



Нотатки до завдання 8

Аналіз тонального тексту

Аналіз тональності (або **sentiment analysis**) у контексті **Big Data** — це процес автоматичного виявлення емоційної забарвленості великої кількості текстових даних. Такий аналіз дозволяє зрозуміти, чи є текст позитивним, негативним або нейтральним, а також виявити емоційні тенденції, думки, ставлення.

Приклади застосування

- **Бізнес:** розуміння клієнтського фідбеку, управління репутацією бренду.
- **Політика:** моніторинг громадської думки під час виборів.
- **Фінанси:** прогнозування ринкових тенденцій на основі новин/відгуків.
- **Соціологія:** аналіз настроїв населення щодо важливих подій.

Виклики

- Сарказм і іронія: складно для машинного розпізнавання.
- Змішані емоції в одному тексті.
- Мова і культура: одні й ті ж слова можуть мати різний сенс.
- Неоднозначність слів (наприклад, "сильно плакав" — добре в контексті сміху, погано — в контексті втрати).

TextBlob

TextBlob — одна з найпростіших у використанні бібліотек для аналізу тональності тексту в Python. Вона підходить для навчання, прототипування, і невеликих проєктів.

```
from textblob import TextBlob
```

```
# Текст для аналізу
```

```
text = "I absolutely love this movie. It was amazing and full of emotions!"

# Створення об'єкта TextBlob
blob = TextBlob(text)

# Аналіз тональності
print("Polarity:", blob.sentiment.polarity) # від -1 до 1
print("Subjectivity:", blob.sentiment.subjectivity) # від 0 до 1
```

Що означають ці значення?

- **Polarity (тональність):**
 - від -1 (дуже негативна)
 - до 1 (дуже позитивна)
 - 0 = нейтральна
- **Subjectivity (суб'єктивність):**
 - від 0 (об'єктивне твердження)
 - до 1 (сильно емоційне/особиста думка)

WordCloud

Бібліотека `wordcloud` в Python використовується для створення **хмар слів** (word clouds) — візуалізацій, де слова представлені у вигляді колажу, а розмір кожного слова залежить від того, скільки разів воно з'являється в тексті.

Це корисно для аналізу текстових даних, пошуку найбільш важливих або повторюваних термінів.

Основні параметри:

- `width` і `height`: Розміри зображення хмари слів.
- `max_words`: Максимальна кількість слів, які будуть відображені на хмарі.
- `background_color`: Колір фону (наприклад, `'white'` або `'black'`).
- `colormap`: Колірна палітра для слів (наприклад, `'Blues'`, `'Greens'`).

- **mask** : Можна вказати форму хмари (наприклад, у вигляді кола чи серця).

Робота з **STOPWORDS :**

- Часто з'являються звичайні слова, такі як "і", "або", "але", які не несуть значущого змісту. Бібліотека **wordcloud** має вбудований набір таких слів (stop words), які можна автоматично виключати з хмари.