

ОСОБЛИВОСТІ АВТОМАТИЗОВАНОГО АНАЛІЗУ ТОНАЛЬНОСТІ ТЕКСТІВ ПІСЕНЬ

Аналіз тональності тексту – це спосіб визначення ставлення автора тексту до предмету, події, людини, або будь-чого іншого, про що йдеться у цьому тексті. Актуальність дослідження полягає у тому, що аналіз тональності текстів пісень не є таким поширеним завданням серед дослідників, як от аналіз тональності заголовків новин, твітів тощо. Дослідження дасть змогу порівняти тональності пісень що належать до різних жанрів та написані різними мовами (англійською та українською).

Аналіз тональності текстів пісень відрізняється від аналізу тональності текстів інших жанрів. Перш за все, тексти відрізняються структурно. Часто в текстах пісень відсутнє римування слів, як у віршах, а більший наголос є на ритмічності. Тому, слова можуть бути скорочені для того щоб текст добре накладався на музику і звучав гарно. Структурна відмінність текстів пісень від, наприклад, відгуків, повідомлень у соціальних мережах, заголовків газет тощо, полягає з-поміж іншого й у тому, що часто в них немає розділових знаків, їх складно розбити на речення .

Для того щоб визначити тональність тексту пісні необхідно цей текст попередньо опрацювати та підготувати для подальшого автоматизованого визначення тональності. Основними етапами цього процесу можна вважати: збір даних (текстів пісень), їх збереження, токенизація, стемінг, визначення частини мови, видалення стоп-слів.

Автоматизований аналіз тональності текстів пісень – завдання непросте. Труднощі виникають зокрема й через те, що в піснях часто використовують крилаті вислови та різні літературні прийоми, які програма не ідентифікує, наприклад “Time to separate the sheep from goats” (Eminem, Fall) або “I found all your skeletons, your closet full of bones” (BANKS, Gemini Feed). Також можуть бути вжиті сленгові слова, наприклад “I see you take pride in bloody eyes, I know you're stoned” (BANKS, Gemini Feed), де словосполучення *bloody eyes* означає *наркотичне сп'яніння*. Іноді в одній фразі одне слово може мати позитивне забарвлення, а інше негативне [3], наприклад “Truна keep it peaceful is a struggle for me” (Drake God's Plan), де *peaceful* – мирний і *struggle* – боротьба. Складно визначити сарказм “Cause you're mentally

retarded, but pretend to be the smartest” (Eminem, The Ringer). Тональність текстів можна визначати на різних рівнях. Оптимальний варіант для нашого дослідження – на рівні тексту однієї пісні. Сукупність текстів пісень дає можливість здійснити аналіз тональності цілого альбому, охарактеризувати період в житті автора, проаналізувати події, які відбувалися з автором у той період тощо [2].

Класифікація тональності документів (Document Sentiment Classification) – це підхід, в якому один документ є єдиним цілим і може мати як позитивну, негативну так і нейтральну тональність. Класифікація тональності на рівні документів передбачає, що документ висловлює думку щодо чогось одного, і думка належать одній людині. Отож, для пісень доцільно використовувати саме такий підхід, а отже розглядати текст однієї пісні як окремий документ [1]. Важливо пам’ятати, що тональність цілого тексту не є сукупністю тональностей окремих його компонентів [3].

Важливо також стосовно кого визначається тональність (мовця, чи того, до кого звертаються/про кого говорять, бо часто це є неоднозначно для одного тексту). Можуть бути речення одночасно позитивні та негативні, наприклад “My biggest flops are your greatest hits”, “But you're losin' the fight you picked” (Eminem, Killshot) [3]. З огляду на те, що метою дослідження є автоматизоване визначення тональності текстів пісень та їх аналіз, нас цікавить ставлення мовця (автора) до того, про що йдеться у тексті пісні. Отже, можемо бачити, що проводити таке дослідження на матеріалі текстів пісень складніше, ніж на матеріалі звичайного тексту (наприклад повідомлення, новини тощо). Необхідно враховувати специфіку структури текстів пісень.

Використані літератури:

1. Paikya K. *Challenges And Techniques For Sentiment Analysis: A Survey* [Електронний ресурс] / K. Paikya, M. Ponn Felciah // *International Journal of Computer Science and Mobile Computing*. – 2015. – Режим доступу до ресурсу: <https://www.ijcsmc.com/docs/papers/March2015/V4I3201599.pdf>.

2. Malheiro R. *Emotion-based Analysis and Classification of Music Lyrics* [Електронний ресурс] / Ricardo Manuel da Silva Malheiro. – 2016. – Режим доступу до ресурсу: http://mir.dei.uc.pt/pdf/Theses/MOODetector/PhD_Thesis_2017_RicardoMalheiro.pdf.

3. Mohammad S. *Challenges in Sentiment Analysis* [Електронний ресурс] / Saif M. Mohammad // *A Practical Guide to Sentiment Analysis*. – 2015. – Режим доступу до ресурсу: <http://www.saifmohammad.com/WebDocs/sentiment-challenges.pdf>.