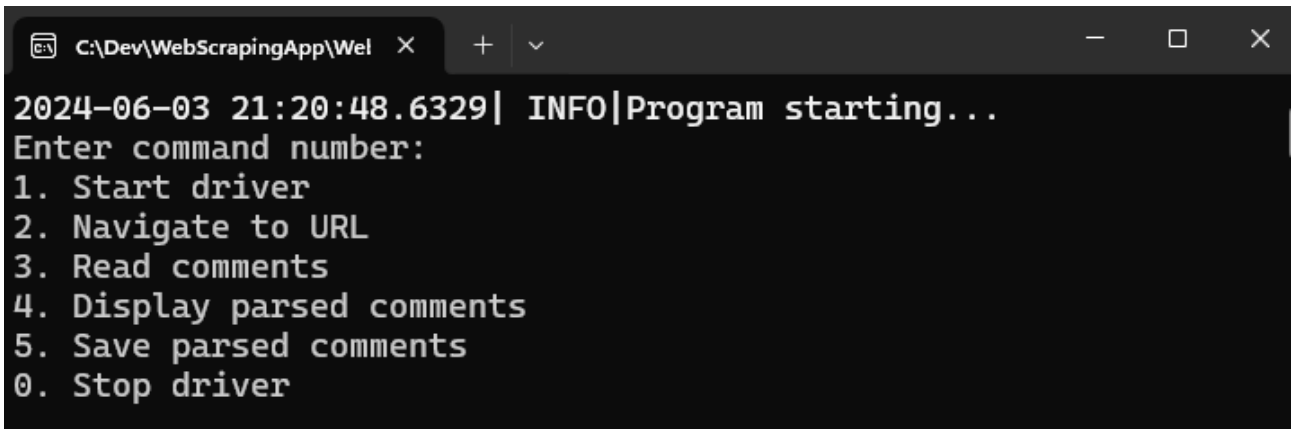


Рисунок 4.2 – Вікно консолі після старту застосунку зі списком команд

4.3.2 Відображення списку команд

Консоль надає вичерпний перелік команд, які дозволяють користувачам ефективно взаємодіяти з системою. Нижче у таблиці 4.1 наведено детальне пояснення кожної команди.

Таблиця 4.1 – Опис команд консольного застосунку

№	Команда	Опис
1	Запуск драйверу	Ініціює виконання веб-драйвера, встановлює з'єднання між застосунком і веб-браузером.
2	Перехід за URL-адресою	Вказує веб-драйверу перейти за вказаною URL-адресою та відкриває відповідну веб-сторінку в браузері.
3	Зчитування коментарів	Зчитує коментарі із вказаної веб-сторінки.
4	Відображення зчитаних даних	Показує зчитані дані у структурованому вигляді в консолі.
5	Збереження даних	Зберігає дані у форматі CSV та відображає шлях до новоствореного файлу.
6	Зупинка драйвера	Завершує роботу веб-драйвера та розриває зв'язок між консоллю і веб-браузером.

4.3.3 Запуск драйвера та відкриття браузера

Наступний крок передбачає запуск веб-драйвера і запуск веб-браузера. Для початку в консолі необхідно ввести перший номер команди (рис. 4.3). Ця дія запускає веб-драйвер, в результаті чого автоматично відкривається браузер (рис. 4.4).

Веб-драйвер виступає сполучною ланкою між скриптами і реальним браузером, полегшуючи взаємодію між ними. Після запуску браузера відкривається вкладка за замовчуванням.

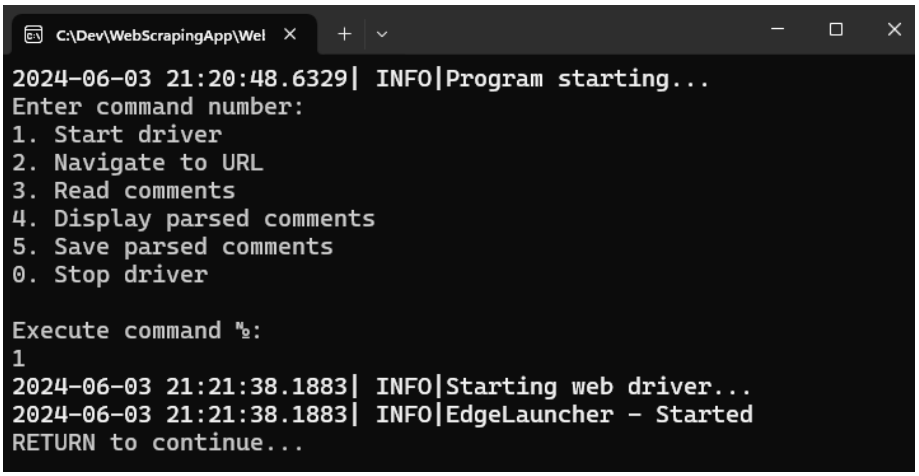


Рисунок 4.3 – Введення команди запуску драйвера

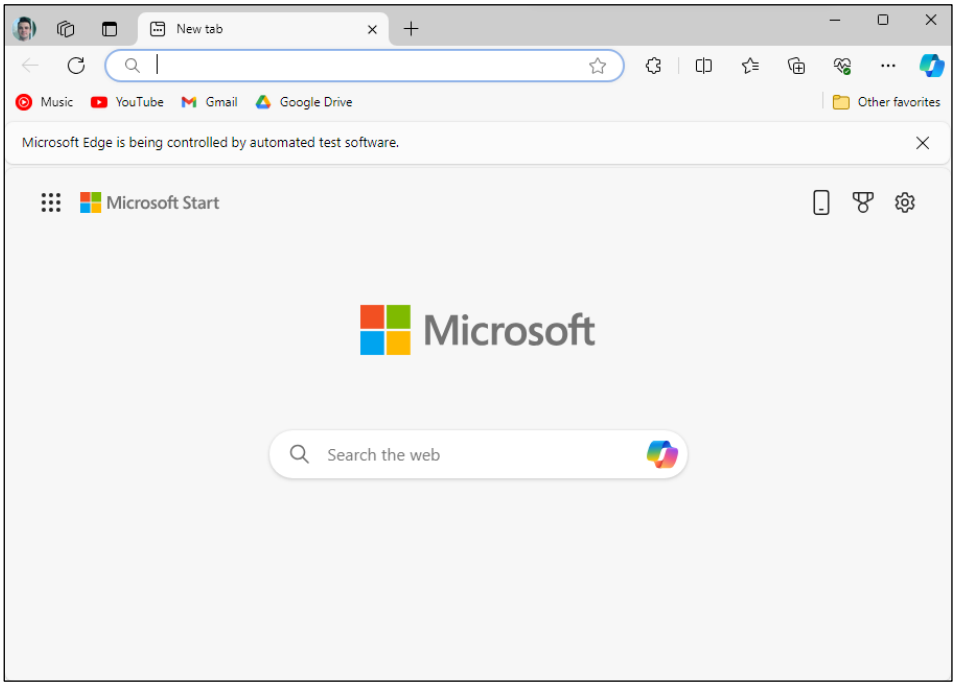


Рисунок 4.4 – Відкриття веб-браузеру після запуску драйвера

4.3.4 Перехід за URL адресою новини

Цей крок передбачає перехід до URL-адреси новини в Twitter. Для цього буде потрібно ввести в консолі другу за списком команду. Після введення номеру, з'явиться запит на введення URL-адреси новини. Вводимо адресу новини з

соціальної мережі Twitter [11].

Важливо зазначити, що URL слід вводити точно так, як вона відображається в адресному рядку браузера. Будь-які помилки в веденні можуть призвести до того, що інструмент не зможе знайти потрібну новину.

Після введення URL-адреси інструмент автоматично перейде на новину в Твіттері. Це дозволить програмі проаналізувати зміст статті і видобути потрібну інформацію (див.рис.4.5).

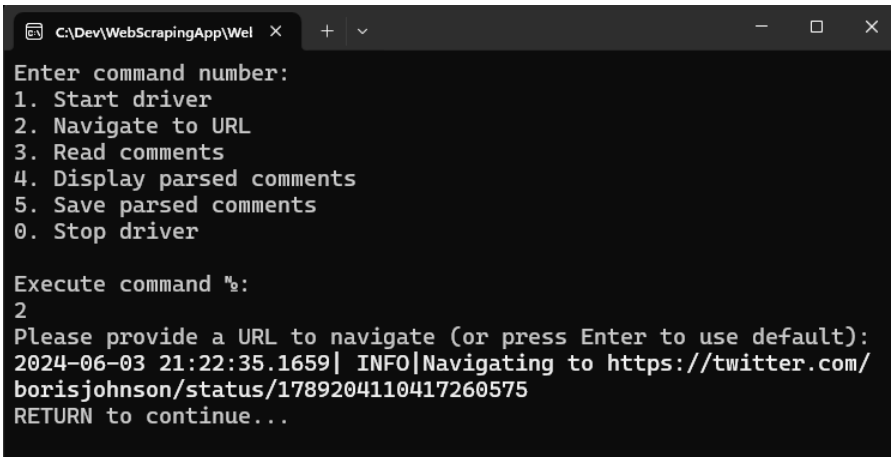


Рисунок 4.5 – Виконання команди переходу до новини

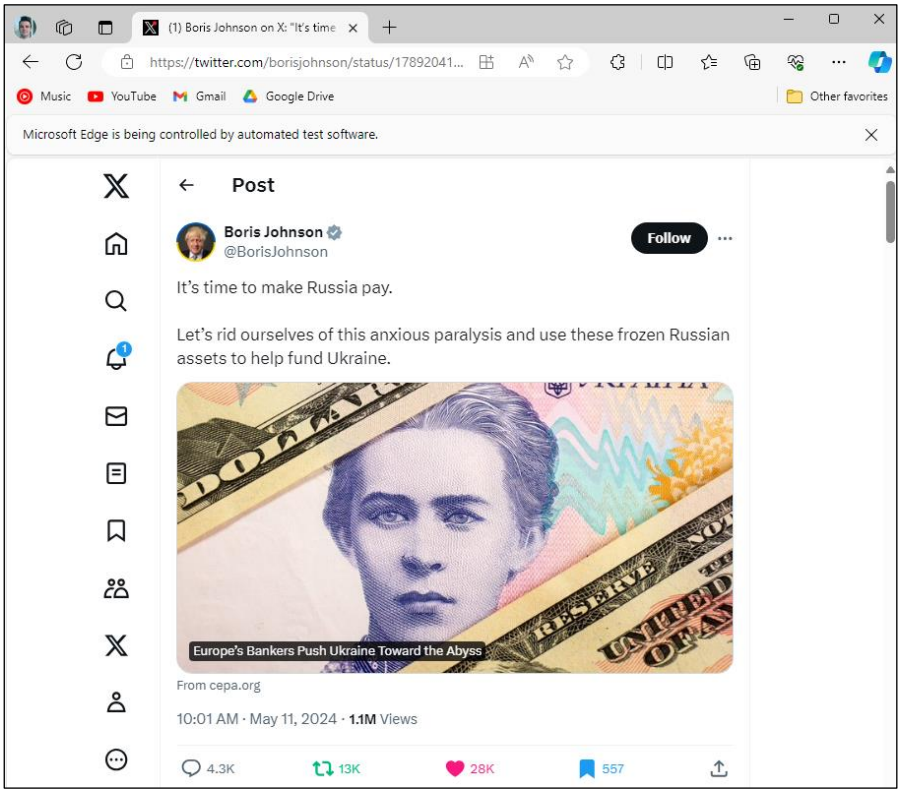


Рисунок 4.6 – Відкриття новини в браузері

4.3.5 Зчитування коментарів

Щоб почати зчитування коментарів під постом у Твіттері, потрібно ввести в консолі наступну команду. Користувачу буде запропоновано вказати кількість коментарів, необхідну для подальшого емоційного аналізу. Для цієї демонстрації буде обрано 50 коментарів.

Після того, як буде введете потрібну кількість коментарів, програма почне зчитування коментарі з веб-сторінки. У консолі буде відображено інформацію про початок прогресу та про закінчення.

Після початку виконання команди буде помітно, що веб-сторінка автоматично прокручується вниз. Це означає, що програма зчитує коментарі зі сторінки. Застосунок продовжить прокручувати сторінку вниз, поки не досягне кінця коментарів або поки не прочитає вказану кількість.

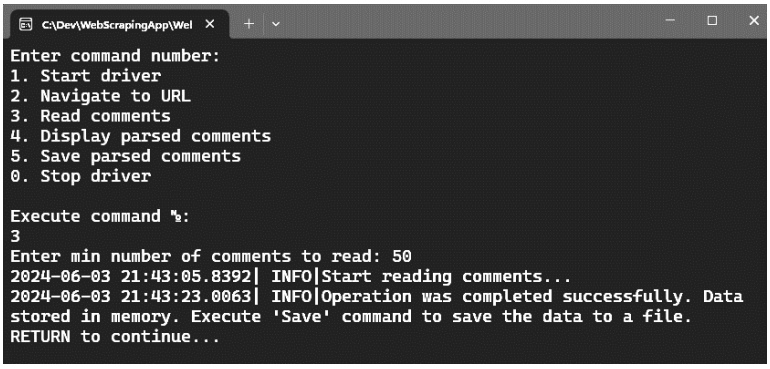


Рисунок 4.7 – Виконання команди зчитування коментарів

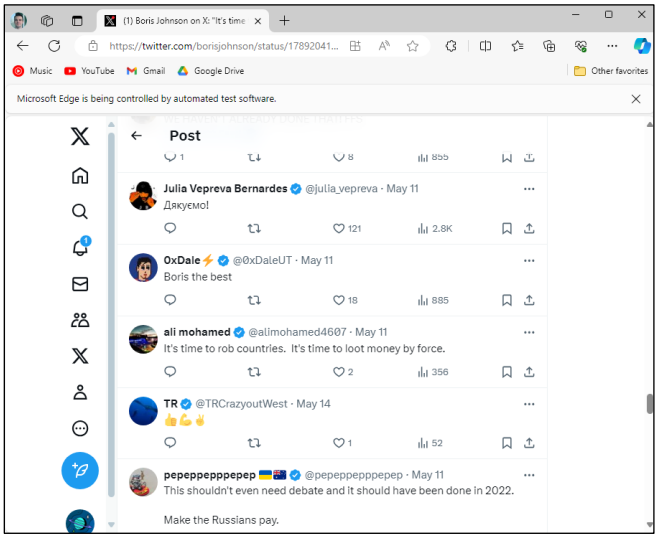
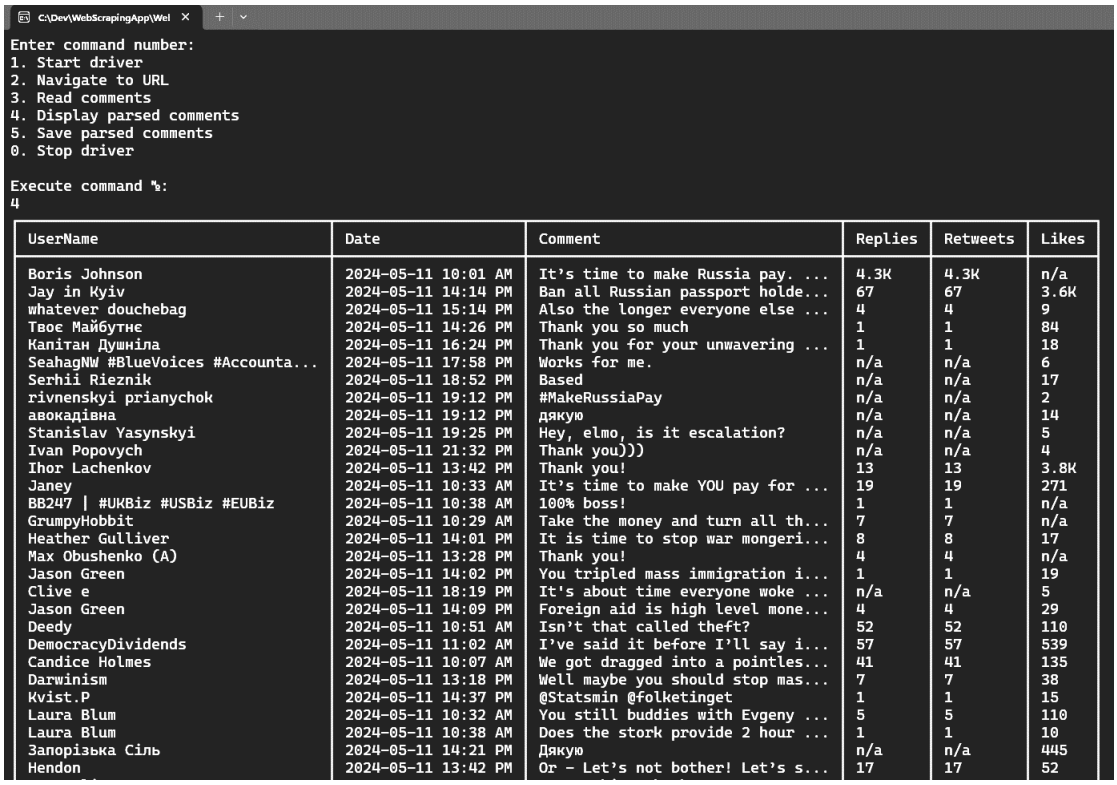


Рисунок 4.8 – Прокручування сторінки під час зчитування коментарів

4.3.6 Перегляд результатів зчитування в консолі

Після того, як коментарі будуть прочитані, їх результати можна переглянути в консолі. Їх можна прокручувати та переглядати в зручній таблиці, але текст коментарів буде урізаний для кращого відображення (див.рис.4.9).



```
C:\Dev\WebScrapingApp\Wel
Enter command number:
1. Start driver
2. Navigate to URL
3. Read comments
4. Display parsed comments
5. Save parsed comments
0. Stop driver

Execute command %:
4
```

UserName	Date	Comment	Replies	Retweets	Likes
Boris Johnson	2024-05-11 10:01 AM	It’s time to make Russia pay. ...	4.3K	4.3K	n/a
Jay in Kyiv	2024-05-11 14:14 PM	Ban all Russian passport holde...	67	67	3.6K
whatever douchebag	2024-05-11 15:14 PM	Also the longer everyone else ...	4	4	9
Твоє Майбутнє	2024-05-11 14:26 PM	Thank you so much	1	1	84
Капітан Душніла	2024-05-11 16:24 PM	Thank you for your unwavering ...	1	1	18
SeahagNW #BlueVoices #Accounta...	2024-05-11 17:58 PM	Works for me.	n/a	n/a	6
Serhii Rieznik	2024-05-11 18:52 PM	Based	n/a	n/a	17
rivnenskyi prianychok	2024-05-11 19:12 PM	#MakeRussiaPay	n/a	n/a	2
авокадівна	2024-05-11 19:12 PM	дякую	n/a	n/a	14
Stanislav Yasynskyi	2024-05-11 19:25 PM	Hey, elmo, is it escalation?	n/a	n/a	5
Ivan Popovych	2024-05-11 21:32 PM	Thank you)))	n/a	n/a	4
Ihor Lachenkov	2024-05-11 13:42 PM	Thank you!	13	13	3.8K
Janey	2024-05-11 10:33 AM	It’s time to make YOU pay for ...	19	19	271
BB247 | #UKBiz #USBiz #EUBiz	2024-05-11 10:38 AM	100% boss!	1	1	n/a
GrumpyHobbit	2024-05-11 10:29 AM	Take the money and turn all th...	7	7	n/a
Heather Gulliver	2024-05-11 14:01 PM	It is time to stop war mongeri...	8	8	17
Max Obushenko (A)	2024-05-11 13:28 PM	Thank you!	4	4	n/a
Jason Green	2024-05-11 14:02 PM	You tripled mass immigration i...	1	1	19
Clive e	2024-05-11 18:19 PM	It‘s about time everyone woke ...	n/a	n/a	5
Jason Green	2024-05-11 14:09 PM	Foreign aid is high level mone...	4	4	29
Deedy	2024-05-11 10:51 AM	Isn’t that called theft?	52	52	110
DemocracyDividends	2024-05-11 11:02 AM	I’ve said it before I’ll say i...	57	57	539
Candice Holmes	2024-05-11 10:07 AM	We got dragged into a pointles...	41	41	135
Darwinism	2024-05-11 13:18 PM	Well maybe you should stop mas...	7	7	38
Kvist.P	2024-05-11 14:37 PM	@Statsmin @folketinget	1	1	15
Laura Blum	2024-05-11 10:32 AM	You still buddies with Evgeny ...	5	5	110
Laura Blum	2024-05-11 10:38 AM	Does the stork provide 2 hour ...	1	1	10
Запорізька Сіль	2024-05-11 14:21 PM	Дякую	n/a	n/a	445
Hendon	2024-05-11 13:42 PM	Or - Let’s not bother! Let’s s...	17	17	52

Рисунок 4.9 – Виконання команди перегляду зчитаних коментарів

4.3.7 Збереження результатів

Щоб зберегти прочитані коментарі в CSV-файл, потрібно ввести в консолі п’яту команду, і коментарі будуть збережені, а програма покаже шлях до вказаного CSV-файлу.

Файл CSV буде містити наступні стовпці:

– автор коментаря;

– дата публікації;

– текст;

– кількість відповідей;

– кількість ретвітів (для Twitter);

– кількість вподобань або реакцій (у випадку Facebook).

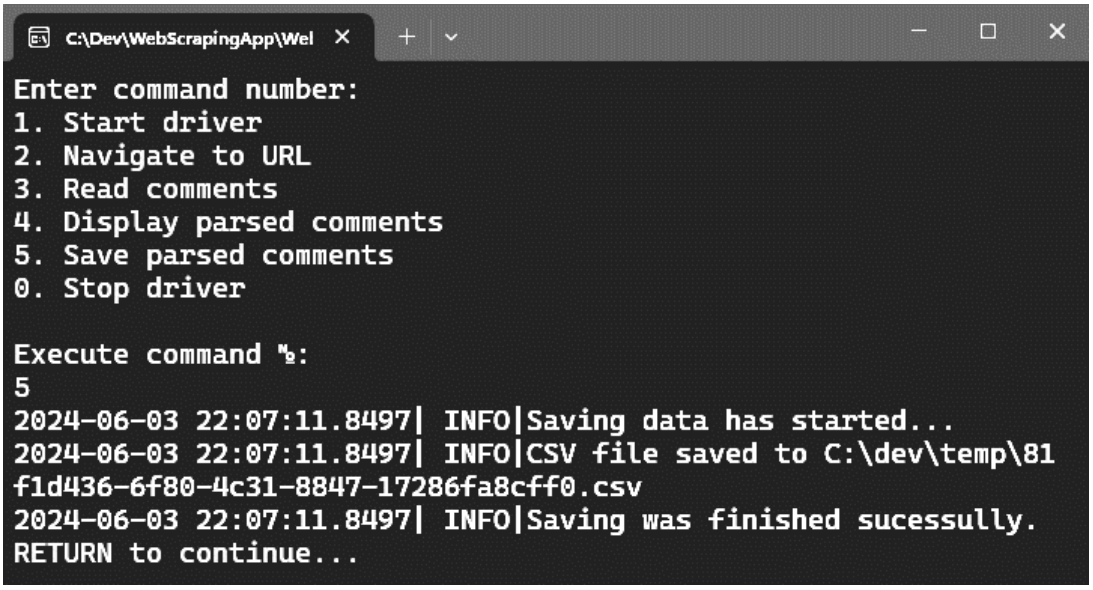
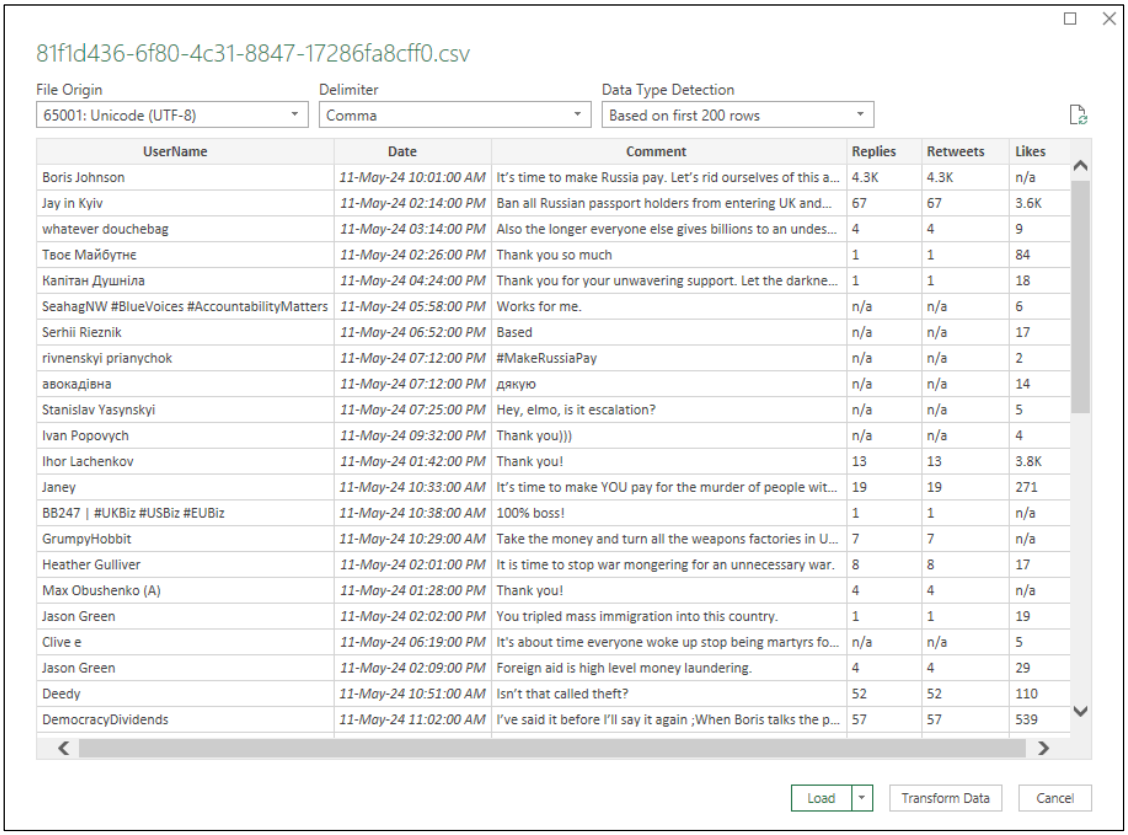


Рисунок 4.10 – Виконання команди збереження зчитаних коментарів

4.3.8 Перегляд результатів у документі Excel

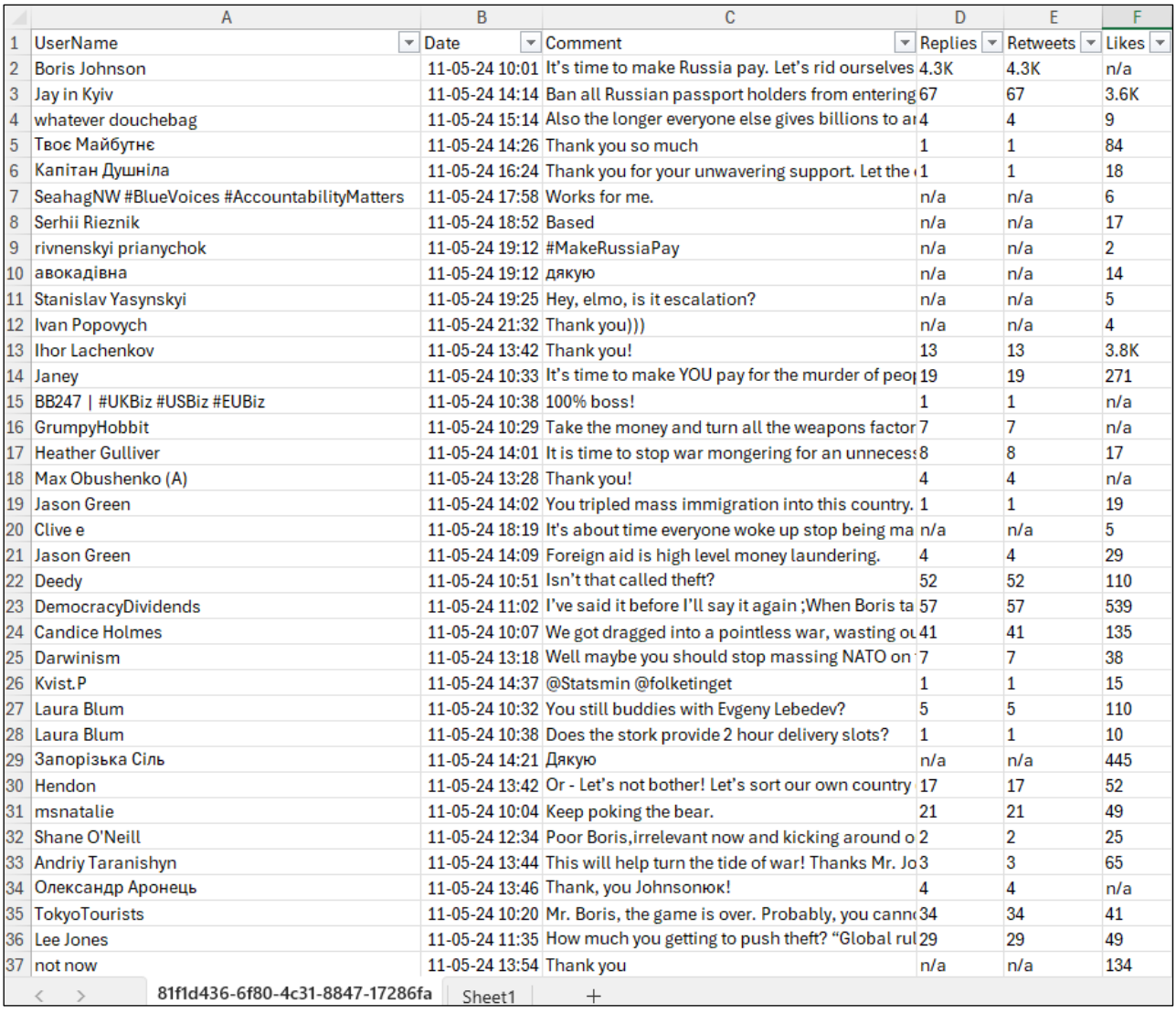
Для прикладу, збережений файл можна завантажити через MS Excel, та подивитися в якому вигляді будуть дані (рис. 4.11).



UserName	Date	Comment	Replies	Retweets	Likes
Boris Johnson	11-May-24 10:01:00 AM	It's time to make Russia pay. Let's rid ourselves of this a...	4.3K	4.3K	n/a
Jay in Kyiv	11-May-24 02:14:00 PM	Ban all Russian passport holders from entering UK and...	67	67	3.6K
whatever douchebag	11-May-24 03:14:00 PM	Also the longer everyone else gives billions to an undes...	4	4	9
Твое Майбутнє	11-May-24 02:26:00 PM	Thank you so much	1	1	84
Капітан Душніла	11-May-24 04:24:00 PM	Thank you for your unwavering support. Let the darkne...	1	1	18
SeahagNW #BlueVoices #AccountabilityMatters	11-May-24 05:58:00 PM	Works for me.	n/a	n/a	6
Serhii Rieznik	11-May-24 06:52:00 PM	Based	n/a	n/a	17
rivnenskyi prianychok	11-May-24 07:12:00 PM	#MakeRussiaPay	n/a	n/a	2
авокадівна	11-May-24 07:12:00 PM	дякую	n/a	n/a	14
Stanislav Yasynskyi	11-May-24 07:25:00 PM	Hey, elmo, is it escalation?	n/a	n/a	5
Ivan Popovych	11-May-24 09:32:00 PM	Thank you)))	n/a	n/a	4
Ihor Lachenkov	11-May-24 01:42:00 PM	Thank you!	13	13	3.8K
Janey	11-May-24 10:33:00 AM	It's time to make YOU pay for the murder of people wit...	19	19	271
BB247 | #UKBiz #USBiz #EUBiz	11-May-24 10:38:00 AM	100% boss!	1	1	n/a
GrumpyHobbit	11-May-24 10:29:00 AM	Take the money and turn all the weapons factories in U...	7	7	n/a
Heather Gulliver	11-May-24 02:01:00 PM	It is time to stop war mongering for an unnecessary war.	8	8	17
Max Obushenko (A)	11-May-24 01:28:00 PM	Thank you!	4	4	n/a
Jason Green	11-May-24 02:02:00 PM	You tripled mass immigration into this country.	1	1	19
Clive e	11-May-24 06:19:00 PM	It's about time everyone woke up stop being martyrs fo...	n/a	n/a	5
Jason Green	11-May-24 02:09:00 PM	Foreign aid is high level money laundering.	4	4	29
Deedy	11-May-24 10:51:00 AM	Isn't that called theft?	52	52	110
DemocracyDividends	11-May-24 11:02:00 AM	I've said it before I'll say it again ;When Boris talks the p...	57	57	539

Рисунок 4.11 – Завантаження збереженого файлу через MS Excel

У кінцевому результаті зчитані дані будуть виглядати в наступному форматі, схожому на ті, як при виводі в консоль (рис. 4.12).



	A	B	C	D	E	F
1	UserName	Date	Comment	Replies	Retweets	Likes
2	Boris Johnson	11-05-24 10:01	It's time to make Russia pay. Let's rid ourselves	4.3K	4.3K	n/a
3	Jay in Kyiv	11-05-24 14:14	Ban all Russian passport holders from entering	67	67	3.6K
4	whatever douchebag	11-05-24 15:14	Also the longer everyone else gives billions to ar	4	4	9
5	Твоє Майбутнє	11-05-24 14:26	Thank you so much	1	1	84
6	Капітан Душніла	11-05-24 16:24	Thank you for your unwavering support. Let the (	1	1	18
7	SeahagNW #BlueVoices #AccountabilityMatters	11-05-24 17:58	Works for me.	n/a	n/a	6
8	Serhii Rieznik	11-05-24 18:52	Based	n/a	n/a	17
9	rivnenskyi prianychok	11-05-24 19:12	#MakeRussiaPay	n/a	n/a	2
10	авокадівна	11-05-24 19:12	дякую	n/a	n/a	14
11	Stanislav Yasynskyi	11-05-24 19:25	Hey, elmo, is it escalation?	n/a	n/a	5
12	Ivan Popovych	11-05-24 21:32	Thank you)))	n/a	n/a	4
13	Ihor Lachenkov	11-05-24 13:42	Thank you!	13	13	3.8K
14	Janey	11-05-24 10:33	It's time to make YOU pay for the murder of peo	19	19	271
15	BB247 | #UKBiz #USBiz #EUBiz	11-05-24 10:38	100% boss!	1	1	n/a
16	GrumpyHobbit	11-05-24 10:29	Take the money and turn all the weapons factor	7	7	n/a
17	Heather Gulliver	11-05-24 14:01	It is time to stop war mongering for an unneces	8	8	17
18	Max Obushenko (A)	11-05-24 13:28	Thank you!	4	4	n/a
19	Jason Green	11-05-24 14:02	You tripled mass immigration into this country.	1	1	19
20	Clive e	11-05-24 18:19	It's about time everyone woke up stop being ma	n/a	n/a	5
21	Jason Green	11-05-24 14:09	Foreign aid is high level money laundering.	4	4	29
22	Deedy	11-05-24 10:51	Isn't that called theft?	52	52	110
23	DemocracyDividends	11-05-24 11:02	I've said it before I'll say it again ;When Boris ta	57	57	539
24	Candice Holmes	11-05-24 10:07	We got dragged into a pointless war, wasting ou	41	41	135
25	Darwinism	11-05-24 13:18	Well maybe you should stop massing NATO on	7	7	38
26	Kvist.P	11-05-24 14:37	@Statsmin @folketinget	1	1	15
27	Laura Blum	11-05-24 10:32	You still buddies with Evgeny Lebedev?	5	5	110
28	Laura Blum	11-05-24 10:38	Does the stork provide 2 hour delivery slots?	1	1	10
29	Запорізька Сіль	11-05-24 14:21	Дякую	n/a	n/a	445
30	Hendon	11-05-24 13:42	Or - Let's not bother! Let's sort our own country	17	17	52
31	msnatalie	11-05-24 10:04	Keep poking the bear.	21	21	49
32	Shane O'Neill	11-05-24 12:34	Poor Boris,irrelevant now and kicking around o	2	2	25
33	Andriy Taranishyn	11-05-24 13:44	This will help turn the tide of war! Thanks Mr. Jo	3	3	65
34	Олександр Аронець	11-05-24 13:46	Thank, you Johnsonюк!	4	4	n/a
35	TokyoTourists	11-05-24 10:20	Mr. Boris, the game is over. Probably, you cann	34	34	41
36	Lee Jones	11-05-24 11:35	How much you getting to push theft? "Global rul	29	29	49
37	not now	11-05-24 13:54	Thank you	n/a	n/a	134

81f1d436-6f80-4c31-8847-17286fa Sheet1

Рисунок 4.12 – Фінальний вигляд зчитаних даних у MS Excel документі

Отже, реалізований програмний комплекс WebScrapingApp ефективно витягує коментарі з новинних постів Twitter. Після запуску програми та взаємодії з консоллю, система представляє ряд команд для управління процесом. Вони включають запуск веб-драйвера, перехід за вказаною URL-адресою, читання коментарів, відображення результатів, збереження зібраних даних і завершення роботи драйвера. Система автоматично прокручує веб-сторінку вниз для збору коментарів і відображає результати в консолі. Зібрані коментарі можна зберегти у форматі CSV та імпортувати в документ Excel для зручного перегляду.

4.3 Отримані результати

Отримані результати демонструють успішну реалізацію системи збору та попередньої обробки коментарів до новин із соціальних мереж. Розроблений програмний інструмент ефективно використовує Selenium WebDriver для автоматизації процесу збору даних, що дозволяє отримувати актуальну інформацію про реакції користувачів на новини.

Подальший розвиток системи може бути спрямований на вдосконалення аналізу емоційного забарвлення коментарів шляхом інтеграції методів машинного навчання, таких як BERT або GPT, що дозволить враховувати контекстуальні нюанси та сарказм. Розширення функціональності системи на інші соціальні платформи та мови забезпечить більш повне розуміння громадської думки та настроїв у різних соціальних групах та регіонах. Впровадження моніторингу та візуалізації в реальному часі дозволить відстежувати динаміку емоційних реакцій, виявляти тренди та приймати обґрунтовані рішення на основі актуальних даних.

Додатковим напрямком розвитку є врахування контексту новин при аналізі емоційного забарвлення коментарів, що підвищить точність інтерпретації. Аналіз зібраних додаткових метаданих, таких як кількість вподобань, репостів та час публікації, дозволить отримати більш детальну інформацію про реакцію користувачів та виявити приховані закономірності.

У цілому, результати дослідження підтверджують ефективність запропонованого підходу та відкривають перспективи для подальшого розвитку системи в напрямку підвищення точності аналізу емоційного забарвлення коментарів, розширення функціональності та застосування в різних сферах дослідження громадської думки та соціальної поведінки.

ВИСНОВКИ

У цьому дослідженні було розроблено програмну систему для збору та попередньої обробки коментарів до новинних публікацій для подальшого аналізу їхнього емоційного забарвлення. Для автоматизації збору коментарів з англомовних соціальних мереж було використано Selenium WebDriver, що забезпечило ефективний та актуальний збір даних. Зібрані коментарі були стандартизовані та конвертовані в уніфікований формат CSV, що дозволяє забезпечити їх подальший процес емоційної обробки. Вибір Selenium WebDriver був обґрунтований його здатністю обробляти динамічний веб-контент і взаємодіяти з сучасними веб-технологіями, що робить його цінним інструментом для збору та аналізу даних у режимі реального часу в контексті онлайн-дискусій та суспільних настроїв.

Розроблена система демонструє практичне застосування методів веб-скрепінгу та обробки даних з метою аналізу громадської думки та настроїв. Автоматизуючи збір та підготовку коментарів, це дослідження робить внесок у розвиток інструментів для розуміння онлайн-дискурсу та його емоційних наслідків. Результати цього дослідження можуть бути застосовані в різних галузях, включаючи маркетинг, соціологію та політологію, щоб отримати уявлення про реакцію громадськості на новини та події. Рішення використовувати CSV-файли для зберігання даних, а не базу даних, було зумовлене насамперед прагненням до простоти та швидкої розробки, що відповідає принципам MVP (Minimum Viable Product - мінімальний життєздатний продукт).

Дане дослідження закладає перспективну основу для майбутніх удосконалень і поліпшень, таких як включення більш складних методів емоційного аналізу, можливість підтримки різних мов і платформ, а також впровадження функцій моніторингу та візуалізації в реальному часі.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Gallagher C., Furey E., Curran K. The Application of Sentiment Analysis and Text Analytics to Customer Experience Reviews to Understand What Customers Are Really Saying. International Journal of Data Warehousing and Mining. 2019. Vol. 15, no. 4. P. 21–47. URL: https://doi.org/10.4018/ijdwm.2019100102 (дата звернення: 28.03.2024).

2. Web Scraping Approaches and their Performance on Modern Websites / A. S. Bale et al. 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 17–19 August 2022. 2022. URL: https://doi.org/10.1109/icesc54411.2022.9885689 (дата звернення: 01.03.2024).

3. Khder M. Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. International Journal of Advances in Soft Computing and its Applications. 2021. Vol. 13, no. 3. P. 145–168. URL: https://doi.org/10.15849/ijasca.211128.11 (дата звернення: 01.03.2024).

4. Contributors to Wikimedia projects. Entropy (information theory) - Wikipedia. Wikipedia, the free encyclopedia. URL: https://en.wikipedia.org/wiki/Entropy_(information_theory) (дата звернення: 05.03.2024).

5. Guo J. Deep learning approach to text analysis for human emotion detection from big data. Journal of Intelligent Systems. 2022. Vol. 31, no. 1. P. 113–126. URL: https://doi.org/10.1515/jisys-2022-0001 (date of access: 08.03.2024).

6. Selenium WebDriver documentation. URL: https://www.selenium.dev/documentation/webdriver/ (дата звернення: 28.03.2024).

7. Задача аналізу тональності тексту Шуляк С.М, Валенда Н.А. Topical issues of the development of modern science // Abstracts of the 9th International scientific and practical conference. Sofia, Bulgaria: ACCENT, 2020. с. 951-956.

8. Мироненко С., Онищенко Є. Порівняльний аналіз методів для вирішення задачі сентимент аналізу тексту. COMPUTER-INTEGRATED TECHNOLOGIES: EDUCATION, SCIENCE, PRODUCTION. 2020. № 40. С. 140–145. URL: https://doi.org/10.36910/6775-2524-0560-2020-40-21 (дата звернення:

28.03.2024).

9. Mathur A., Kubde P., Vaidya S. Emotional Analysis using Twitter Data during Pandemic Situation: COVID-19. *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, м. COIMBATORE, India, 10–12 черв. 2020 р. 2020. URL: https://doi.org/10.1109/icces48766.2020.9138079 (дата звернення: 21.04.2024).

10. Бугай Д.Ю., Валенда Н.А., Застосування веб-скрапінгу та аналізу тональності коментарів для дослідження реакцій на новини // 28-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті», Харків, ХНУРЕ, 2024. С. 314-316.

11. Приклад новини для тесту зчитування коментарів у соціальній мережі Twitter. URL: https://twitter.com/borisjohnson/status/1789204110417260575 (дата звернення: 20.05.2024)