

## **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. ПІ

м. Харків, Україна

[dmytro.buhai@nure.ua](mailto: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].

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

їхнім тематичним змістом, що допоможе зробити подальший аналіз більш систематизованим та об'єктивним. Використання методів веб-скреїпінгу 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.