

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Київський національний університет імені Тараса Шевченка
Навчально-науковий інститут філології
Кафедра української мови та прикладної лінгвістики

**АВТОМАТИЧНИЙ АНАЛІЗ ЕМОЦІЙНОГО ЗАБАРВЛЕННЯ
УКРАЇНСЬКОМОВНИХ ТЕКСТІВ СОЦМЕРЕЖ**

Кваліфікаційна робота бакалавра
студентки IV курсу
ОПП «Прикладна (комп'ютерна)
лінгвістика та англійська мова»,
спеціальності 035 «Філологія»,
спеціалізації 035.10 «Прикладна
лінгвістика»,
галузі знань 03 «Гуманітарні науки»
Марії-Соломії ДМИТРОЦІ

Науковий керівник:
док.філол.н., проф. кафедри
української мови
та прикладної лінгвістики
Наталія ДАРЧУК

«Допущено до захисту»
Протокол засідання кафедри
української мови та прикладної лінгвістики
від 28.05.2025 № 15
Завідувач кафедри _____ **Сергій РІЗНИК**

АНОТАЦІЯ

Кваліфікаційна робота присвячена актуальній проблемі автоматичного аналізу емоційного забарвлення україномовних текстів, що генеруються користувачами соціальних мереж. Метою проєкту є створення системи автоматичного емоційного забарвлення українськомовного тексту соціальних мереж на основі лексичного підходу і розробка рекомендацій з покращення її роботи на підставі кількісного критерію - коефіцієнта Яніса.

У роботі розглянуто теоретичні засади аналізу емоційного забарвлення тексту, а саме: що таке емоції, яка їхня ключова роль у нашому житті, як вони можуть впливати на інтерпретацію текстів і як емоційний контекст може змінювати сприйняття інформації та впливати на реакцію аудиторії. Приділено увагу і способам кодування емоцій в різних мовах, що зумовлює складність їх порівняльного аналізу й інтерпретації. Досліджується вплив тематики, жанру та контексту на точність емоційного аналізу.

Також у роботі досліджено актуальність створення саме українськомовних рішень у сфері аналізу емоцій з огляду на обмежену кількість відповідних досліджень української мови.

Обґрунтовано теоретичні засади класифікації емоцій як основу автоматичного аналізу емоцій у текстах. Головну увагу приділено огляду різних методів автоматичного аналізу емоцій, а саме методу на основі правил, методу на основі ключових слів і технологіям машинного навчання. Описуються результати роботи системи автоматичного аналізу емотивності українськомовного тексту, починаючи з етапу формування корпусу текстів популярних соціальних мереж і завершуючи коефіцієнтом Яніса як одного з індикаторів емоційності тексту. У висновках наводяться результати аналізу недоліків й обмежень, з якими стикається програма, а також шляхи оптимізації та покращення точності аналізу українськомовного тексту. Запропоноване рішення може стати основою для подальших досліджень у сфері аналізу настроїв, інформаційної безпеки та медіаграмотності, а також сприяти створенню україномовних продуктів у галузі обробки природної мови.

Ключові слова: українська мова, емоції, емоційний аналіз, емоційний тон, соціальні мережі, автоматична класифікація, коефіцієнт Яніса, класифікація емоцій.

ABSTRACT

This bachelor's thesis addresses the pressing issue of automatically analyzing the sentiment of Ukrainian-language texts generated by users of social media platforms. The project aims to develop a system for automatic emotional analysis of Ukrainian social media texts based on a lexical approach, and to formulate recommendations for improving its performance using a quantitative indicator — the Janis coefficient.

The thesis explores the theoretical foundations of sentiment analysis, including the nature of emotions, their key role in human life, their impact on text interpretation, and how emotional context can influence the perception of information and audience response. Attention is also paid to the ways emotions are encoded in different languages, which creates challenges for comparative analysis and interpretation. The influence of topic, genre, and context on the accuracy of emotion detection is examined.

The work also investigates the importance of developing Ukrainian-language solutions in the field of sentiment analysis, considering the limited availability of linguistic resources for the Ukrainian language.

The theoretical basis for emotion classification is substantiated as a foundation for automatic sentiment analysis in texts. The main focus is on reviewing various sentiment analysis methods, including rule-based methods, keyword-based methods, and machine learning technologies. The thesis describes the results of the system's performance, from the stage of compiling a text corpus from popular social networks to the application of the Janis coefficient as one of the indicators of emotional content.

The conclusions present an analysis of the limitations and challenges faced by the system, along with possible ways to optimize and improve the accuracy of the emotional analysis of Ukrainian-language texts. The proposed solution may serve as a foundation for further research in the field of sentiment analysis, information security, and media literacy, and contribute to the development of Ukrainian-language tools for natural language processing.

Keywords: Ukrainian language, emotions, emotional analysis, emotional tone, social networks, automatic classification, Janis coefficient, emotion classification.

ЗМІСТ

ВСТУП	7
РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ЕМОЦІЙНОГО ЗАБАРВЛЕННЯ ТЕКСТУ	9
1.1. Роль емоцій у комунікації та їх вплив на інтерпретацію тексту	9
1.2. Способи вираження емоційного стану мовними засобами	11
1.3. Емоційна мова у комунікаційних практиках: ЗМІ, реклама, соцмережі	12
1.4. Способи вираження емоцій в різних мовах	14
1.5. Вплив тематики, жанру та соціального контексту тексту на точність емоційного аналізу	15
1.6. Теоретичні основи класифікації та автоматичного аналізу емоцій у текстах.	
Методи автоматичного аналізу емоцій	17
1.6.1. Теоретичні аспекти класифікації емоцій	17
1.6.2. Автоматичний аналіз емоцій у текстах	17
1.6.3. Методи автоматичного аналізу емоцій	18
Висновок до розділу 1	19
РОЗДІЛ 2. ПРАКТИЧНІ АСПЕКТИ АВТОМАТИЧНОГО АНАЛІЗУ ЕМОЦІЙ	20
2.1. Використання сучасних технологій машинного навчання та аналізу тексту для автоматичного визначення емоцій	20
2.2. Аналіз поширених проблем та помилок при автоматичному виявленні емоцій	22
2.3. Огляд та дослідження наявних баз даних	23
Висновок до розділу 2	26
РОЗДІЛ 3. РОЗРОБКА ПРОГРАМИ АВТОМАТИЧНОГО АНАЛІЗУ ЕМОЦІЙНОСТІ ТЕКСТІВ СОЦМЕРЕЖ	27
3.1. Формування корпусу текстів для аналізу (соцмережі: Telegram, Facebook, Instagram і так далі)	27
3.2. Коефіцієнт Яніса як індикатор емоційності	28

3.3. Розробка та застосування програми для автоматичного виявлення емотивних слів	29
3.4. Виявлення недоліків і обмежень у програмі	36
3.5. Оптимізація та покращення точності аналізу у майбутньому	38
Висновок до розділу 3	39
ВИСНОВОК	40
ДЖЕРЕЛА	43
ДОДАТКИ	48
Додаток 1	48
Додаток 2	49
Додаток 3	50

ВСТУП

У сучасному світі, де є потреба у постійному оновленні інформації, соцмережі стали одним з головних способів комунікації, самовираження та обміну емоціями. Щодня мільйони користувачів публікуються в різних платформах соцмереж, виражаючи свої власні думки, реакції на різні події та переживання. Саме тому різні соціальні платформи стали не лише засобом масової комунікації, але й джерелом для вивчення настроїв людей на різні події та виклики як в межах держави, так і на міжнародному рівні. У цьому контексті особливо **актуальною** є тема автоматичного виявлення емоційного забарвлення текстів у соцмережах, яке набуває особливої ваги для широкого спектру застосувань, включаючи моніторинг суспільних настроїв, аналіз різних ситуацій і навіть маркетингових кампаній. Завдання автоматичного виявлення емоційного забарвлення у соцмережах є не тільки важливим для лінгвістики, а й для психології та інформаційних технологій.

Метою проєкту є створення системи автоматичного емоційного забарвлення українськомовного тексту соціальних мереж на основі лексичного підходу і розробка рекомендацій з покращення її роботи на підставі кількісного критерію - коефіцієнта Яніса.

Ідея дослідження полягає в інтеграції лінгвістичного та комп'ютерного підходів для створення алгоритму, здатного виявляти позитивні, негативні емоції в українськомовних текстах соціальних мереж шляхом ідентифікації емотивних слів, оцінки їхньої кореляції та виявлення закономірностей і особливостей емоційності за допомогою інструментів комп'ютерної лінгвістики.

Об'єктом дослідження є україномовні тексти, розміщені різними користувачами в соціальних мережах (зокрема Instagram, Facebook та Threads).

Предметом дослідження виступають лінгвістичні маркери емоційності, а також виявлення емоційних забарвлень публікацій у соціальних мережах.

Матеріалом дослідження слугує попередньо сформований корпус текстів, який включає 201 текст на різні теми.

Актуальність теми зумовлена значним обсягом емоційно забарвленого контенту в соціальних мережах; потребою в автоматизованих або автоматичних

інструментах їх аналізу, оскільки ручний аналіз досить трудомісткий; недостатньою кількістю розроблених інструментів для автоматичного аналізу українськомовних текстів, оскільки існуючі моделі емоційного аналізу в основному орієнтовані на англomовний контент і не враховують особливості української мови.

Для досягнення мети були сформульовані такі **завдання**:

1. проаналізувати теоретичні основи емоційного забарвлення тексту та його вплив на комунікацію в соціальних мережах;
2. дослідити роль емоцій у мовленні, зокрема в україномовному сегменті соціальних платформ;
3. проаналізувати сучасні підходи та методи автоматичного аналізу емоцій у текстах;
4. сформувати корпус текстів із соціальних мереж для подальшого аналізу;
5. розробити програму для автоматичного виявлення емоційного забарвлення текстів соціальних мереж українською мовою, впровадити коефіцієнт Яніса для оцінки забарвлення текстів;
6. здійснити тестування й оцінити ефективність розробленого інструменту на прикладі сформованого тексту, а також визначити обмеження програми та перспективи удосконалення цієї програми.

Роботу виконано за допомогою таких методів: метод емпіричного дослідження, аналізу і синтезу, кількісний метод.

Робота складається зі вступу, 3 розділів, висновків, літератури й додатків. 1 розділ має 6 підрозділів, 2 розділ має 3 підрозділи й 3 розділ має 5 підрозділів.

РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ЕМОЦІЙНОГО ЗАБАРВЛЕННЯ ТЕКСТУ

1.1. Роль емоцій у комунікації та їх вплив на інтерпретацію тексту

Емоції впливають на те, як ми сприймаємо світ, як ми думаємо і поведимося у суспільстві. Поняття ЕМОЦІЯ дуже складне і трактується по-різному. В загальнолінгвістичному плані емоція – це переживання людиною свого ставлення до дійсності, до особистого й навколишнього життя; душевне переживання, почуття людини [9]. Існує багато сучасних наук, в яких досліджуються емоції, наприклад психологія, в якій вивчається природа емоцій; соціологія, якій емоції цікаві в контексті соціуму, лінгвістика, в полі зору якої засоби вираження емоцій у мові, філософія, для якої емоції цікаві у контексті теорії пізнання тощо.

За основу дослідження сформульовано таке визначення: емоція – це засіб вираження внутрішнього стану людини через мову. Дане твердження дозволяє розглядати емоції не тільки як внутрішній стан мовця, але і як комунікативну одиницю. Відповідно, методика аналізу передбачає вивчення синтаксичних, лексичних та дискурсивних засобів вираження емоцій.

Є різні класифікації емоцій:

- за характером впливу на людину: емоції бувають позитивні, такі як радість, щастя, задоволення, тобто відчуття, пов'язані із задоволенням потреб людини;
- негативні: страждання, горе, страх, гнів, ненависть – те, що викликає почуття невдоволення і вимагає зміни середовища;
- стенічні та астенічні: стенічні емоції активізують організм людини, спонукають до діяльності й піднімають настрій, наприклад, захват або гнів. Астенічні емоції, наприклад смуток, сором і печаль, пригнічують стан людини й активність організму.

Найчастіше у полі зору дослідників потрапляють такі емоції: щастя, гнів, страх, смуток і огида [21].

Вивчення функціонування емоцій є надзвичайно важливим у лінгвістиці, оскільки вони відіграють ключову роль у комунікації, впливаючи на сприйняття та

інтерпретацію текстів. Крім того, емоції демонструють і навіть допомагають уточнити значення висловлювання та надати йому певного сенсу. Наприклад, вираз *“Чудово, ти знову запізнився!”* є нейтральним виразом, а вираз *“О, чудово, ти знову запізнився! Як завжди.”* є саркастичним та іронічним через додавання вигуку *“О”* та акценті на слові *“знову”*. А вираз *“Як завжди”* ніби підсилює іронічність зауваження. Такі маркери в письмовій формі допомагають передати емоційне забарвлення, яке в усному мовленні забезпечувала б інтонація [34].

Вивчення того, як емоції впливають на міжособистісну комунікацію, має вирішальне значення, адже емоції можуть вплинути на те, як люди сприймають, розуміють та оцінюють їх взаємодію з іншими. На відміну від когнітивних процесів, таких як сприйняття, пам'ять, мислення, емоції відображають реальність у вигляді переживань. Щодня, стикаючись з різними подіями, ми реагуємо на них емоційно, а саме радісно, сумно, позитивно чи негативно. Уміння розпізнавати емоції відкриває шлях до кращого розуміння стану співрозмовника та підтримування ефективної комунікації. Тому так важливо навчитися розуміти емоції інших та будувати своє спілкування так, щоб передати саме позитивний посил. Це проявляється в емпатії [7]. Емпатія – це здатність розуміти емоційний стан іншої людини та співпереживати їй [11]. Саме її наявність забезпечує емоційний комфорт співрозмовників.

В українськомовних текстах емоції виявляються у двох основних формах: наприклад, в художній літературі це емоції, описувані у творі, або які емоції відчуває певний персонаж, друга форма – це емоційні реакції читачів на зміст тексту, що можна спостерігати в текстах медійного стилю. Сприйняття тексту залежить від багатьох чинників, серед яких важливу роль відіграє внутрішній голос та емоційне забарвлення, з яким людина читає. Під час читання людина сама вирішує, з яким емоційним тоном читати текст – негативним чи позитивним. Наприклад, читачі звертаються до поезії, романів чи п'єс для того, щоб отримати певне емоційне задоволення. Позитивний емоційний настрій здатен зробити текст приємнішим для сприйняття, але водночас зменшити критичність оцінювання. Негативні емоції (на кшталт страху або тривоги) можуть звузити коло уваги й змусити концентруватися на конкретних деталях.

Емоційний стан читача безпосередньо впливає на інтерпретацію змісту, наприклад, нейтральний вислів може видатися агресивним, якщо читач перебуває у роздратованому стані. Люди, котрі відчують радість, схильні інтерпретувати текст оптимістично, а ті, хто засмучений, можуть бачити більше негативу навіть у нейтральних висловлюваннях. Тексти, що відповідають актуальному емоційному стану читача, краще сприймаються та запам'ятовуються. Використання емоційно забарвленої лексики чи риторичних прийомів дозволяють посилити або змінити значення тексту, викликаючи певну реакцію у читача [39]. Саме здатність текстів зворушувати, надихати, впливати, дарувати насолоду, співпереживати є їхньою невід'ємною характеристикою [27].

1.2. Способи вираження емоційного стану мовними засобами

Мова — це не просто засіб передачі інформації, а потужний інструмент впливу на емоційний стан, поведінку та соціальну активність людини. Вона тісно пов'язана з психічним, емоційним і навіть фізіологічним станом особистості. За допомогою мови людина входить у соціум, засвоює культурні цінності та норми поведінки, формується не лише свідомість, а й емоційна ідентичність людини [10].

Одним із перших учених, які досліджували вплив мови на емоційний стан людини, був Едвард Сепір. Його гіпотеза полягала в тому, що мова, якою користується людина, формує її мислення та сприйняття реальності, а не навпаки [8]. Наприклад, у японській чи корейській мовах існує високий рівень ввічливості у звертанні до інших, що безпосередньо відбивається на поведінці людей, зокрема у підвищеній увазі до контексту та емоцій співрозмовника.

Мова має не лише раціональну, а й емоційну складову. Лексичні засоби, інтонація, ритм, акценти — усе це впливає на емоційне сприйняття висловлювань. Це особливо яскраво проявляється у таких сферах, як реклама, політика, медіа, де емоційний вплив є цілеспрямованим і навіть маніпулятивним. Правильно підібрані мовні засоби здатні викликати широкий спектр емоцій – від довіри й натхнення до страху або агресії. Мова також може програмувати свідомість: через використання мовних кліше, повторів, імперативних конструкцій формується певна модель поведінки. Негативно забарвлена лексика, токсичні наративи чи мова ворожнечі не

лише формують агресивну поведінку, а й завдають шкоди психіці [10]. Отже, мова є не лише засобом комунікації, але й важливим інструментом для вираження та формування емоцій. Зміни в мовленні свідчать про зміну емоційного стану, і водночас певні мовні стратегії можуть свідомо використовуватися для впливу на емоційний стан іншої особи.

1.3. Емоційна мова у комунікаційних практиках: ЗМІ, реклама, соцмережі

Як було зазначено раніше, емоції відіграють ключову роль у різних сферах життя. Емоційна мова – це висловлювання, наповнене емоційно забарвленими словами та засобами, які передають ставлення до певного фрагмента реальності. У мові це можуть бути:

- органічно емоційні слова (наприклад: *кохання, недруг, боягуз*);
- суфіксально-префіксальні форми (*дівчин-к-а, стеж-ечк-а, най-лют-іш-ий*);
- дитячі слівця (*цяця, їстоньки*).

Водночас різні культури мають свої унікальні емоційні слова, що не мають буквального перекладу, наприклад: *gigil* (філіппінська), *shinrin-yoku* (японська). Науковці розглядають емоційну мову як особливий стиль комунікації, спрямований на трансляцію емоційно забарвлених суджень, що є засобом впливу на аудиторію, підсилюючи або спотворюючи сприйняття реальності [3].

У ЗМІ емоційна мова використовується як потужний інструмент впливу на сприйняття інформації й почуття людей. Емоції, що передаються через різні зображення, звуки та слова, мають вагомий вплив на сприйняття навколишнього світу. Вони викликають у нас певні реакції, впливаючи на наше ставлення до поданої інформації та наше розуміння подій. Медіа намагаються викликати широкий спектр емоцій – від зацікавленості та захоплення до роздратування та страху.

Емоційна наповненість у медіа визначає реакції аудиторії на події, про які ми чуємо або читаємо. Наприклад, позитивні емоції, викликані медіа, підвищують інтерес та віру у можливість змін, тоді як негативні можуть призвести до відчуття безпорадності або навіть агресії.

Люди емоційно залучаються до споживання інформації, емоції посилюють переконливість та запам'ятовуваність контенту. Емоційно забарвлений контент

створює ілюзію співучасті, співчуття до героїв чи ведучих, передбачення розгортання подій у сюжетах (напруження, катарсис) дають змогу глядачеві переживати події, як власні [38]. Прикладом використання емоційної мови у ЗМІ є заголовки новин з емоційним забарвленням: *“Шокуючі подробиці катастрофи”*. У цьому заголовку слова *“шокуючі”* і *“катастрофа”* викликають страх і тривогу. Заголовок *“Смілива жінка врятувала дитину”* викликають повагу та захоплення завдяки слову *“смілива”*.

У рекламі, емоції – це центральний механізм впливу, що визначає сприйняття, пам’ять і поведінкові реакції споживачів. У рекламі використовують такі емоції, як:

- Страх — дієвий, за наявності чіткого розв’язання проблеми.
- Гнів — потенційно може спонукати до активної поведінки (наприклад, захистити себе чи стати сильнішим).
- Вина — викликає бажання виправити ситуацію (наприклад, зробити пожертву).
- Захоплення — привертає увагу, відкриває до нової інформації, формує більш спокійне та уважне ставлення.

Позитивні та негативні емоції мають різні поведінкові наслідки. Високоемоційний контент (захват, гнів, тривога) підвищує залученість, стимулює поширення та реакції аудиторії. Відтак дієвість реклами залежить від правильно обраних емоційних впливів [35].

Прикладом може бути гасло *“Відчуй справжнє щастя разом з нашою кавою!”*, де фраза *“справжнє щастя”* викликає позитивні емоції.

Гасло *“Відкрий для себе нові горизонти разом із нашим смартфоном!”* викликає цікавість.

У цифрову епоху емоції стали важливим фактором поширення контенту в соцмережах. Згідно з дослідженням професора Yoesoep Edhie Rachmad поширення емоційного контенту пояснюється кількома теоріями:

- Emotional Contagion Theory (ECT) – емоції передаються від людини до людини в онлайн-середовищі.
- Social Amplification Theory (SAT) демонструє, як кожен лайк, репост, коментар посилює та розширює охоплення емоційно насиченої інформації.

- Content Virality Framework Theory (CVFT), в якій стверджується, що контент із більшим емоційним впливом має більший шанс стати вірусним.
- User-Driven Viral Diffusion Theory (UDVDT), в якій акцентується увага на користувачеві, який не є пасивним споживачем, а активним учасником розповсюдження емоційного контенту.
- Influencer Impact Theory (ІІТ) доводить, що емоції, поширені лідерами думок, мають значно вищу ефективність і вірусний потенціал. [32]

Сучасні платформи віддають перевагу дописам, які знаходять відгук у конкретній аудиторії. Тому навіть акаунти з невеликим охопленням можуть стати вірусними, якщо їхній контент відповідає запитам користувачів. Глибоке розуміння інтересів і болів підписників дозволяє створювати зміст, який дійсно корисний і цікавий та викликає відповідний емоційний відгук.

1.4. Способи вираження емоцій в різних мовах

Емоції — це загальнолюдський досвід, проте способи, якими ми їх розуміємо, іменуємо й виражаємо, залежать від культури та мови [29]. Відмінність емоцій у різних мовах і культурах полягає не лише в перекладі слів, але й у тому, як ці емоції відчуються, інтерпретуються та проявляються [24]. На прикладі слова “*щасливий*” можна побачити цю різницю між англійською та польською мовами. Станіслав Баранчук у своїй роботі досліджує це питання. Він вважає, що це слово у польській мові, як і в будь-якій слов’янській, має набагато глибше значення. Зазвичай це стан глибокого задоволення від важливих речей: кохання, родини, сенсу життя. Натомість в англійській мові слово “*happy*” має ширше і часто поверхневе значення, використовується як легке задоволення чи зручність. Наприклад, якщо поставити питання англійською мовою “*Is everybody happy?*”, що швидше означає “*чи всім комфортно*”, “*чи всім весело*”. А от польське *szczęśliwy* (як і українське щасливий) має глибше значення — це не щось буденне, а стан глибокого задоволення від важливих речей: кохання, родини, сенсу життя. Тож запитання “*Чи всі щасливі?*” польською прозвучало б як фраза з глибшим сенсом.

Подібну ситуацію можна побачити у випадку слів “*Anxious*” і “*Боюсь*”. В англійському контексті “*anxious*” часто асоціюється із самоконтролем та

психологічним аналізом, вказуючи на тривожність як на об'єкт спостереження. А от польське (й українське) *боюся* — це просте й природне визнання емоції без осуду чи необхідності контролювати її [29].

У різних мовах емоції можна передавати двома способами: методом перекладу і метод мапування. Метод перекладу полягає в тому, щоб знайти еквівалент для терміну у двох мовах, наприклад, англійській та японській. Однак тут існує проблема: ми можемо лише виявити те, що вже закладено в припущеннях перекладу. Наприклад, японське слово, яке ми вважаємо еквівалентом англійського *"angry"*, насправді може мати вужчий або ширший емоційний спектр, ніж англійське слово.

Метод мапування полягає в тому, що дослідники демонструють носіям різних мов ті самі емоційні стимули, наприклад, фото обличчя з певними емоціями, і просять схарактеризувати емоцію, котру вони бачать. Потім порівнюють подібність відповідей. З цього можна зробити такі висновки, що англомовні, іспанські, польські та італійські респонденти надавали схожі відповіді. А от у представників народу Шуар (Амазонія) суттєва частина відповідей була не назвами емоцій, а описами поведінки: *"він посміхається"*, *"він дивиться"*. Їхня система емоцій більше зосереджується на контексті та діях, аніж на внутрішньому почутті, як у західних мовах. [24]

1.5. Вплив тематики, жанру та соціального контексту тексту на точність емоційного аналізу

Тематика тексту прямо впливає на те, як алгоритми емоційного аналізу або людина-аналітик трактує емоційне забарвлення.

Різні методи розпізнавання емоцій демонструють різну ефективність залежно від типу текстів. Наприклад у статті "SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods" авторів Філіпе Н Рібейро, Матеус Араужо, Полліанна Гонсалвес, Маркос Андре Гонсалвес та Фабрісіо Беневенуто, було проведено порівняння 24 сучасних методів аналізу настроїв. Результати показали, що жоден із методів не забезпечував стабільно високої точності на всіх наборах даних, що, своєю чергою, підтверджує, що універсального методу емоційного аналізу не існує.

Окрім тематики, важливу роль відіграють жанр тексту та соціальний контекст його створення [18]. Прикладом може бути метод Vader (Valence Aware Dictionary and sEntiment Reasoner) – словник та інструмент аналізу настроїв на основі правил, спеціально розроблений для текстів, що містять неформальну мову, такими як пости в соціальних мережах і відгуки. На відміну від традиційних методів, VADER краще пристосований до коротких фрагментів тексту, таких як твіти, відгуки про товари або будь-який створений користувачем контент, що може містити сленг, емодзі та аббревіатури. В ньому використано попередньо створений словник слів, пов'язаних зі значеннями настроїв, і набір правил для підрахунку балів настроїв [33]. Водночас, хоча цей інструмент дуже добре працює з текстами з соцмереж, його ефективність значно знижується при застосуванні на інших типах текстів.

Важливо зазначити, що методи, натреновані на текстах певної тематики, наприклад на відгуках про продукти, можуть показувати низьку точність при застосуванні для аналізу коментарів у новинах. Так само метод, який добре працює на рецензіях до фільмів, може бути малоефективним у політичному дискурсі або в технічних оглядах. Тому перед застосуванням певного методу до текстів нового жанру обов'язково потрібно провести попереднє тестування, щоб оцінити коректність результатів [18].

Особливо складним є емоційний аналіз текстів із соціальних мереж. Такі тексти відзначаються великою кількістю скорочень, неформальної мови, сленгу та різних хештегів. Також у соцмережах дуже часто використовують емотикони [12] – графічне зображення людських емоцій у формі спеціального малюнка і письмового знака [4]. Ці особливості суттєво відрізняють мовне середовище соціальних мереж від формальних текстів. Дослідження показали, що додавання емотиконів до набору даних допоможе покращити результати емоційного аналізу. Так, CNN-модель (згортова нейронна мережа) – це різновид мережі глибокого навчання, яка використовує послідовність згорткових та об'єднувальних шарів для вилучення інформації, обробляє вхідні зображення, присвоюючи значення у вигляді ваг і відхилень різним ознакам або об'єктам на зображенні. Потім ці характеристики подаються в один або кілька повністю пов'язаних шарів, де робиться прогноз або

класифікація [16]. Дана модель досягла точності 95.28% на звичайному наборі даних і 95.37% після додавання емотиконів. [12]

1.6. Теоретичні основи класифікації та автоматичного аналізу емоцій у текстах. Методи автоматичного аналізу емоцій.

1.6.1. Теоретичні аспекти класифікації емоцій

Теоретичні аспекти класифікації емоцій відіграють ключову роль у розумінні емоційної багатогранності емоційного досвіду людини. Дослідження емоцій є важливим для глибшого аналізу текстових даних, що у майбутньому стане основою для створення ефективних систем аналізу емоцій з використанням методів машинного навчання та обробки природної мови. Однак класифікація емоцій у текстових даних стикається з низкою труднощів, які суттєво ускладнюють розробку точних та ефективних алгоритмів. Однією з основних проблем є розпізнавання емоцій, що вимагає не лише високої точності алгоритмів, а й глибокого розуміння контексту та суб'єктивності людських емоцій. Емоції сильно залежать від контексту, їх значення може змінюватися залежно від ситуації, культурних особливостей та особистісних характеристик мовця. Додаткову складність становлять такі явища, як іронія та сарказм, які містять приховану емоційну оцінку й потребують спеціального підходу для коректної інтерпретації.

1.6.2. Автоматичний аналіз емоцій у текстах

Автоматичний аналіз емоцій у текстах є одним із пріоритетних напрямів сучасної лінгвістики [27]. Завдяки розвитку NLP було досягнуто суттєвого прогресу у цій сфері, що дозволяє системам автоматично виявляти та класифікувати емоції, які автор прагнув передати у тексті. Машини можуть навчатися на заздалегідь розмічених даних, де емоції вже вказані, або ж після певного навчання можуть самостійно визначати емоції в інших текстах без попередньої розмітки. При такому типі аналізу машина можна виявляти широкий спектр емоцій – від базових, таких як радість, сум, гнів, страх, до складніших, таких як смуток або розчарування. Аналіз емоцій у текстах має широке застосування в різних галузях, зокрема відстеження емоцій у публікаціях, виявлення емоційного підтексту в політичних промовах або

новинах, а також аналіз клієнтських відгуків з метою покращення обслуговування та якості продуктів [6].

1.6.3. Методи автоматичного аналізу емоцій

Одним з підходів автоматичного визначення емоцій є метод на основі словника, також відомий як лексичний метод. Передбачається використання спеціальних словників слів, де кожне слово має оцінку настрою – позитивну чи негативну. Сума або середнє значення цих оцінок дозволяє визначити загальний емоційний тон речення або документа. Для створення словника збираються слова певної мови у систематизованому вигляді з їх контекстами вживання, потім словник оновлюється шляхом пошуку їх синонімів та антонімів [31]. Лексичні методи ґрунтуються на застосуванні заздалегідь підготовлених переліків слів або виразів, що пов'язані з певними емоціями чи настроями [27].

Ще одним підходом є метод на основі ключових слів, в якому поєднуються лінгвістичні правила та аналіз структури речень. Спочатку здійснюється попередня обробка тексту, що включає видалення стоп-слів, токенизацію і лематизацію. Далі виявляються ключові слова, пов'язані з емоціями, кожному реченню призначається відповідна емоційна мітка. Якщо у реченні немає ключового слова, це інтерпретується як відсутність емоційного змісту. Наприклад, є два речення “Сьогодні я склав іспит на відмінно” і “Ура! Сьогодні я склав іспит з відзнакою”. Ці два речення вказують на емоцію радість, проте перше речення з більшою ймовірністю буде не виявлене як речення, в якому є емоція, а друге речення буде виявлене як таке, в якому є емоція радість завдяки слову “Ура” [15],

Метод на основі правил спрямований на створення граматичних та логічних правил для виявлення емоцій у текстах. Для невеликої кількості документів такі правила можуть бути сформовані вручну, але з великою кількістю документів можуть виникнути труднощі. Підхід до побудови правил поєднує методи розпізнавання ключових слів (KR) та лексичної спорідненості [19]. Він починається з попередньої обробки тексту, що включає в себе вилучення слів, POS-тегування, токенизація тощо. Потім за допомогою лінгвістичних та статистичних концепцій формулюються

правила емоційного аналізу. Ці правила застосовуються для розмітки текстових даних емоційними категоріями [15].

Висновки до розділу 1

У першому розділі було зазначено, що емоції відіграють важливу роль у нашій комунікації, впливають на трактування інформації, на міжособистісне спілкування. Емоції визначаються як психофізіологічні стани, які змінюють наше сприйняття дійсності та мають універсальне, культурно специфічне вираження, а також є невід'ємною частиною комунікативного акту, що виконує не лише інформативну, а й впливову та маніпулятивну функцію. Різні культури не тільки по-різному інтерпретують і виражають емоції, але й надають їм різних значень. У мовленні емоції можуть передаватися як вербальними, так і невербальними методами. Застосування лексичних, а також візуальних елементів у контексті, в тому числі цифровому (емодзі, меми, шрифтові маркери) може створити багаторівневу емоційну структуру повідомлення, яка активно вплине на сприйняття й інтерпретацію змісту. Особливо виразно це простежується в сучасних медіапрактиках, де емоційно забарвлений контент у ЗМІ, рекламі та соціальних мережах виконує не лише інформаційну, а й маніпулятивну функцію, часто використовується для залучення уваги аудиторії, формування громадської думки або стимулювання певних поведінкових реакцій. Різні мови по-різному виражають емоції не лише на лексичному, а й на концептуальному рівні. Було доведено, що тематика, жанр і соціальний контекст впливають на точність автоматичного аналізу емоцій. Розвиток NLP і технологій машинного навчання відкриває нові можливості для автоматизації такого аналізу та підтверджує необхідність подальшого вивчення емоційної компоненти мови в міждисциплінарному контексті, а саме з урахуванням даних психології, когнітивної лінгвістики, соціальних комунікацій та цифрового дискурсу.

РОЗДІЛ 2. ПРАКТИЧНІ АСПЕКТИ АВТОМАТИЧНОГО АНАЛІЗУ ЕМОЦІЙ

2.1. Використання сучасних технологій машинного навчання та аналізу тексту для автоматичного визначення емоцій

Використання сучасних технологій машинного навчання для автоматичного розпізнавання емоцій – це одна з провідних тенденцій в аналізі тексту та людино-машинній взаємодії. Оскільки емоції відіграють важливу роль у спілкуванні, їх точне розпізнавання у тексті дозволяє машинам краще розуміти емоційний стан людини, адаптувати свою реакцію та покращувати якість взаємодії.

У межах одного з досліджень було використано текстові дані з соціальних мереж, що містили емоційний контент. Розглядалися шість основних емоцій: любов, радість, здивування, смуток, гнів і страх. Дані були розділені на три підмножини: навчальна множина (для навчання моделі), тестовий набір (для тестування моделі) та валідаційний набір (для оцінки моделі на етапі її гіперпараметрів моделі на етапі її налаштування).

У дослідженні “Detection of emotion by text analysis using machine learning” було проаналізовано декілька підходів до автоматичного визначення емоцій, першим був лексичний підхід, цей підхід можна використовувати як для аналізу настроїв, так і для розпізнавання емоцій. У випадку лексичного підходу перед початком обробки необхідно підготувати якісний словник, який міститиме репрезентативні слова, що будуть досить точно анотовані з числовим ступенем належності до аналізованого класу, наприклад позитивність, токсичність або тип емоції. Проте цей метод має свої мінуси, оскільки для більшості слів належність до певної емоції є нечіткою, а деякі можна віднести до більш ніж одного емоційних класів. Тобто цей підхід передбачає попереднє створення словників – спеціальних слів, які асоціюються з певними емоціями. Наприклад, до списку радісних слів можуть входити такі слова як *щасливий, усмішка, радість*, ці слова зазвичай використовуються для вираження позитивних емоцій. До списку сумних слів входять такі слова як – *сум, печаль, сльози, втрачати*, що зазвичай передають негативні або пригнічені емоційні стани.

Аналогічно створюються словники для інших емоцій, таких як злість, страх, здивування тощо. Надалі ці списки використовуються для автоматичного виявлення емоцій у тексті шляхом аналізу, які саме слова і з яких списків у ньому присутні. Потім машина бере текст, який потрібно проаналізувати, шукає в цьому тексті слова зі списків емоцій і для кожного знайденого слова встановлюється, до якого списку (якої емоції) воно належить. В кінці своєї роботи відбувається підрахунок слів і встановлюється, з якого емоційного списку в тексті найбільше слів. Таким чином, цей метод визначає, яка емоція, найімовірніше, виражена в тексті, просто рахуючи емоційно забарвлені слова, які в ньому є.

Підхід глибокого навчання – це метод машинного навчання, котрий використовує штучні нейронні мережі з великою кількістю шарів для отримання властивостей з «сирих» даних. Замість ручної розробки ознак для навчання моделі, алгоритми глибокого навчання автоматично “вивчають” ієрархію ознак, де кожний наступний шар вивчає складніші абстракції з попереднього. Цей метод є більш точним. Тут слова представляють у вигляді векторів, вони фіксують семантичні зв'язки між словами [26].

У Сірак'юському університеті провели дослідження з виявлення емоцій у Reddit. Reddit — це американський вебсайт, агрегатор соціальних новин, рейтингування контенту та вебфорум. [2] Дослідники Сірак'юського університету створили набір даних “GoEmotions”, в якому зібрали близько 58000 коментарів з цього сайту, кожен з цих коментарів був анотований за 27 емоційними критеріями. Ці критерії містять такі емоції, як захоплення, забава, гнів, роздратування, схвалення, турбота, розгубленість і цікавість. Для спрощення класифікації 27 оригінальних категорій співвіднесені з шістьма базовими емоціями Екмана. В цьому дослідженні були певні виклики, оскільки Reddit-мова — неформальна, здебільшого містить сленг, скорочення, емоційні скорочення та сарказм. Це ускладнює завдання класифікації емоцій, оскільки такі сигнали важко інтерпретувати навіть людині. Також частина коментарів містить декілька емоцій водночас.

Було випробувано різні підходи для класифікації емоцій: машинне навчання; глибоке навчання; ансамблеві методи. З методів машинного навчання було

використано Logistic Regression, Naive Bayes, SVM, XGBoost тощо, з глибокого навчання LSTM та з ансамблевих методів - Voting, Bagging, Stacking.

Серед них Stacking Classifier, що поєднує декілька моделей із метакласифікатором (Logistic Regression), продемонстрував найкращі результати з огляду на точність та F1-метрику. Для демонстрації практичності було створено вебзастосунок на Streamlit, що дозволяє в реальному часі вводити текст та здобувати емоцію з відповідною ймовірністю та емоذجі [28].

2.2. Аналіз поширених проблем та помилок при автоматичному виявленні емоцій

Автоматичне розпізнавання емоцій – важке завдання, яке має ряд проблем і хиб на різних стадіях обробки даних, аналізу тексту, зображення чи звуку. Одна з основних проблем – це двозначність та контекстуальність мови. Наприклад, сарказм, іронія, гумор чи навіть певні культурні особливості мови можуть збивати алгоритми визначення емоцій [13], оскільки вони змінюють прямий зміст висловлювання. Наприклад: *“Обожнюю стояти в заторі годинами!”* — позитивний вислів, на перший погляд, але фактично негативний. Унікальне рішення — корпус iSarcasm, де автори самі позначають, що саме вони мали на увазі, тобто маркування відбувається на основі передбаченого сарказму. До 2020 року набір даних формувався лише на основі дописів у твітері, проте пізніше дані почали брати з новин, книг, YouTube і Reddit. Підхід, коли анотацію здійснювали самі автори, дозволяє краще навчити моделі розрізняти справжній сарказм, адже він заснований на тому, що людина хотіла дійсно висловити, а не на тому, як це здалося читачеві. Цей датасет зацікавив нас тим, що, на відміну від більшості інших, у ньому тексти анотовані самими авторами. У звичайних датасетах це роблять сторонні анотатори, які можуть неправильно інтерпретувати авторський задум, що часто призводить до помилок. Натомість підхід, реалізований у iSarcasm, де саме автори вказують, що саме мали на увазі, робить цей ресурс еталонним для дослідження сарказму в текстах [14].

Наступною проблемою автоматичного емоційного аналізу є те, що емоції часто виражаються непрямо, а значення слів залежать від контексту, який необхідно враховувати. Тексти можуть одночасно передавати кілька емоцій одночасно, і моделі

не завжди здатні правильно їх розпізнати. Особливо це стосується лексичних підходів, що використовуються в багатьох базах даних: вони часто мають обмежені словники, не враховують синоніми й полісемію, а також цей підхід має труднощі з обробкою нових або маловживаних слів, скороченнями і сленгом.

Додатковою складністю є мовна обмеженість більшості існуючих систем. Вони орієнтовані переважно на англійську мову, яка домінує в сучасному цифровому просторі. Для інших мов бракує якісних корпусів, адаптованих моделей, лексиконів та морфологічних інструментів. У результаті автоматичні моделі, натреновані на англійськомовному контенті, не можуть бути ефективно застосовані до інших мов без адаптації.

Ще одна проблема при автоматичному визначенні емоцій – це обмеженість емоцій. Деякі емоції зустрічаються частіше, наприклад, радість чи сум, тоді як інші – рідше (огода, страх). Крім того, анотації цих даних можуть бути доволі суб'єктивними, якщо вони анотуються людьми, оскільки люди мають різні інтерпретації емоцій у тексті.

Крім того, якість результатів значною мірою залежить від попередньої обробки тексту. Помилки на цьому етапі, зокрема некоректна токенізація, нормалізація, очищення тексту, можуть призводити до втрати важливої семантичної інформації. Наприклад, неправильне розпізнавання емодзі, скорочень чи інтернет-сленгу здатне суттєво спотворити результати емоційного аналізу, особливо в неформальних текстах, які є характерними для соціальних мереж [13].

Автоматичне виявлення емоцій стикається з низкою викликів — від мовної невизначеності та обмежених ресурсів до складності контексту. Покращення можливе через якісніші, багатомовні корпуси; врахування контексту автора й ситуації; застосування мультимодальних та трансформерних моделей; розгляд культурної специфіки відчуття емоцій [14].

2.3. Огляд та дослідження наявних баз даних (WordNet, EmoLex, General Inquirer)

Існує дуже багато лексичних баз даних. Розглянемо найбільш поширені.

WordNet — це онлайн-лексична платформа англійської мови, створена для застосування програмами обробки природної мови. Вона поєднує традиційну лексикографічну інформацію з можливостями сучасної обчислювальної технології [20]. WordNet також є вільною та публічно доступною для завантаження. Структура WordNet робить її корисним інструментом для комп'ютерної лінгвістики та обробки природної мови. WordNet зовні нагадує тезаурус, оскільки групує слова на підставі їх значень. Основним зв'язком між словами у WordNet є синонімія. Синоніми – слова, які позначають те ж саме поняття і є взаємозамінними в багатьох контекстах, згруповані в неупорядковані множини – синсети. Кожен з 117 000 синонімів WordNet пов'язаний з іншими синонімами за допомогою невеликої кількості “концептуальних зв'язків”. Крім того, синоніми супроводжуються короткими визначеннями, містять одне або кілька коротких речень, які ілюструють вживання членів синонімічного ряду [40].

Проте, у WordNet є певні проблеми:

1. Визначення значення слова залежить від контексту.
2. WordNet лише надає варіанти значень.

Для розв'язання цієї проблеми створюються семантичні конкорданси — корпуси текстів, у яких кожне слово пов'язане з конкретним значенням у WordNet [20].

EMOLEX (NRC Word-Emotion Association Lexicon) — це великий науковий проєкт, спрямований на вивчення лексики емоцій у п'яти європейських мовах: французькій, німецькій, англійській, російській та іспанській [23]. База даних включає 8 базових емоцій: гнів, страх, очікування, довіра, здивування, смуток, радість і відраза та двома почуттями негативним і позитивним [25]. Його мета — збудувати багатомовну базу даних емоційної лексики, розбираючи слова за семантичними, синтаксичними та дискурсивними рисами. Усе базується на великих корпусах — понад 700 млн слів у журналістських і літературних текстах. EMOLEX містить відкриту багатомовну базу даних EmoBase і складається з 3 модулів: EmoConc — запити до корпусів; EmoLing — статистика та порівняння емоційних слів; EmoProf — навчальний модуль для викладачів французької як іноземної.

Основними цілями EMOLEX є виділення лексико-синтаксичних патернів, що виражають емоції. Вони поєднують два підходи – належність до певного концепту і вживаність у текстах [23].

Набір даних TESS (Toronto Emotional Speech Set) – це набір даних для класифікації тренувальних емоцій, складається із записів 200 речень, вимовлених акторами. Цікаво те, що цей набір даних лише жіночий і має дуже якісний звук. Фрази виражають різні емоції, зокрема щастя, смуток, гнів, страх, здивування та огиду. Набір даних організований таким чином, що емоції двох акторок, які начитували текст, містяться в окремих папках, що дає можливість знайти аудіофайл усіх 200 цільових слів. Набір даних містить записи різних груп акторів, що забезпечує широкий спектр емоційних проявів. Кожен актор записав однаковий набір речень, що дозволяє послідовно порівнювати різні емоційні стани. Емоції, представлені в наборі даних, ретельно відібрані, щоб охопити широкий спектр, що важливо для навчання надійних моделей розпізнавання емоцій. Набір даних TESS широко використовується в різних додатках, наприклад:

- при навчанні моделей машинного навчання для розпізнавання емоцій на основі мовних патернів;
- для підвищення емоційної виразності синтезованого мовлення у віртуальних помічниках та інших додатках;
- вивчення нюансів емоційної експресії та її впливу на комунікацію.

Набір даних призначений для полегшення вивчення емоційної експресії в мовленні, що робить його особливо корисним для розробки та тестування систем розпізнавання емоцій.

Важливо підкреслити, що набір даних TESS створено спеціально для аналізу усного мовлення, а не тексту. Записи в наборі відображають живу емоційну інтонацію, темп, ритм і акустичні характеристики, притаманні саме усному мовленню [36].

Також існує база даних з унікальним корпусом соціальних мереж українською мовою, яка називається Ukrainian Resilience Dataset, дані для цієї бази даних були зібрані під час війни між Україною та Росією. Основним призначенням цієї бази є

виявлення публікацій у соцмережах, в яких люди шукають допомоги під час війни. Такі публікації шукаються за ключовими словами, наприклад “війна”, “допомога”. Ці пости класифікують за такими оцінками: 1 – є заклики по допомогу, 0 – немає цього заклику. Анотація відбувалася за таким принципом: кожен пост був перевірений двома анотаторами, якщо оцінки збігалися, то ця оцінка приймалася [30].

Висновки до розділу 2

У другому розділі було досліджено важливі практичні аспекти застосування сучасних технологій машинного навчання для автоматичного визначення емоцій у текстах, що сформувало новий напрям сучасної лінгвістики, штучного інтелекту та взаємодії людини й машини. Проаналізовано як класичні лексичні методи, так і сучасні підходи, які засновані на глибокому навчанні та трансформерах. Справжнє розпізнавання емоцій стає можливим, коли поєднуються високоякісні дані, потужні алгоритми та методи, чутливі до контексту та культури. Глибокі нейронні мережі дозволяють досягти значно вищої точності в задачах класифікації емоцій, особливо в умовах складного контексту, багатозначності та мультиемоційності тексту. Розвиток таких технологій має практичне значення в царинах медіа, маркетингу, освіти, соціальних мереж, розвитку емоційно чутливих віртуальних помічників. Проте їх успішна реалізація вимагає подолання основних викликів: технічних, мовних та семантичних перешкод через постійні дослідження, адаптацію до нових мов і жанрів та етичну відповідальність у їх застосуванні. Подолання цих викликів можливе шляхом використання мультимодальних моделей, багатомовних корпусів, а також урахування культурної специфіки.

Сучасні досягнення у сфері цифрової лінгвістики та штучного інтелекту відкривають нам нові перспективи для глибокого розуміння емоцій у мові, що є надзвичайно актуальним для розвитку людино-машинної взаємодії, психологічних досліджень і соціальних комунікацій.

РОЗДІЛ 3. ОЦІНКА ЯКОСТІ АВТОМАТИЧНОГО АНАЛІЗУ ЕМОЦІЙНОСТІ ТЕКСТІВ СОЦМЕРЕЖ

3.1. Формування корпусу текстів для аналізу

Формування репрезентативного та якісного корпусу текстів є ключовим етапом будь-якого дослідження, націленого на аналіз мовних даних. Для проведення свого дослідження було взято публікації з трьох актуальних соціальних платформ: Instagram, Facebook та Threads.

Instagram [22] – це соцмережа, в якій основна увага приділяється фотографіям та відеоконтенту, проте тут також є підписи під фото та відео і коментарі. В цих текстових фрагментах користувачі виражають власну думку, а також поєднують розмовну мову, інколи формальну. Ці тексти увійшли до корпусу.

Facebook [17] – платформа, орієнтована на ширше коло тем, включаючи в себе особисті думки й публічні обговорення.

Threads [37] – це новітня платформа, чимось подібна до Twitter, в основному тут переважають короткі публікації. У цих текстах акцент робиться на висловленні власної думки людей, часто з високим ступенем емоційної насиченості, а також мовних явищ, характерних для цифрового дискурсу, наприклад сленг, емодзі тощо.

Ці платформи обрано з огляду на їх популярність серед користувачів та значний обсяг україномовного контенту. Особливу увагу приділено саме неформальним текстам, які створені звичайними користувачами, що дозволяло простежити актуальні мовні тенденції.

До фінальної версії корпусу було включено 201 текст, який ретельно відібрано вручну. Тексти відрізняються за обсягом, жанром, тематикою та стилем. Основними критеріями вибору текстів були:

- наявність емоційно забарвленої лексики, як позитивної, так і негативної;
- тематичне різноманіття, що включає тексти з особистими переживаннями з різних політичних тем;
- лише україномовний контент без перекладу.

В результаті корпус склав 18302 слововживання, що є репрезентативним обсягом для проведення якісного аналізу. Зібраний матеріал є надійною основою для подальшого аналізу емотивної лексики та висвітлення певних особливостей цифрового дискурсу. З повною версією корпусу текстів можна ознайомитися у Додатку 1.

3.2. Коефіцієнт Яніса як індикатор емоційності

Для дослідження емоційного змісту мовлення, зокрема в межах контент-аналізу, існують різні процедури для кількісного підрахунку результатів. У нашому дослідженні буде використано коефіцієнт Яніса. Цей коефіцієнт дає змогу обчислити співвідношення позитивних і негативних висловлювань у межах певної тематичної площини. У контексті нашої роботи ми будемо застосовувати коефіцієнт Яніса для визначення емоційності у текстах зібраних з різних соцмереж: Instagram, Facebook та Threads. Цей коефіцієнт можна обчислити за такими формулами:

- якщо, кількість позитивних оцінок перевищує кількість негативних, коефіцієнт ми рахуємо за такою формулою

$$c = \frac{f^2 - f \cdot n}{r \cdot t}$$

- якщо, переважають негативні оцінки, тоді обчислюємо за такою формулою

$$c = \frac{f \cdot n - n^2}{r \cdot t}$$

де, f - це кількість позитивних оцінок; n – кількість негативних оцінок; r – кількість емоційно забарвлених слів; t - загальна кількість слів у тексті, який досліджується.

Інтерпретація коефіцієнта Яніса залежить від його числового значення та співвідношенню між кількістю позитивних і негативних слів у тексті. Інтерпретувати результати можна так: якщо значення коефіцієнта > 0 , тоді переважає позитивна емоційність, чим вище значення, тим сильніше текст насичений позитивними словами.

Якщо < 0 , тоді у тексті переважає негативна емоційність: чим нижче значення, тим більш негативно забарвлений текст.

Якщо ≈ 0 , то такий результат свідчить про те, що текст є нейтральним або має емоційний баланс (кількість позитивних і негативних висловлювань приблизно однакова) [5].

Коефіцієнт Яніса буде інтегровано до функціоналу програми, розробленої в межах цього дослідження. Його застосування дозволить не лише виявляти емоційно забарвлену лексику у текстах, але й автоматично обчислювати ступінь емоційної поляризації повідомлення або публікації на основі кількісного співвідношення позитивних і негативних оцінок.

Застосування коефіцієнта Яніса в програмному середовищі дасть змогу формалізувати підсумки контент-аналізу та забезпечити об'єктивну метрику емоційності для кожного тексту в корпусі. Це буде важливим кроком на шляху до створення повноцінного інструменту для автоматизованого аналізу цифрового дискурсу.

3.3. Розробка та застосування програми для автоматичного виявлення емотивних слів

Метою даного етапу дослідження є створення програмного інструменту, який дозволяє автоматизовано ідентифікувати емотивно забарвлені слова в українськомовних текстах у соцмережах. Програма має забезпечувати швидкий аналіз великих обсягів даних із виведенням результатів, які можуть бути використані для подальших лінгвістичних досліджень. Основними завданнями, які має виконувати програма, є:

- розробка словника емотивної лексики;
- створення алгоритму аналізу текстів на наявність емотивних слів;
- реалізація графічного інтерфейсу для завантаження текстів і перегляду результатів;
- тестування програми на корпусі текстів.

Для виявлення емотивної лексики обрано словниковий підхід, коли кожне слово зіставляється зі списком емотивних слів, поділеним на позитивні (рис.1) й

негативні (рис.2) елементи. З повною версією словників можна ознайомитися у Додатку 2.

```
"Гарних",  
"Спокій",  
"абсолютно впевнений",  
"адекватна",  
"адекватна людина",  
"адекватний",  
"адекватно",  
"адекватный",  
"адекватні",  
"аналізую",  
"ангел",  
"ангела",  
"аромат",  
"атмосфера",  
"атмосферу",  
"багато",
```

Рисунок 1. Словник з позитивними словами

```
"«не пхай носа»",  
"ЖАХ",  
"Лицемірство",  
"НЕідеально",  
"оце",  
"Цькування",  
"абияк",  
"абсурд",  
"аб'юз",  
"агресивний",  
"агресія",  
"агресії",
```

Рисунок 2. Словник з негативними словами

Такий підхід дозволяє точно ідентифікувати заздалегідь визначені емоційно забарвлені одиниці без залучення машинного навчання, що є доцільним на початковому етапі дослідження. Словник було сформовано на основі ручного визначення емотивних слів.

Програма виконує такі функції:

- завантаження тексту з постами;

- аналіз кожного посту на наявність позитивних та негативних слів;
- підрахунок коефіцієнта Яніса;
- виведення результату в інтерфейс;
- додавання нових слів до позитивного або негативного словника;
- візуалізація словникових даних за допомогою графіка;
- збереження результатів у файлі.

Код працює на основі імпорту модулів, таких як:

- tkinter – для побудови графічного інтерфейсу користувача;
- ttkbootstrap – для дизайну інтерфейсу;
- filedialog, messagebox, scrolledtext – компонентів інтерфейсу (вікна для відкриття файлів, повідомлення, прокручування тексту);
- re – для очищення тексту регулярними виразами;
- matplotlib.pyplot — для побудови графіків;
- json – для роботи зі словниками;
- os — для перевірки наявності файлів.

Далі створюються файли словників у форматі JSON – текстовому форматі обміну даними між комп'ютерами. JSON базується на тексті, може бути прочитаним людиною. Формат дає змогу описувати об'єкти та інші структури даних. Цей формат використовується переважно для передавання структурованої інформації через мережу [1]. У цих файлах зберігаються списки слів що мають позитивне або негативне емоційне забарвлення (рис. 3). Програма використовує ці файли для того, щоб визначити настрій тексту під час аналізу.

```
# Шлях до словників
POSITIVE_WORDS_FILE = "positive_words.json"
NEGATIVE_WORDS_FILE = "negative_words.json"
```

Рисунок 3. Фрагмент коду для збереження лексики з емоційним забарвленням у вигляді окремих файлів (позитивна та негативна)

Якщо файл існує, то відкривається і завантажується список слів, далі слова зберігаються у файлі у форматі JSON.

Далі код обробляє тексти: видаляє зайві символи, зводить все до нижнього регістру, токенізує слова і залишає лише ті слова, що містять лише літери без цифр (рис. 4).

```
def preprocess(text):  
    text = re.sub(pattern: r'[\W\s]', repl: ' ', text)  
    text = re.sub(pattern: r'\s+', repl: ' ', text)  
    text = text.lower()  
    words = text.split()  
    return [word for word in words if word.isalpha()]
```

Рисунок 4. Фрагмент коду для попередньої обробки текстів: видалення зайвих символів, перетворення до нижнього регістру, токенізація та фільтрація лише літерних слів

Програма аналізує емоції за допомогою коефіцієнта Яніса, який показує емоційне забарвлення. Далі зчитується текстовий файл у форматі .txt, де кожен пост відокремлений двома новими рядками. Для кожного з цих постів обчислюється позитивні й негативні слова, коефіцієнт Яніса і результат зберігається. Всі результати можна зберегти у .txt файл.

Програма після запуску створює графічний інтерфейс. Створюються кнопки: “Завантажити текстовий файл”, “Зберегти результати”, “Побудувати графік”, “Додати слово”. Також є поле для введення слова, вибір позитивного чи негативного типу цього слова і вікно виводу результатів (рис. 5).

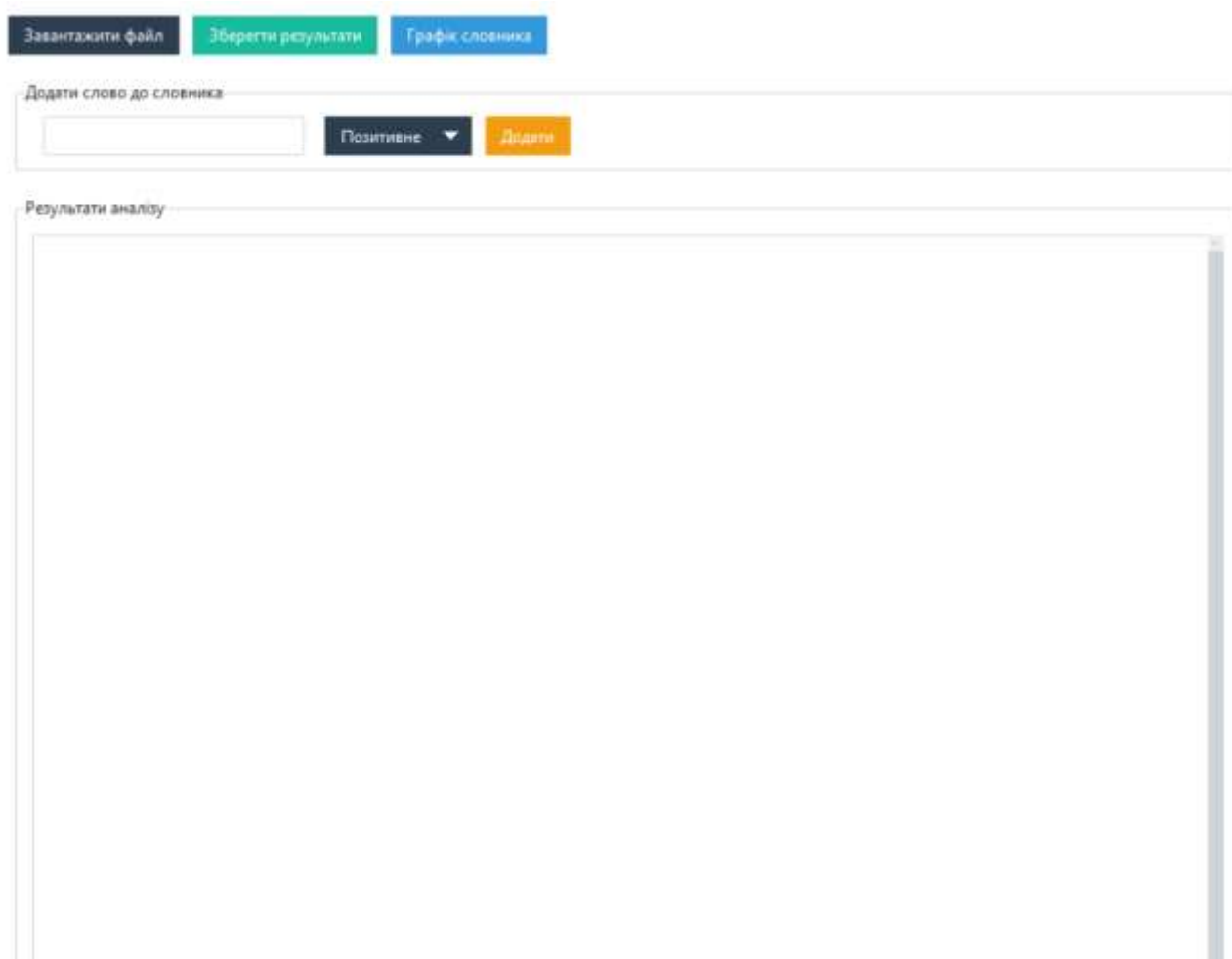


Рисунок 5. Вікно графічного інтерфейсу користувача, що містить функціональні елементи: кнопки керування (завантаження, збереження, побудова графіка, додавання слова), поле введення слова, вибір типу емоції та вивід результатів

Якщо слово відсутнє у словнику, його можна додати вручну через графічний інтерфейс через команду `def add_word()`: для цього слова можна вибрати категорію – позитивну чи негативну. Вона додає лематизоване слово до позитивного або негативного словника та зберігає його там (рис. 6).

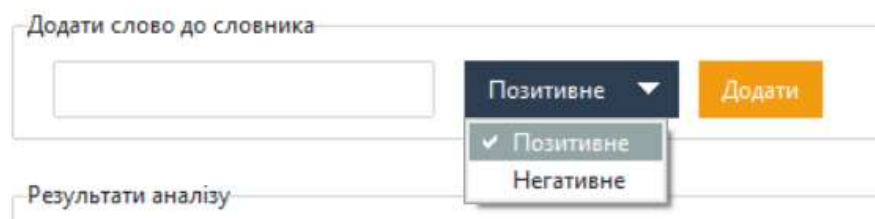


Рисунок 6. Механізм додавання слова до позитивного або негативного словника через графічний інтерфейс

Коли слово додано у словник то з'являється ось таке повідомлення (рис. 7).

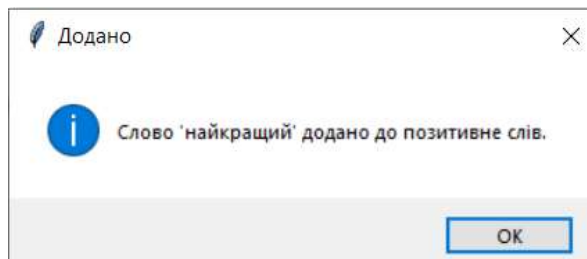


Рисунок 7. Повідомлення про успішне додавання слова до позитивного або негативного словника через графічний інтерфейс

Також можна побудувати графік який буде показувати кількість слів у кожному словнику (рис. 8).

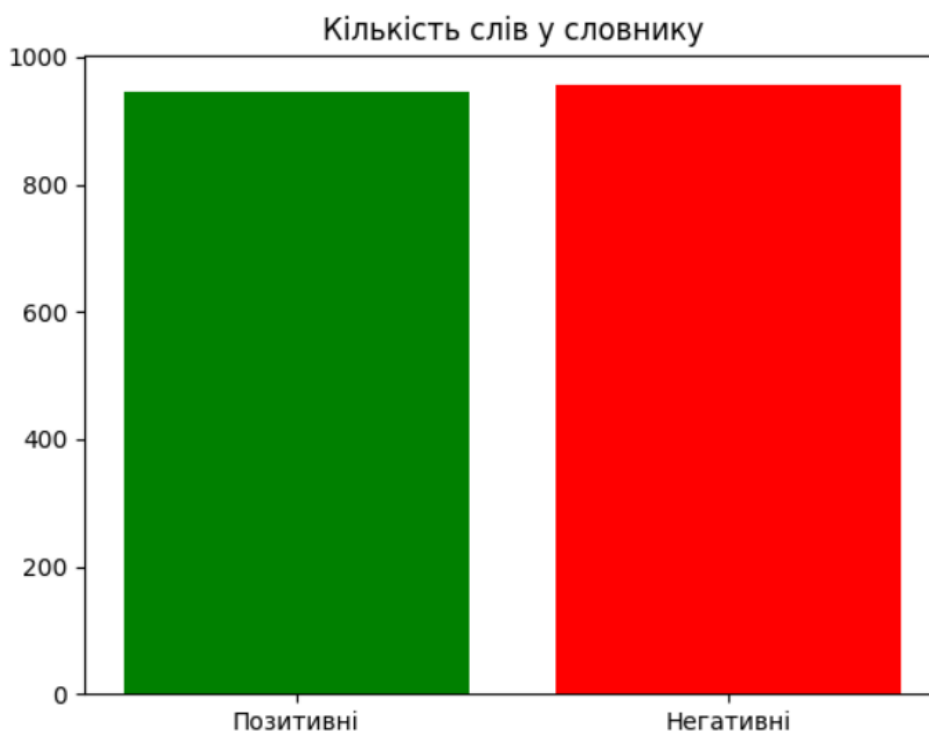


Рисунок 8. Графік кількості слів у позитивному та негативному словниках

У результаті розробки було створено програму, яка здатна автоматично виявляти емотивні слова в тексті (рис. 9).

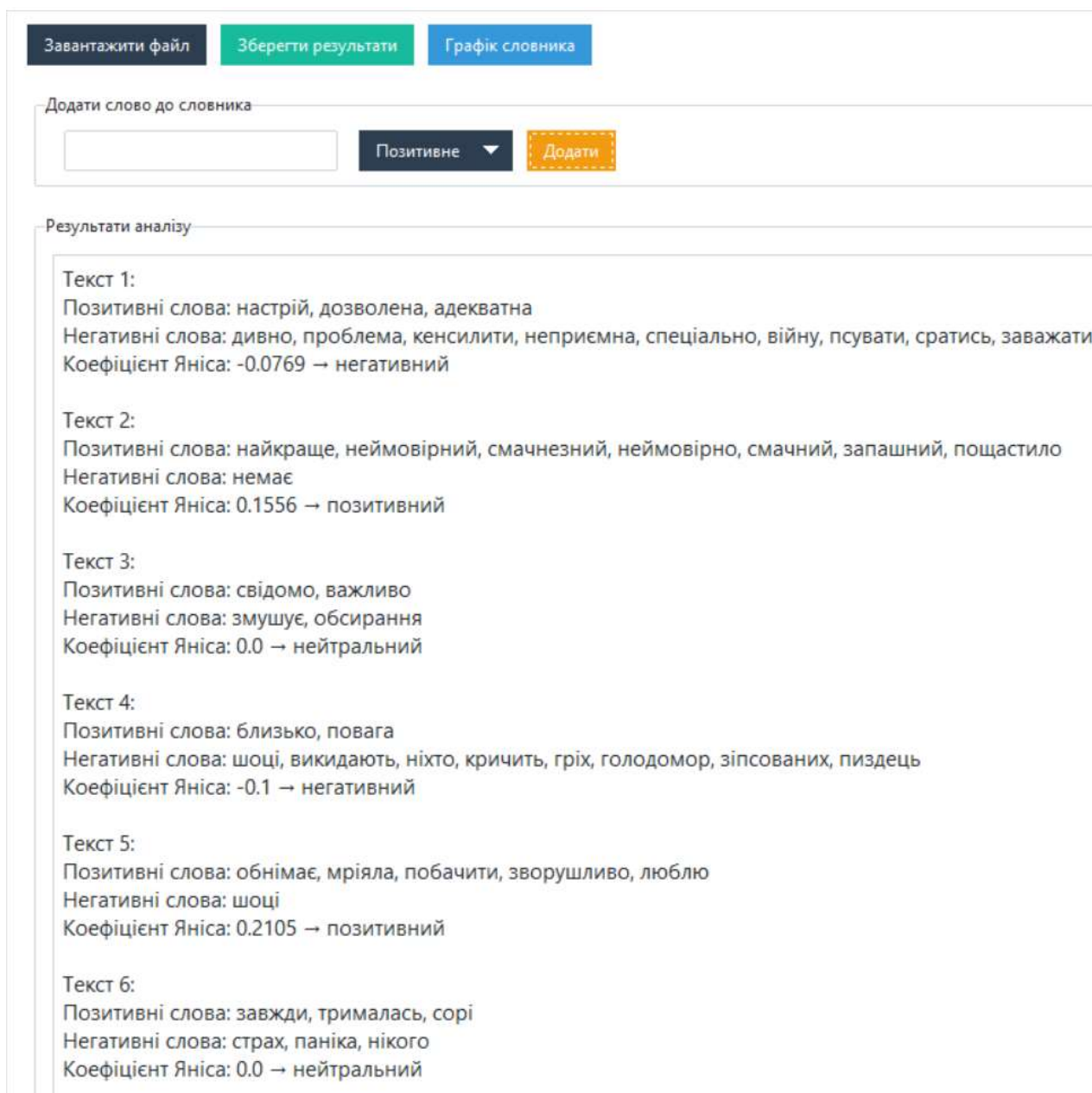


Рисунок 9. Приклад роботи програми з автоматичного виявлення емотивної лексики в текстових даних

Як зазначалося вище, результат роботи програми може бути збережений у вигляді текстового файлу. Це дає змогу зафіксувати отримані дані для подальшого аналізу, зберігати історію оброблених текстів або здійснювати повторне звернення до результатів без необхідності повторного запуску програми. Після успішного збереження файлу на екрані з'являється повідомлення, яке інформує користувача про завершення операції (рис.10). На рисунку 11 наведено приклад збереженого файлу з результатами аналізу у форматі txt.

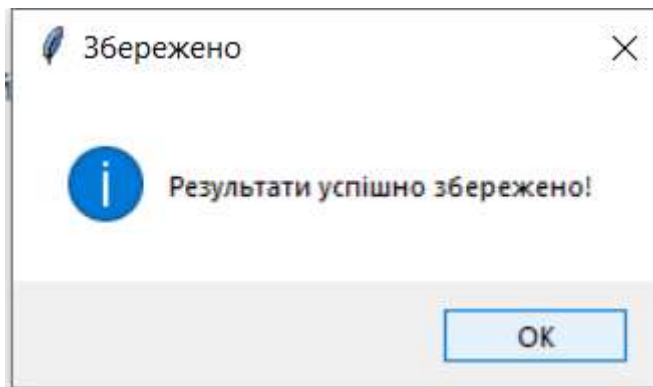


Рисунок 10. Повідомлення про успішне збереження результатів емоційного аналізу у текстовий файл

```
Текст 1:  
Позитивні слова: настрій, дозволена, адекватна  
Негативні слова: дивно, проблема, кенсилити, неприємна, спеціально, війну, псувати, сратись, заважати  
Коефіцієнт Яніса: -0.0769 → негативний  
  
Текст 2:  
Позитивні слова: найкраще, неймовірний, смачнезний, неймовірно, смачний, запашний, пощастило  
Негативні слова: немає  
Коефіцієнт Яніса: 0.1556 → позитивний  
  
Текст 3:  
Позитивні слова: свідомо, важливо  
Негативні слова: змушує, обсирання  
Коефіцієнт Яніса: 0.0 → нейтральний  
  
Текст 4:  
Позитивні слова: близько, повага  
Негативні слова: щоці, викидають, ніхто, кричить, гріх, голодомор, зіпсованих, пиздешь  
Коефіцієнт Яніса: -0.1 → негативний  
  
Текст 5:  
Позитивні слова: обнімає, мріяла, побачити, зворушливо, люблю  
Негативні слова: щоці  
Коефіцієнт Яніса: 0.2105 → позитивний  
  
Текст 6:  
Позитивні слова: завжди, трималась, сорі  
Негативні слова: страх, паніка, нікого  
Коефіцієнт Яніса: 0.0 → нейтральний  
  
Текст 7:  
Позитивні слова: особливого, щира, чесна, кумиром  
Негативні слова: нічого, вивертало, матюкається, пиячить, огидно, огидно, гірше  
Коефіцієнт Яніса: -0.0526 → негативний
```

Рисунок 11. Приклад текстового файлу з результатами емоційного аналізу

Повну версію коду можна переглянути в Додатку 3.

3.4. Виявлення недоліків і обмежень у програмі і їх вирішення

У процесі розробки та тестування програми для автоматичного емоційного аналізу постів із соціальних мереж було виявлено низку недоліків та обмежень у програмі, які впливали на точність результатів роботи програми.

Перша проблема – це обмежений базовий словник емоційної лексики. На початковому етапі роботи програма орієнтувалася лише на базовий набір позитивних і негативних слів. Цей набір був обмеженим та не охоплював специфічну лексику, яка є притаманна для соцмереж (різний сленг, аббревіатури, запозичення з різних мов, суржик, терміни та скорочення). Унаслідок цього частина емоційно забарвлених слів залишалася нерозпізнаною, що знижувало точність визначення емоційного тону текстів. Для розв'язання цієї проблеми було прийнято рішення впровадити власний словник, який користувач може редагувати вручну, додаючи нові слова до позитивної або негативної категорії. Це значно підвищило адаптивність програми до живої, розмовної мови, а також до молодіжного та інтернет жаргону. Словники можна знайти у Додатку 2.

Друга проблема – це неправильне розпізнавання кількості постів, що було суттєво. Тобто, програма не розпізнавала всі тексти у вхідному файлі. Замість очікуваних 200 постів, було оброблено лише 107. Після перегляду текстового файлу з постами з'ясувалося, що деякі тексти не були розділені за допомогою порожнього рядка. І тому тексти, які не мали такого розділення, сприймалися програмою, як один об'єднаний фрагмент. Рішенням стало - додавання порожніх рядків між кожним текстом відповідно до вимог програми.

На початковому етапі розробки програми для обробки текстових даних було вирішено застосувати бібліотеку `rumorphy2` для лематизації слів. Однак у процесі тестування було виявлено кілька суттєвих недоліків:

- помилкова лематизація великої кількості слів. Наприклад, деякі слова було зведено до хибної леми;
- недостатня адаптованість до сучасної української мови;
- помилкове розпізнавання деяких українських слів як російських. У багатьох випадках українські слова лематизувалися не в українському варіанті, а в російському.

Після аналізу результатів у зв'язку з цими обмеженнями було прийнято рішення відмовитися від використання `rumorphy` у проєкті.

Ще однією складністю є те, що програма орієнтована лише на українські слова. І якщо в постах трапляються слова іншими мовами, наприклад - англійські слова, суржик або якісь сленгові слова, вони не враховуються. Але їх можна додати вручну, через програму.

Також, програма не враховує контекст і тому деякі слова, які можуть вживатися і в позитивному значенні, і в негативному, не розпізнаються. Наприклад, в тексті 122 є фраза “Я в шоці”, програма розпізнала слово “в шоці” як негативне, хоча, якщо брати до уваги контекст, то ця фраза тут має позитивне значення. Такий самий приклад можна побачити в тексті 7 зі словом “кумир”, яке саме по собі має позитивне значення, але в нашому контексті воно є негативним.

3.5. Оптимізація та покращення точності аналізу у майбутньому

Подальший розвиток програми емоційного аналізу текстів із соціальних мереж передбачає впровадження більш гнучких методів обробки природної мови, що дозволить значно підвищити точність результатів та адаптивність до різноманітних комунікативних ситуацій. Наприклад, можна застосувати сучасні мовні моделі на основі нейронних мереж, наприклад GPT. Це дозволить аналізувати не тільки слова за словником, але й контекстуальну складову, а також виявляти складні емоційні конструкції, такі як сарказм, іронія або приховане емоційне забарвлення.

Щоб уникнути швидкого застаріння ручного словника, доцільно реалізувати механізм його автоматичного оновлення на основі аналізу великої кількості текстів. Тоді б програма могла б виявити нові емоційно забарвлені слова.

Також можна інтегрувати різні емодзі в систему, оскільки вони часто змінюють або підсилюють емоційні висловлення, а їх інтеграція в алгоритм сприятиме точнішому аналізу настрою користувача. Для цього можна буде створити окремий емоційний словник з цими емодзі.

У майбутньому варто було б реалізувати підтримку змішання слів з різних мов для постів, у яких є англійські, французькі, німецькі тощо слова, а також суржик, який широко використовується у соціальних мережах. Це дозволить краще охопити ширший спектр комунікації та зробити аналіз кращим.

Розширення аналізу до семантичного рівня, тобто до значення слів у контексті, дозволить точніше оцінювати загальний зміст повідомлення.

Висновки до 3 розділу

У межах третього розділу було описано автоматичний аналіз емоційного забарвлення текстів соціальних мереж. Сформовано корпус текстів, який відображає різноманітність онлайн комунікації і який є основою для подальшої автоматизованої обробки на предмет виявлення емоційно забарвленої лексики.

Важливим елементом став розрахунок коефіцієнта Яніса – спеціального індикатора, що дозволяє кількісно оцінити рівень емоційності тексту з урахуванням частоти позитивних і негативних лексем.

Також була реалізована програма з графічним інтерфейсом з автоматичного пошуку емотивних слів. Крім того, користувачеві надається можливість розширювати словник позитивної та негативної лексики. У ході тестування цієї програми було виділено декілька недоліків й окреслено перспективи удосконалення системи, зокрема через впровадження нейромережових моделей.

ВИСНОВКИ

В даній кваліфікаційній роботі було здійснено комплексне дослідження теоретичних та практичних аспектів автоматичного аналізу емоційного забарвлення україномовних текстів соціальних мереж. Актуальність теми зумовлена зростанням обсягів цифрової комунікації та необхідністю ефективного аналізу емоцій у текстовому контенті, що є важливим як для лінгвістики, так і для суміжних дисциплін, наприклад психології, соціології, філософії.

Встановлено, що емоції відіграють важливу роль у письмовій комунікації, а також впливають на сприйняття, інтерпретацію тексту та реакцію читача на цей текст. Емоційно насичений текст здатен не лише передавати якусь емоцію, а також може формувати ставлення, викликати довіру у читача, об'єднати або роз'єднати аудиторію.

У теоретичній частині роботи було проаналізовано ключові аспекти емоцій у мовленні. Було з'ясовано, що емоції є ключовим елементом комунікації. Уміння правильно розпізнавати емоції у письмовому мовленні має вирішальне значення для ефективної взаємодії, особливо актуальним є це в умовах цифрової комунікації.

Розглянуто вплив тематики та жанру на точність емоційного аналізу, що є важливим при роботі з текстами у соціальних мереж. Також наголошено, що емоції інтерпретуються по-різному у різних мовах і культурах. І виокремлено основні методи автоматичного аналізу емоцій, зокрема:

- словникові методи;
- методи на основі ключових слів;
- метод правил.

Крім того, окреслено складності у класифікації емоцій, зокрема врахування полісемії, іронії, сарказму.

Окрему увагу приділено практичним аспектам автоматизованої та автоматичної емоційної обробки текстів, розглянуто наявні бази даних. Здійснено аналіз типових помилок і викликів в автоматичному розпізнаванні емоцій. Загалом, сучасний автоматичний аналіз емоційних станів демонструє значні досягнення, проте залишається далеким від повної інтерпретаційної гнучкості, притаманної людині.

Існують виклики, пов'язані з багатозначністю, культурними кодами, вираженням складних або змішаних емоцій.

Практична частина дослідження була присвячена реалізації програми автоматичного аналізу емоційного забарвлення українських текстів соціальних мереж. Цей процес включав у собі такі етапи:

- по-перше, створено корпус текстів, цей етап став фундаментом цього дослідження. Було зібрано 201 текст з трьох соціальних мереж, ці тексти відображають сучасний стан мовлення у соцмережах, мають неформальну мову, сленг, а також високий ступінь емоційної насиченості текстів.
- по-друге, розроблено програму автоматичного визначення емотивних слів. Програма з графічним інтерфейсом, яка виявляє позитивно та негативно забарвлені слова, обчислює коефіцієнт Яніса та класифікує текст за цим коефіцієнтом. Програма дозволяє швидко здійснювати аналіз постів, додавати нові слова у словники, візуалізувати результати та зберігати їх.

Проте, у ході реалізації було виявлено низку проблем, наприклад:

- обмеженість словника;
- некоректне розпізнавання кількості постів;
- помилки в лематизації, відсутність контекстного аналізу.

Подальші перспективи вдосконалення програми вбачаємо у розширенні мультимовної підтримки, впровадження LLM, а також автоматичному оновленні словників.

Загалом дослідження підтвердило актуальність теми, доцільність використання автоматизованих та автоматичних методів виявлення та аналізу емоційного забарвлення, а також окреслило перспективи подальшого розвитку таких систем у галузях лінгвістики, психології та соціальних комунікацій. Створена програма є ефективним інструментом для автоматизованого аналізу емоційного забарвлення текстів. Розроблений проєкт є скромним внеском у галузь комп'ютерної лінгвістики,

а також демонструє потенціал міждисциплінарного підходу до вивчення емоцій у текстах сучасного цифрового простору.

ДЖЕРЕЛА

1. Вікіпедія - JSON. URL: <https://uk.wikipedia.org/wiki/JSON>
2. Вікіпедія - Реддіт. URL: <https://uk.wikipedia.org/wiki/Reddit>
3. Віталій Крикун. «ЕМОЦІЙНА МОВА» ЯК ЗАСІБ ЕФЕКТИВНОГО ВПЛИВУ В КОМУНІКАЦІЇ. *Людинознавчі студії: збірник наукових праць Дрогобицького державного педагогічного університету імені Івана Франка. Серія «Філософія».* 2021. URL: https://www.academia.edu/66101429/%D0%95%D0%9C%D0%9E%D0%A6%D0%86%D0%99%D0%9D%D0%90_%D0%9C%D0%9E%D0%92%D0%90_%D0%AF%D0%9A_%D0%97%D0%90%D0%A1%D0%86%D0%91_%D0%95%D0%A4%D0%95%D0%9A%D0%A2%D0%98%D0%92%D0%9D%D0%9E%D0%93%D0%9E_%D0%92%D0%9F%D0%9B%D0%98%D0%92%D0%A3_%D0%92_%D0%9A%D0%9E%D0%9C%D0%A3%D0%9D%D0%86%D0%9A%D0%90%D0%A6%D0%86%D0%87
4. “Горох”. Онлайн бібліотека словників української мови. URL: <https://goroh.pp.ua/%D0%A2%D0%BB%D1%83%D0%BC%D0%B0%D1%87%D0%B5%D0%BD%D0%BD%D1%8F/%D0%B5%D0%BC%D0%BE%D1%82%D0%B8%D0%BA%D0%BE%D0%BD>
5. Л.В. Чирун, В.А. Висоцька. ЗАСТОСУВАННЯ КОНТЕНТ-АНАЛІЗУ ТЕКСТОВОЇ ІНФОРМАЦІЇ В СИСТЕМАХ ЕЛЕКТРОННОЇ КОМЕРЦІЇ. 2010. *Наукові журнали та конференції.* С. 343 - URL: <https://science.lpnu.ua/sites/default/files/journal-paper/2019/apr/16291/vis689ism-332-347.pdf>
6. Машинне навчання для аналізу емоцій у текстах. 2025. URL: [Машинне навчання для аналізу емоцій у текстах - Блог Spashka](#)
7. Олена Смаліцька. Роль емоцій у міжособистісному спілкуванні. *НАУКА. ОСВІТА. МОЛОДЬ.* 2023. URL: [122.pdf](#)
8. Солдатова Л. П. СУТНІСТЬ ПОНЯТТЯ «МОВА» У ДОСЛІДЖЕННЯХ ЕДВАРДА СЕППРА. *Науковий вісник Міжнародного гуманітарного*

- університету. 2019. Вип. 40, ч. 2. С. 125–126 - URL: http://www.vestnik-philology.mgu.od.ua/archive/v40/part_2/31.pdf
9. Тлумачний словник української мови. URL: <https://slovnyk.ua/index.php?sword=%D0%B5%D0%BC%D0%BE%D1%86%D1%96%D1%8F>
10. Шульжук Наталія Василівна. ЕМОЦІЇ У КОНТЕКСТІ ПРОБЛЕМНОГО ПОЛЯ КОГНІТИВНО-ДИСКУРСНОЇ МОВОЗНАВЧОЇ ПАРАДИГМИ ТА АНТРОПОЗОРІЄНТОВАНОЇ ЛІНГВОДИДАКТИКИ. *Наукові записки Національного університету «Острозька академія»: серія «Філологія»*. Острог. 2023. С. 316 – 320. - URL: <https://journals.oa.edu.ua/Philology/article/view/3866/3514>
11. Що таке емпатія та як її розвинути?. 2022. URL: <https://vmpdu.edu.vn.ua/novyny/shcho-take-empatiya-ta-yak-yiyi-rozvynuty>
12. Aadil Gani Ganie. Presence of informal language, such as emoticons, hashtags, and slang, impact the performance of sentiment analysis models on social media text?. URL: <https://arxiv.org/pdf/2301.12303>
13. ABDULLAH AL MARUF, FAHIMA KHANAM¹, MD. MAHMUDUL HAQUE, ZAKARIA MASUD JIYAD, M. F. MRIDHA, (Senior Member, IEEE), AND ZEYAR AUNG. Challenges and Opportunities of Text-Based Emotion Detection: A Survey. *Magazine IEEE Access*. 2024.
14. Alexandru-Costin Băroiu and Ștefan Trăușan-Matu. Automatic Sarcasm Detection: Systematic Literature Review. *Magazine Information*. 2022.
15. Ashritha R Murthy, Anil Kumar K M. A Review of Different Approaches for Detecting Emotion from Text. *IOP Conference Series: Materials Science and Engineering*. 2021. URL: [pdf](#)
16. Convolutional Neural Network. *IT wiki/data science*. 2023. URL: <https://itwiki.dev/data-science/ml-reference/ml-glossary/convolutional-neural-network#:~:text=%D0%97%D0%B3%D0%BE%D1%80%D1%82%D0%BA%D0%BE%D0%B2%D0%B0%20%D0%BD%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%20%D0%BC%D0%B5%D1%80%D0%B5%D0%B>

[6%D0%B0%20\(CNN\)%20%2D,%D0%B4%D0%B5%20%D1%80%D0%BE%D0%B1%D0%B8%D1%82%D1%8C%D1%81%D1%8F%20%D0%BF%D1%80%D0%BE%D0%B3%D0%BD%D0%BE%D0%B7%20%D0%B0%D0%B1%D0%BE%20%D0%BA%D0%BB%D0%B0%D1%81%D0%B8%D1%84%D1%96%D0%BA%D0%B0%D1%86%D1%96%D1%8F](https://www.facebook.com/)

17. Facebook. URL: <https://www.facebook.com/>
18. Filipe N Ribeiro, Matheus Araujo, Pollyanna Goncalves, Marcos Andre Goncalves, Fabrício Benevenuto. SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science volume 5, Article number: 23*. 2016. URL: <https://arxiv.org/pdf/1512.01818>
19. Francisca Adoma Acheampong, Chen Wenyu, Henry Nunoo-Mensah. Text-based emotion detection: Advances, challenges, and opportunities. *Engineering Reports*. 2020. p. 9-10 - URL: <Text-based emotion detection: Advances, challenges, and opportunities>
20. George A. Miller. WordNet: A Lexical Database for English. *Communications of the ACM*. URL: <https://dl.acm.org/doi/pdf/10.1145/219717.219748>
21. Henrik Nordström. Emotional Communication in the Human Voice. Stockholm. 2019. p. 13-17 - URL: <https://www.diva-portal.org/smash/get/diva2:1304804/FULLTEXT01.pdf>
22. Instagram. URL: <https://www.instagram.com/>
23. Iva Novakova. EMOLEX Le lexique des émotions dans cinq langues européennes : sémantique, syntaxe et dimension discursive. *Agence Nationale de la Recherche*. France, 2023. URL : <https://hal.univ-grenoble-alpes.fr/hal-02526412/file/Compte-rendu%20final-EMOLEX.pdf>
24. James S. Boster. Handbook of Categorization in Cognitive Science. *Emotion Categories across Languages*. 2017. URL: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=a0f4bfa39deaf1c3bf4a408534d67e19dd5d6c0e>
25. Jérémie Clos, Anil Bandhakavi, Nirmalie Wiratunga, Guillaume Cabanac. Predicting Emotional Reaction in Social Networks. *39th European Colloquium on Information*

- Retrieval*. Aberdeen, United Kingdom, 2018. p. 3-4 - URL: <https://hal.science/hal-01809392/document>
26. Kristína Machová¹, Martina Szabóová¹, Ján Paralič and Ján Mičko. Detection of emotion by text analysis using machine learning. *Frontiers in Psychology “Emotion Science”*. 2023. p. 5-6 – URL: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1190326/full>
27. Leila Kerkeni, Youssef Serrestou, Mohamed Mbarki, Kosai Raoof, Mohamed Ali Mahjoub, Catherine Cléder. Automatic Speech Emotion Recognition Using Machine Learning. *Social Media and Machine Learning*. 2022. URL: <https://hal.science/hal-02432557/document>
28. Maliheh Alaeddini. Emotion Detection in Reddit: Comparative Study of Machine Learning and Deep Learning Techniques. *Computation and Language*. 2024. URL: <https://arxiv.org/pdf/2411.10328>
29. Mary Besemeres. Different Languages, Different Emotions? Perspectives from Autobiographical Literature. *Journal of Multilingual and Multicultural Development*. P. 140-145 - URL: https://www.researchgate.net/profile/Jean-Marc-Dewaele/publication/249024475_Languages_and_Emotions_A_Crosslinguistic_Perspective/links/567bfccb08ae19758383a2c9/Languages-and-Emotions-A-Crosslinguistic-Perspective.pdf#page=50
30. MSVPJ Sathvik, Abhilash Dowpati, Srreyansh Sethi. Ukrainian Resilience: A Dataset for Detection of Help-Seeking Signals Amidst the Chaos of War. *Findings of the Association for Computational Linguistics: EMNLP 2024*. 2024. URL: <https://aclanthology.org/2024.findings-emnlp.16.pdf>
31. Nandwani, P., Verma, R. A review on sentiment analysis and emotion detection from text. 2021
32. Ruiz, Jose Tronch, Silvia Sanz Blas. The Role of Emotions and Social Influence. 2016. URL: https://d1wqtxts1xzle7.cloudfront.net/117304097/BOOK_2023_003_NEW_97_end-libre.pdf?1723123234=&response-content-

[disposition=inline%3B+filename%3DThe_Role_Of_Emotions_And_Social_Influen
c.pdf&Expires=1744661452&Signature=XoiZQx1dxlrV1ye2kWd5iZY-lpvSjPC-
nOhIZoFo-gLJf3BVtnTARBb15k-
U1QThMDereENYo00waA2e3Qr6nCrbxwKGDzlahWQTetO1yjQUCGcvMeSyk7l
KlaLNeHn9KF~c8quAYIBFKCa0Q6~8ftEbdynwKv5Q8gdUoJQ3-
iHFAIR9PEZCSl89zPd3w8oA8~2AFqaE-
WjJNFxuyIufBVG5wTY3BfbpsGb5Ov6zRgaY0u6pQ3hx~eZlgZ3NInb7KZ1SLs1
Z-
1lwpAgTPngEdqUIFjCmLczC3dm3pSL9hjSk7jqvYeOTZcnp7T7eui~iELn9E~nFe
pW-Id95JpGKQw_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA](https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/)

33. Sentiment Analysis using VADER – Using Python. 2024. URL: <https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/>
34. Shashidhar G., Koolagudi K., Sreenivasa Rao. Emotion recognition from speech: a review. *Institute of Electrical and Electronics Engineers*. 2012. URL: [koolagudi12.pdf](https://www.researchgate.net/publication/312212121)
35. Siegfried Dewitte. The Role of Emotions in Advertising: A Call to Action Karolien Poels. 2019
36. Tess Dataset for Speech Recognition. 2025. URL: <https://www.restack.io/p/speech-recognition-answer-tess-dataset-cat-ai>
37. Threads. URL: <https://www.threads.com/?hl=en>
38. Werner Wirth and Holger Schramm. Media and Emotions. Centre for the Study of Communication and Culture. 2005.
39. Willie van Peer, Anna Chesnokova. What literature does to our emotions, and how do we know? Empirical studies will tell. *Synopsis: text, context, media*. 2019. URL: <https://synopsis.kubg.edu.ua/index.php/synopsis/article/view/345/298>
40. WordNet: A Lexical Database for English. *Princeton University*. URL: <https://wordnet.princeton.edu/>

ДОДАТКИ

Додаток 1. Корпус текстів

Посилання на текстовий корпус, що став основою для формування словників емоційно забарвленої лексики. До корпусу увійшли пости, коментарі та дописи, зібрані з відкритих джерел таких соціальних мереж, як Facebook, Instagram та Threads.

URL:

https://drive.google.com/drive/folders/1tFxu2bqvTdrJPhRJdgeCwLVbCkU_C8Fd?usp=sharing

Додаток 2. Словники лексики

Посилання на словники позитивної та негативної лексики, сформовані в результаті автоматизованої обробки текстових даних. URL:
https://drive.google.com/drive/folders/17TdVx4OLTR2fbzF7K5DsizNuMw_vpek5?usp=sharing

Додаток 3. Код програми

Код програми, розробленої для автоматизованого аналізу текстів з метою виявлення емоційно забарвленої лексики

```
import tkinter as tk

from ttkbootstrap import Style

from ttkbootstrap.widgets import Frame, LabelFrame, Button, Entry, OptionMenu

from ttkbootstrap.constants import *

from tkinter import filedialog, messagebox

from ttkbootstrap.scrolled import ScrolledText

import re

import json

import os

import matplotlib.pyplot as plt

# Шлях до словників

POSITIVE_WORDS_FILE = "positive_words.json"

NEGATIVE_WORDS_FILE = "negative_words.json"

def load_words(filename):

    if os.path.exists(filename):

        with open(filename, 'r', encoding='utf-8') as f:

            return set(json.load(f))

    return set()

positive_words = load_words(POSITIVE_WORDS_FILE)

negative_words = load_words(NEGATIVE_WORDS_FILE)
```

```
def save_words(words, filename):

    with open(filename, 'w', encoding='utf-8') as f:

        json.dump(sorted(list(words)), f, ensure_ascii=False)

def preprocess(text):

    text = re.sub(r'^\w\s', ' ', text)

    text = re.sub(r'\s+', ' ', text)

    text = text.lower()

    return [word for word in text.split() if word.isalpha()]

def analyze_text(words):

    pos = [w for w in words if w in positive_words]

    neg = [w for w in words if w in negative_words]

    r = len(pos) + len(neg)

    t = len(words)

    y = ((len(pos) - len(neg)) / r) * (r / t) if r > 0 and t > 0 else 0

    sentiment = "нейтральный"

    if y > 0.01:

        sentiment = "позитивный"

    elif y < -0.01:

        sentiment = "негативный"

    return pos, neg, round(y, 4), sentiment

def process_file():

    file_path = filedialog.askopenfilename(filetypes=[("Text Files", "*.txt")])
```

```
if not file_path:

    return

with open(file_path, 'r', encoding='utf-8') as f:

    texts = f.read().strip().split("\n\n")

output.delete('1.0', tk.END)

global_total_pos = 0

global_total_neg = 0

all_results = []

for idx, post in enumerate(texts):

    words = preprocess(post)

    pos, neg, score, sentiment = analyze_text(words)

    global_total_pos += len(pos)

    global_total_neg += len(neg)

    result = f"Текст {idx+1}:\n"

    result += f"Позитивні слова: {' '.join(pos) if pos else 'немає'}\n"

    result += f"Негативні слова: {' '.join(neg) if neg else 'немає'}\n"

    result += f"Коефіцієнт Яніса: {score} → {sentiment}\n\n"

    output.insert(tk.END, result)

    all_results.append(result)

summary = f"\nЗагалом позитивних: {global_total_pos}, негативних:
{global_total_neg}\n"
```

```
output.insert(tk.END, summary)

last_results.clear()

last_results.extend(all_results)

last_results.append(summary)

def save_results():

    file_path = filedialog.asksaveasfilename(defaultextension=".txt")

    if not file_path:

        return

    with open(file_path, 'w', encoding='utf-8') as f:

        f.write('\n'.join(last_results))

    messagebox.showinfo("Збережено", "Результати успішно збережено!")

def add_word():

    word = word_entry.get().strip().lower()

    if not word:

        return

    category = var.get()

    if category == "Позитивне":

        positive_words.add(word)

        save_words(positive_words, POSITIVE_WORDS_FILE)

    else:

        negative_words.add(word)

        save_words(negative_words, NEGATIVE_WORDS_FILE)

    word_entry.delete(0, tk.END)
```

```
messagebox.showinfo("Додано", f"Слово '{word}' додано до {category.lower()}  
слів.")  
  
def show_plot():  
  
    labels = ['Позитивні', 'Негативні']  
  
    values = [len(positive_words), len(negative_words)]  
  
    plt.bar(labels, values, color=['green', 'red'])  
  
    plt.title("Кількість слів у словнику")  
  
    plt.ylabel("Кількість")  
  
    plt.show()  
  
# --- СТИЛІ та ВІКНО ---  
  
style = Style(theme="flatly")  
  
root = style.master  
  
root.title("Емоційний аналіз постів")  
  
root.geometry("950x750")  
  
# Верхня панель кнопок  
  
button_frame = Frame(root, padding=10)  
  
button_frame.pack(fill=X)  
  
Button(button_frame, text="Завантажити файл", bootstyle=PRIMARY,  
command=process_file).pack(side=LEFT, padx=5)  
  
Button(button_frame, text="Зберегти результати", bootstyle=SUCCESS,  
command=save_results).pack(side=LEFT, padx=5)  
  
Button(button_frame, text="Графік словника", bootstyle=INFO,
```

```
command=show_plot).pack(side=LEFT, padx=5)

# Додавання слова

word_frame = LabelFrame(root, text="Додати слово до словника", padding=10)

word_frame.pack(fill=X, padx=20, pady=10)

word_entry = Entry(word_frame, width=30)

word_entry.pack(side=LEFT, padx=10)

var = tk.StringVar(value="Позитивне")

OptionMenu(word_frame, var, "Позитивне", "Позитивне", "Негативне").pack(side=LEFT,
padx=5)

Button(word_frame, text="Додати", bootstyle=WARNING,
command=add_word).pack(side=LEFT, padx=5)

# Виведення результатів

output_frame = LabelFrame(root, text="Результати аналізу", padding=10)

output_frame.pack(fill=BOTH, expand=True, padx=20, pady=10)

output = ScrolledText(output_frame, font=("Segoe UI", 11), height=30)

output.pack(fill=BOTH, expand=True)

last_results = []

root.mainloop()
```