

## **АВТОМАТИЧНЕ ВИЗНАЧЕННЯ ТОНАЛЬНОСТІ ТЕКСТОВИХ ДАНИХ**

**Савченко М.В.**

*Національний технічний університет  
«Харківський політехнічний інститут», м. Харків*

Системи такого типу можуть бути використані для відстеження настроїв покупців щодо певних продуктів, швидкого реагування на інциденти, конкурентні дослідження між різноманітними корпораціями, дослідження популярності продуктів від різних виробників на ринку та інше.

Робота присвячена реалізації алгоритму аналізу тональності україномовних текстових даних [1] на базі нейронної мережі. Було реалізовано наступний алгоритм:

- Відбір текстових даних на вхід до алгоритму.
- Оцінювання вхідних даних на наявність тексту та необхідність подальшого аналізу.
- Проведення попереднього оброблення текстових даних шляхом приведення його до формату, що може бути токенизований. Оптимізація тексту включає приведення останнього до нормальної форми (виділення html тегів, прибирання розділових знаків, приведення регістру до одного формату, тощо).
- Токенізація даних для подальшого оброблення.
- Обробка токенів, що включає процеси коли відповідні токени мають бути проведені через процеси видалення стоп-слів, лематизації або стемінгу, опціонально з пропуском даного кроку, та потім, також опціонально, через алгоритм виділення значущих афіксів.
- Аналіз тональності токенів. Цей етап проводиться з використанням нейронної мережі, отримані результати потім обчислюються з використанням оцінок, отриманих на етапі виділення значущих афіксів.
- Підсумовування оцінки для загального тексту.

В роботі описано загальну архітектуру системи на базі описаного алгоритму та розглянуто модулі програмного забезпечення.

Проведено тестування розробленого методу, проаналізовано ефективність роботи методу в залежності від різних типів вхідних даних.

### **Література:**

1. Тональний словник української мови [Електронний ресурс]. – Режим доступу: <https://lang.org.ua/uk/dictionaries/>