ДОДАТОК А

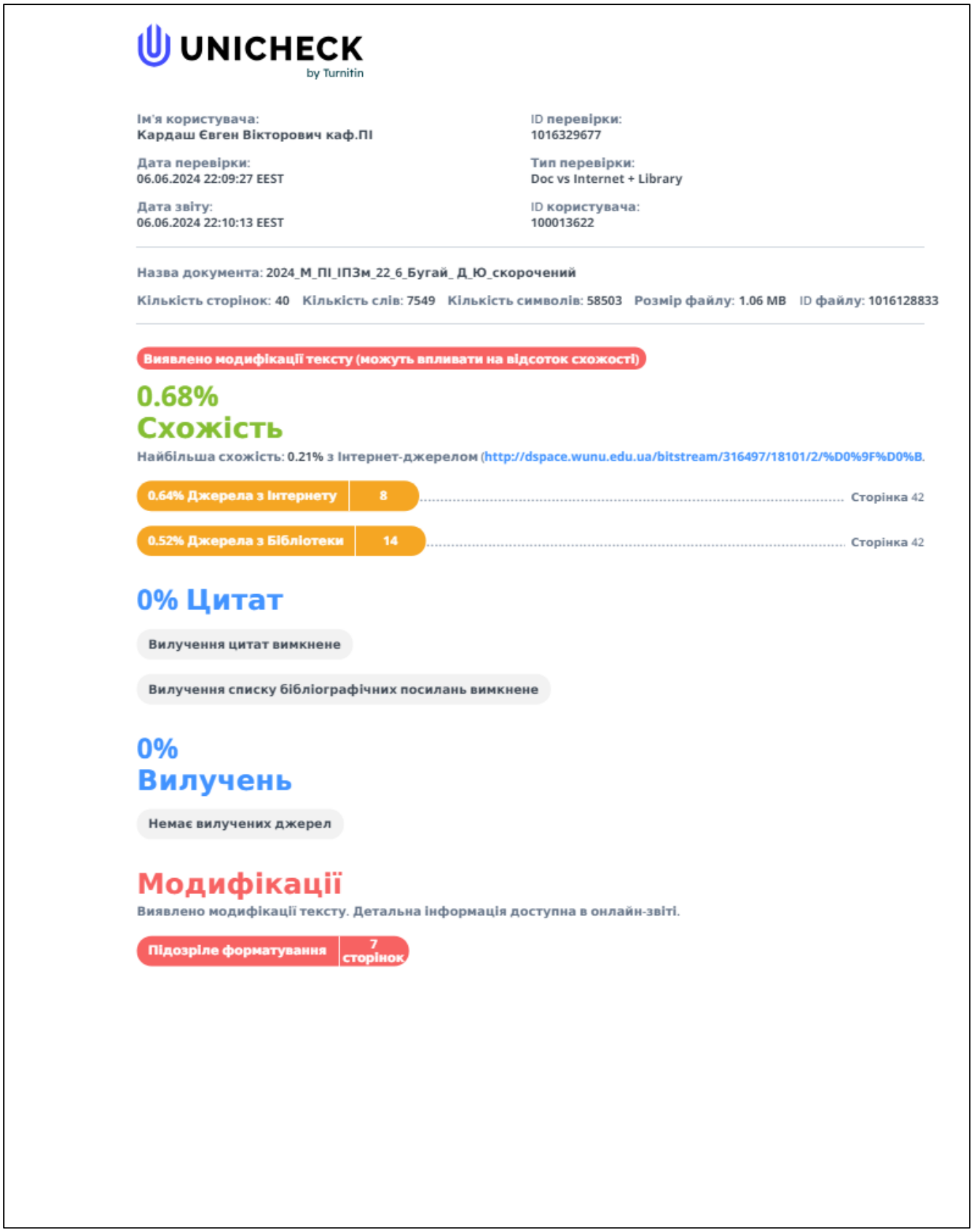
Перелік джерел посилання за науковими напрямами керівника та науковців кафедри програмної інженерії

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.

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.

ДОДАТОК Б

Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ



UNICHECK
by Turnitin

Ім'я користувача:
Кардаш Євген Вікторович каф.ПІ

ID перевірки:
1016329677

Дата перевірки:
06.06.2024 22:09:27 EEST

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
06.06.2024 22:10:13 EEST

ID користувача:
100013622

Назва документа: 2024_М_ПІ_ІПЗм_22_6_Бугай_ Д_Ю_скорочений

Кількість сторінок: 40 Кількість слів: 7549 Кількість символів: 58503 Розмір файлу: 1.06 MB ID файлу: 1016128833

Виявлено модифікації тексту (можуть впливати на відсоток схожості)

0.68%
Схожість

Найбільша схожість: 0.21% з Інтернет-джерелом (http://dspace.wunu.edu.ua/bitstream/316497/18101/2/%D0%9F%D0%B.

0.64% Джерела з Інтернету 8 ..... Сторінка 42

0.52% Джерела з Бібліотеки 14 ..... Сторінка 42

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 7 сторінок

ДОДАТОК В

Слайди презентації

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

NURE ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ

Дослідження методів аналізу емоційного забарвлення коментарів. Підготовка даних

Бугай Дмитро Юрійович,
група: ІПЗм-22-6

Науковий керівник: доцент каф. ПІ
Валенда Наталя Анатоліївна

18 червня 2024

SE software engineering

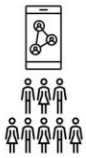
Дослідження

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

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

SE software engineering

Актуальність



Значний вплив соціальних медіа-платформ на формування громадської думки

Потреба в розумінні настроїв населення та тенденцій їх зміни

Відкриття шляхів для планування, як покращити емоційне самопочуття населення

Постановка задачі

- аналіз методів збору коментарів під новинами;
- створення єдиної форми даних;
- збереження результатів для подальшої передачі на аналіз;
- валідація на реальних даних.

Очікуваний результат – зробити вагомий внесок у розвиток емоційного аналізу настроїв населення, продемонструвати практичне застосування методів веб-скрепінгу коментарів під постами новин та обробки зібраних даних.

Аналіз існуючих рішень

Найбільш поширені готові рішення для взаємодії з HTML-сторінками

Порівняння існуючих рішень

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

Використані технології

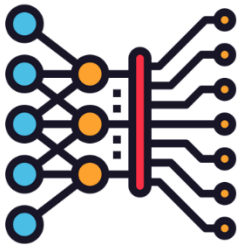
На основі аналізу існуючих рішень було обрано технологію Selenium, для взаємодії з веб-сторінками

Теорія емоційного аналізу тексту

Dictionary

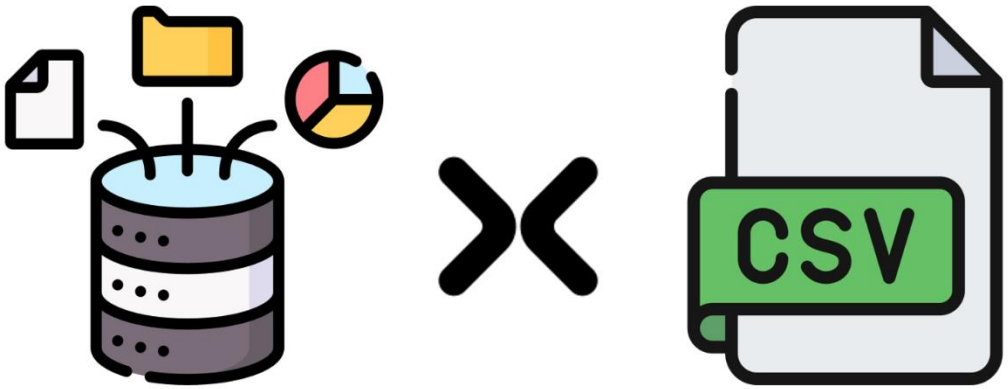
Continuous learning

Machine learning



Deep learning

Збереження даних



Архітектура системи

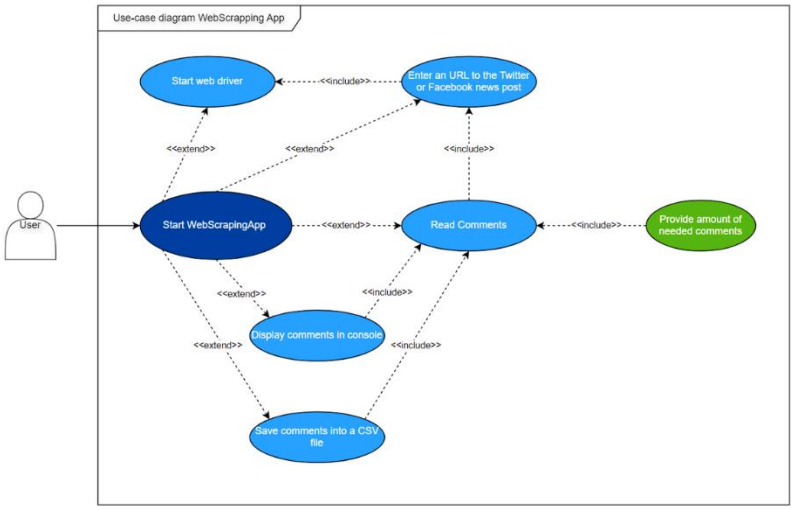
Web Scraping App

Керує загальним робочим процесом збору та попередньої обробки даних

Selenium Scraper

Отримує команди, виконує веб-скрепінг та взаємодіє з HTML документами

Демонстрація Use-case сценаріїв



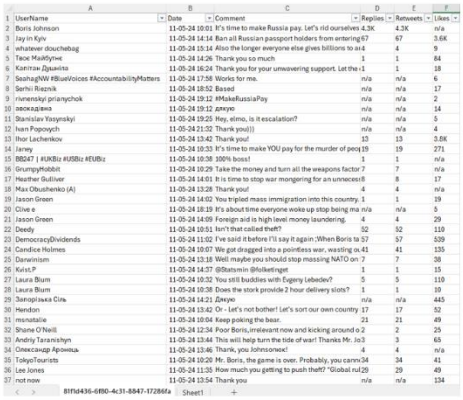
Результати зчитування даних

```
Enter command number:
1. Start driver
2. Navigate to URL
3. Read comments
4. Display parsed comments
5. Save parsed comments
6. 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
Таос Майбутнє	2024-05-11 14:26 PM	Thank you so much	1	1	84
Капітан Дуакіла	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 Riezzik	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
авожадіма	2024-05-11 19:12 PM	дякую	n/a	n/a	14
Stanislav Yasynskyi	2024-05-11 19:25 PM	Hey, else, 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 Січ	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

Консольний вигляд



	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 a	4	4	9
5	Таос Майбутнє	11-05-24 14:26	Thank you so much	1	1	84
6	Капітан Дуакіла	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 Riezzik	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	авожадіма	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 unnecess	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 to	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 Січ	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	29
33	Andriy Taranishyn	11-05-24 13:44	This will help turn the tide of war! Thanks Mr. Jo	3	3	65
34	Олександр Арончык	11-05-24 13:46	Thank, you Johnsonwoc!	4	4	n/a
35	TokyoTourists	11-05-24 10:20	Mr. Boris, the game is over. Probably, you canno	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

Відкритий CSV файл у Excel

Аналіз отриманих результатів

- У цьому дослідженні було розроблено програмну систему, яка автоматизує збір та попередню обробку коментарів із соціальних мереж для подальшого аналізу їх емоційного забарвлення. Система використовує Selenium WebDriver для веб-скрейпінгу, обробляє дані та зберігає їх у CSV-файлі для подальшого аналізу.

- Майбутні дослідження можуть бути спрямовані на розширення мовного та культурного контексту, включення більш досконалих методів аналізу емоцій, підтримку різних мов, впровадження функцій моніторингу та візуалізації в реальному часі, а також врахування контексту новин та додаткових метаданих для підвищення точності аналізу

Публікація дослідження

XXVIII Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті»

Підсумки

- розроблено програмну систему для збору, попередньої обробки та підготовки коментарів під новинами для подальшого аналізу емоційного забарвлення;
- система демонструє практичне застосування методів веб-скрепінгу та обробки зчитаних даних;
- результати дослідження можуть бути застосовані в різних сферах, включаючи маркетинг, соціологію та політологію.

ДОДАТОК Г

Текст наукової публікації за темою кваліфікаційної роботи

УДК 004.89 DOI: https://doi.org/10.30837/IYF.IIS.2024.321

ЗАСТОСУВАННЯ ВЕБ-СКРАПІНГУ ТА АНАЛІЗУ ТОНАЛЬНОСТІ КОМЕНТАРІВ ДЛЯ ДОСЛІДЖЕННЯ РЕАКЦІЙ НА НОВИНИ

Бугай Д. Ю.
Науковий керівник – к.т.н., доцент Валенда Н. А.
Харківський національний університет радіоелектроніки, каф. ІІІ
м. Харків, Україна
dmytro.buhai@nure.ua

In the digital age, it is essential for stakeholders to comprehend societal reactions to news events. This paper investigates the use of web scraping techniques and sentiment analysis to examine the emotional responses conveyed in user comments. By automating data collection from online platforms and utilizing natural language processing algorithms, this thesis aims to analyze the nuanced emotional tones expressed in user-generated content. The engineering approach involves overcoming technical challenges, such as navigating through dynamic web pages and addressing ethical concerns regarding data privacy.

Веб-збирання даних (веб-скрейпінг) – це процес автоматичного збирання інформації з веб-сайтів. Цей процес зазвичай включає в себе використання спеціальних програм або скриптів, які аналізують структуру веб-сторінок і витягують з них потрібні дані. Інформація може бути отримана з текстових елементів, таблиць, зображень або будь-яких інших елементів, що містяться на веб-сторінках [1].

Веб-збирання даних використовується для різноманітних цілей, таких як аналіз ринку, збір конкурентної інформації, створення баз даних, моніторинг цін, автоматизація завдань та багато іншого. Цей процес дозволяє ефективно і швидко отримувати великі обсяги даних з веб-ресурсів, що дає можливість компаніям та дослідникам використовувати цю інформацію для прийняття рішень та виконання різних аналітичних завдань.

Найбільш вдалим інструментом для веб-збирання даних є Selenium. Це універсальний інструмент який, автоматизує взаємодію з веб-сайтами, полегшуючи вилучення даних з динамічних сторінок, що містять багато JavaScript коду. Він відомий завдяки емуляції користувацької взаємодії та пропонує такі функції, як сумісність з різними браузерами, динамічна обробка сторінок, емуляція користувацької взаємодії, підтримка декількох мов, паралельне виконання, перегляд сторінок без заголовків і безперервні оновлення. Він інтегрується з веб-драйвером браузера для розширеної взаємодії, забезпечуючи крос-платформну сумісність з широкою підтримкою спільноти та документації. Ці можливості роблять Selenium потужним і гнучким інструментом для веб-скрейпінгу, особливо для веб-сайтів з динамічним контентом і складною взаємодією з користувачем [2].

Рисунок В.1 – Тези магістерського дослідження з 28-го Міжнародного молодіжного форуму «Радіоелектроніка та молодь у XXI столітті». Перша сторінка

Наприклад, ставлячи собі на меті дізнатися про реакцію суспільства на події в країні, можна навчити штучний інтелект розпізнавати емоційне забарвлення коментарів під постами новин. Веб-збирання даних за допомогою Selenium інструменту, у контексті даного прикладу, може бути досить доречно використане для автоматизованого збору текстової інформації, яка містить коментарі користувачів з веб-сайтів новинних порталів. Цей процес передбачає аналіз HTML-коду веб-сторінок та екстракцію текстових даних, що містять відгуки користувачів. Подальший емоційний аналіз отриманих даних може бути здійснений за допомогою методів обробки природної мови (Natural Language Processing, NLP). Використання NLP дозволяє автоматично визначати емоційний тон коментарів, враховуючи їх семантичне значення та вживані мовні конструкції. Це дозволяє розділити коментарі за потрібними категоріями, наприклад, за їх позитивним, негативним або нейтральним забарвленням [3].

Найсучаснішим методом вирішення різних задач обробки природних мов є використання попередньо тренованих моделей та нейронних мереж з архітектурою «трансформер». Найвідоміший приклад такої моделі – 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 [4].

Отримані результати емоційного аналізу можуть бути використані для розуміння загальної суспільної реакції на конкретну подію або новину. Аналізуючи велику кількість коментарів, можна встановити тенденції в емоційній реакції користувачів та виявити потенційні групи, які сприймають події з різних точок зору. Ці дані можуть бути використані для прогнозування реакції суспільства на подібні події в майбутньому. Таким чином, веб-збирання даних у поєднанні з емоційним аналізом відгуків користувачів може слугувати важливим інструментом для дослідження та передбачення реакції суспільства на новини та події.

Після визначення емоційного тону коментарів можна звернутися до додаткового аналізу контексту та відносин між користувачами. Це може включати дослідження ключових тем, які порушуються в коментарях, а також ідентифікацію взаємодії між різними користувачами. Наприклад, можна вивчити, які конкретні аспекти новин або подій найбільше привертають увагу аудиторії та які теми викликають найбільший емоційний відгук. Це допоможе отримати більш глибоке розуміння того, як суспільство сприймає різні аспекти свого оточення.

Крім того, застосування методів машинного навчання для класифікації коментарів за різними категоріями або за ступенем емоційного навантаження може бути корисним для автоматизації аналізу та обробки великих обсягів даних. Наприклад, можна навчити модель класифікувати коментарі за їхнім ступенем емоційної інтенсивності або за

315

Рисунок В.2 – Тези магістерського дослідження з 28-го Міжнародного молодіжного форуму «Радіоелектроніка та молодь у XXI столітті». Друга сторінка

їхнім тематичним змістом, що допоможе зробити подальший аналіз більш систематизованим та об'єктивним. Використання методів веб-скрейпінгу за допомогою Selenium та аналізу емоційного тону коментарів, використовуючи NLP інструменти, відкриває широкі можливості для дослідження реакцій західного суспільства та міжнародного співтовариства на новини, що стосуються України. Це дозволяє здійснити докладний аналіз міжнародного ставлення до ключових подій, політичних рішень, соціальних змін та інших аспектів життя країни.

Аналіз емоційного тону коментарів дозволяє розуміти ставлення міжнародної аудиторії до подій в Україні на емоційному рівні. Це може включати виявлення позитивних або негативних емоцій, які виражаються у коментарях, а також виявлення основних тематичних категорій або ключових аспектів, які викликають емоційну реакцію серед іноземної аудиторії. За допомогою цих методів дослідження можна отримати глибоке розуміння міжнародної думки щодо подій в Україні.

Список використаних джерел:

1. 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 (date of access: 01.03.2024).

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 (date of access: 01.03.2024).

3. 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: 01.03.2024).

4. 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.

Рисунок В.3 – Тези магістерського дослідження з 28-го Міжнародного молодіжного форуму «Радіоелектроніка та молодь у XXI столітті». Третя сторінка

ДОДАТОК Д

Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008: 2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент (посада) | програмної інженерії (кафедра) | ІПЗм-22-6 (група)

Бугай Дмитро Юрійович

(прізвище, ім'я, по батькові)

Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
	7.3 Нумерація сторінок звіту	
	7.4 Нумерація розділів, підрозділів, пунктів, підпунктів	
	7.5 Рисунки	
	7.6 Таблиці	
	7.7 Переліки	
	7.8 Примітки	
	7.9 Виноски	
	7.10 Формули та рівняння	
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні позначки	
	7.15 Додатки	

зауважень немає

Експерт _____ (підпис) Олена ОЛІЙНИК (прізвище, ініціали)

09.06.2024