

УДК 004.62

Куликовська Н.А.¹, Тіменко А.В.^{1,2}, Трохимчук В.Є.³, Тіменко К.І.²

¹ старш. викл. НУ «Запорізька політехніка»

² викл. Запорізький авіаційний фаховий коледж ім. О. Г. Івченка

³ студ. гр. КНТ-522 НУ «Запорізька політехніка»

ПОРІВНЯЛЬНИЙ АНАЛІЗ ЕФЕКТИВНОСТІ МЕТОДІВ NLTK ТА TEXTBLOB ДЛЯ АНАЛІЗУ ТОНАЛЬНОСТІ УКРАЇНОМОВНИХ ТЕКСТІВ

Стрімкий розвиток соціальних мереж та збільшення об'єму текстового контенту в інтернеті створює потребу в ефективних інструментах автоматизованого аналізу тональності текстів. Особливо актуальним є питання обробки україномовного контенту, що вимагає адаптації існуючих методів та оцінки їх ефективності [1]. Аналіз тональності тексту знаходить широке застосування у моніторингу соціальних медіа, аналізі відгуків користувачів, оцінці суспільних настроїв та інших сферах [2].

Дослідженню методів аналізу тональності присвячено значну кількість робіт. Зокрема, активно вивчаються підходи на основі машинного навчання та нейронних мереж [3], досліджуються лінгвістично-орієнтовані методи з урахуванням морфологічних особливостей конкретних мов [4]. Особлива увага приділяється аналізу текстів соціальних мереж та відгуків користувачів [5]. Порівняльний аналіз ефективності різних методів та бібліотек, зокрема NLTK та TextBlob, проводиться за показниками точності та швидкодії [6].

Водночас недостатньо дослідженими залишаються питання ефективності застосування популярних бібліотек аналізу тональності для обробки україномовних текстів, зокрема порівняльного аналізу можливостей NLTK та TextBlob у цьому контексті.

В рамках дослідження проведено порівняльний аналіз бібліотек NLTK та TextBlob для задачі аналізу тональності україномовних текстів. Порівняння здійснювалося за критеріями точності виявлення емоційного забарвлення, швидкодії та зручності використання.

З бібліотеки NLTK використано інструмент "Аналіз настроїв VADER", що генерує оцінки позитивної, негативної та нейтральної тональності тексту, та токенизатор "word_tokenize" для розбиття тексту на менші одиниці. Для морфологічного аналізу україномовних текстів, через відсутність відповідного корпусу в NLTK, застосовано бібліотеку rymorphy2.

Загальна блок-схема процесу аналізу тональності представлена на рис. 1.

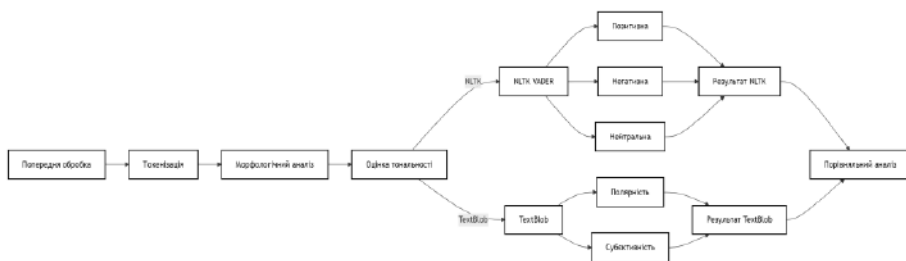


Рисунок 1 – Блок-схема аналізу тональності тексту

Експериментальне порівняння методів проведено на наборі україномовних текстів. Для порівняння двох методів був розроблено програмний модуль, який рахує час роботи кожного алгоритму за однаковими вхідними текстовими даними (рис. 2).

```

11
12 # Аналіз тональності з використанням NLTK
13 start_time = time.time()
14 sentiment_scores = sia.polarity_scores(text)
15 end_time = time.time()
16 # Визначення тональності з використанням NLTK
17 if sentiment_scores['compound'] >= 0.05:
18     sentiment = "Positive"
19 elif sentiment_scores['compound'] <= -0.05:
20     sentiment = "Negative"
21 else:
22     sentiment = "Neutral"
23 print(f"NLTK Sentiment: {sentiment}")
24 print(f"NLTK Execution Time: {end_time - start_time:.6f} seconds")
25 # Аналіз тональності з використанням TextBlob
26 start_timet = time.time()
27 blob = TextBlob(text)
28 sentiment = blob.sentiment.polarity
29 end_timet = time.time()
30 # Визначення тональності з використанням TextBlob
31 if sentiment > 0:
32     sentiment_label = "Positive"
33 elif sentiment < 0:
34     sentiment_label = "Negative"
35 else:
36     sentiment_label = "Neutral"
37 print(f"TextBlob Sentiment: {sentiment_label}")
38 print(f"TextBlob Execution Time: {end_timet - start_timet:.6f} seconds")

```

Рисунок 2 - Порівняння за швидкістю методи NLTK та TextBlob

Результати оцінювання за основними критеріями наведено в табл. 1. NLTK демонструє вищу точність аналізу (80%) та кращі можливості налаштування, проте вимагає більших зусиль у конфігурації та має нижчу швидкодію. TextBlob, маючи дещо нижчу точність (70%), відзначається простотою використання та вищою швидкістю обробки даних.

Таблиця 1 - Порівняння методів аналізу тональності

Метод	Точність виявлення	Зручність використання	Гнучкість та налаштованість
NLTK	80%	70%	90%
TextBlob	70%	90%	60%

У роботі проведено порівняльний аналіз ефективності методів NLTK та TextBlob для аналізу тональності україномовних текстових даних. В результаті експериментального дослідження встановлено, що NLTK забезпечує вищу точність аналізу (80%) та кращі можливості налаштування, проте вимагає більших зусиль при конфігурації. TextBlob демонструє нижчу

точність (70%), але відзначається простотою використання та вищою швидкістю обробки даних. Експериментально підтверджено, що TextBlob виконує аналіз тексту в середньому вдвічі швидше за NLTK.

Практична цінність отриманих результатів полягає у визначенні оптимального інструменту аналізу тональності залежно від конкретних вимог: для задач, що потребують високої точності, доцільно використовувати NLTK, а для швидкого аналізу великих обсягів даних - TextBlob. Подальші дослідження варто спрямувати на розробку гібридних методів, що поєднуюватимуть переваги обох підходів, та їх адаптацію для специфіки україномовних текстів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Куликовська Н. А. Дослідження методів аналізу тональності текстових даних / Н. А. Куликовська, А. В. Тіменко, В. Є. Трохимчук, М. Б. Ільяшенко // Вісник Херсонського національного технічного університету. 2024. № 1(88). С. 235-230.
2. Mukasheva A. Tasks and methods of text sentiment analysis. Scientific journal of astana IT university. 2021. No. 7. P. 55-62.
3. Abonizio H. Q., Paraiso E. C., Barbon Junior S. Toward text data augmentation for sentiment analysis. IEEE transactions on artificial intelligence. 2021. P. 1.
4. Yao J. Automated sentiment analysis of text data with NLTK. Journal of physics: conference series. 2019. Vol. 1187, no. 5. P. 60-78.
5. Jin Y. et al. A review of text sentiment analysis methods and applications. Frontiers in business, economics and management. 2023. Vol. 10, no. 1. P. 58-64.
6. Poria S., Hussain A., Cambria E. Concept extraction from natural text for concept level text analysis. Multimodal sentiment analysis. Cham, 2018. P. 79-84.