

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Київський національний університет імені Тараса Шевченка
Інститут філології
Кафедра української мови та прикладної лінгвістики

**Автоматичне визначення тональності
українськомовного тексту (на матеріалі новин ТСН)**

Кваліфікаційна робота магістра
за спеціальністю 035 «Філологія»,
спеціалізацією 035.10 «Прикладна
лінгвістика»,
галузі знань 03 «гуманітарні науки»
ОП "Прикладна лінгвістика
(редакторсько-перекладацька
та експертна діяльність)"

БЛОУС Валерії Вячеславівни

науковий керівник:

к.філол.н.,доц. Оксана ЗУБАНЬ

рецензент: д.філол.н. Наталія ДАРЧУК

ЗМІСТ

ВСТУП.....	4
РОЗДІЛ 1. ЕМОТИВНИЙ АНАЛІЗ ТЕКСТУ: МЕТОДОЛОГІЧНИЙ АСПЕКТ.....	7
1.1.Емотивний складник тексту.....	7
1.2. Сентимент-аналіз: історія розвитку та сучасний стан.....	11
1.3. Методи тонального аналізу тексту.....	15
1.4. Тональні словники української мови.....	19
1.5.Синтаксичні зв'язки в реченні для визначення емотивності тексту.....	22
1.6. Висновки до розділу 1.....	25
РОЗДІЛ 2. СИСТЕМА АВТОМАТИЧНОГО ВИЗНАЧЕННЯ ТОНАЛЬНОСТІ УКРАЇНСЬКОМОВНИХ ТЕКСТІВ.....	27
2.1. Укладання тонального резидентного словника української мови.....	28
2.2 Укладання корпусу текстів.....	29
2.3 Розроблення програмного забезпечення системи автоматичного визначення тональності українськомовних текстів новин ТСН.....	31
2.4 Результати роботи системи автоматичного визначення тональності українськомовних текстів.....	41
2.5. Зіставлення автоматичної тональної оцінки з тональною оцінкою людини.....	46
2.6. Тональна класифікація українськомовних текстів.....	56
2.7. Висновки до розділу 2.....	57
РОЗДІЛ 3. ТОНАЛЬНИЙ СЛОВНИК СУЧАСНОЇ УКРАЇНСЬКОЇ МОВИ: ЛЕКСИЧНА СИСТЕМА ТА ТЕКСТОВА РЕАЛІЗАЦІЯ	59
3.1. Лінгвостатистична характеристика реєстрів резидентного тонального словника та тонального словника текстів новин.....	59
3.2. Тональна статистична структура тексту.....	61
3.3. Висновки до розділу 3.....	66
ВИСНОВКИ.....	68

ВИКОРИСТАНІ ДЖЕРЕЛА.....	70
ДОДАТОК 1.....	75
ДОДАТОК 2.....	85

ВСТУП

Актуальність дослідження : У зв'язку з глобалізацією сучасного світу, виникає необхідність розроблення ефективних алгоритмів аналізу і класифікації інформації.

Упродовж останніх десяти років у галузі опрацювання природної мови широкої популярності набув емоційно-смісловий аналіз. Про це свідчить численна кількість досліджень у цій сфері. Фундаментальний науковий вклад у дослідження загальних теоретичних аспектів емотивного та емоційного спілкування зробили такі вчені-лінгвісти, як: Б. Волек [8], В. Шаховський [39], О. Виноградов [7], Л. Бабенко [3] та ін.

Емоційно-смісловий аналіз набуває популярності у зв'язку з розвитком різних платформ для оцінювання (сайти про фільми, одяг або техніку).

На сьогоднішній день, існує досить велика кількість систем для визначення тональності тексту англійської, російської мов, наприклад:

- SentiWordNet ;
- WordNet-Affect;
- SenticNet;
- ВААЛ;

Для української мови створювалися окремі програми, у яких аналізувалися тексти однієї тематики. З'явилися такі цікаві проекти, як дослідження емоційного забарвлення відгуків про готелі [46], ресторани; коментарів про фільми [55], продукти; повідомлень у блогах і соціальних мережах про політичні події тощо.

Для української мови не створено жодної системи сентимент аналізу, тому постає завдання розроблення системи автоматичного визначення емотивної тональності українськомовних текстів.

Метою дослідження є створення автоматичної системи визначення лексичної тональності українськомовних текстів.

Завдання:

- проаналізувати сучасні лінгвістичні концепції тональності тексту;
- дослідити основні методи емотивного аналізу тексту;
- розробити алгоритм автоматичного визначення тональності українськомовних текстів на лексичному рівні організації;
- укласти тональний словник української лексики;
- укласти базу даних публіцистичних текстів української мови;
- розробити програмне забезпечення автоматичного аналізу тональності текстів;
- проаналізувати ефективність роботи системи;
- Провести статистичне дослідження емотивної лексики публіцистичних текстів.

Об'єктом дослідження є емотивна іменникова, прикметникова, прислівникова та дієслівна лексика української мови.

Предмет дослідження: лексико-емотивна статистична структура публіцистичних текстів української мови.

Теоретичне та практичне значення роботи:

Теоретична значущість виконаної роботи полягає в тому, що в ній теоретично обґрунтовано та практично реалізовано алгоритм створення системи автоматичного визначення тональності українськомовного тексту.

Практичне значення одержаних результатів полягає у створенні тонального словника української мови та розробленні автоматичної системи емотивного аналізу українськомовних текстів. Ця система може бути використана для класифікації тексту (наприклад, в судово-криміналістичному аналізі), так і в інших задачах штучного інтелекту.

Методи дослідження:

Основними методами дослідження є метод компонентного аналізу, метод комп'ютерного моделювання, метод статистичного аналізу та кількісних

підрахунків, метод алгоритмів.

У роботі було використано методи об'єктно-орієнтованого програмування, система керування базами даних SQLServer.

Матеріал дослідження: тексти новин ТСН, які разом складають дві зонально хронологічні вибірки текстів обсягом 232 010 слововживань (тексти 2019 та 2020 років) та 181 275 слововживань (тексти 2021 року). Матеріал добирався з сайту новин ТСН [29] за допомогою бібліотеки scrapy [5].

РОЗДІЛ І

ЕМОТИВНИЙ АНАЛІЗ ТЕКСТУ: МЕТОДОЛОГІЧНИЙ АСПЕКТ

Упродовж останніх десяти років у галузі опрацювання природної мови широкої популярності набув емоційно-смісловий аналіз, про що свідчить численна кількість досліджень у цій сфері.

Емоційно-смісловий аналіз може використовуватися для класифікації тексту (наприклад, в судово-криміналістичному аналізі), так і в інших задачах штучного інтелекту. Розпізнавання емоційного забарвлення також називають аналізом тональності тексту. Аналіз тональності тексту (сентимент-аналіз) - галузь комп'ютерної лінгвістики, що займається виділенням із текстів емоційно забарвленої лексики або емоційної оцінки автора.

1.1. Емотивна складова тексту

У лінгвістиці тексту розглядають як мову опису емоцій і мову вираження емоцій. В. Телія стверджує, що «на мовному рівні емоції трансформуються в емотивність, тобто емоції є психологічна категорія, а емотивність – мовна» [34].

В. Шаховський [41] стверджував, що «емоція постає способом оцінки предметів і явищ реальної дійсності, виникнення оцінного ставлення є умовою виникнення емоції. Емотивність тексту є однією з основних властивостей тексту, яка актуалізується через емотивно навантажені текстові компоненти, що втілюють авторські емоційні наміри й моделюють ймовірні емоції адресата».

У лінгвістиці не існує повного й однозначного визначення поняття емотивності.

В. Шаховський визначає «емотивність як семантичну ознаку, притаманну мові, виражати системою своїх засобів емоційність за допомогою мовних одиниць у семантиці» [39, с. 23]. Дослідник також зазначає, що «емотивність – це емоційність у мовному значенні, тобто чуттєва оцінка об'єкта, вираження мовними або мовленнєвими засобами відчуттів, настроїв, переживань людини» [41, с.153]. Емотивність завжди експресивна та оцінювальна, але не навпаки.

О. Селіванова [31, с.248] під емотивністю розуміє «складову конотативного компонента у семантичній структурі мовної одиниці, який репрезентує емоційне ставлення носіїв мови до позначуваного. Емотивність

може формувати й денотат значення слова, що створює суперечність у розгляді денотата й конотата у встановленні межі між ними».

І. Літвінчук називає емотивністю «результат інтелектуальної інтерпретації емоційності, що транслюється в мові та мовленні» [17, с. 1].

М. Гамзюк розуміє під емотивністю «мовне вираження емоцій» [9, с. 37].

Поняття Emotionalität у словнику Kleines Wörterbuch der Stilkunde розуміють як, емоційність, тобто «можливе позначення суми емоційних, тобто пов'язаних із душевним переживанням, елементів тексту» [54, с. 39].

Є. Мягкова використовує поняття «емоційне навантаження слова», що визначає як «ті якості слів, завдяки яким вони виражають і називають емоції» [22, с. 10].

Сучасні дослідники визначають емотивність як текстову категорію, яка виражає емоційне ставлення автора, його функції у тексті, дійових осіб, ймовірність емоцій реального чи, модельованого авторською свідомістю, гіпотетичного читача щодо описуваних подій, явищ, персонажів, їхньої поведінки й аналізується за допомогою емоційно маркованих текстових компонентів.

Об'єктом категорії емотивності є емоції. Дослідники зазначають, що саме детальний аналіз цього об'єкта дає змогу глибше проникнути в сутність емотивності [9, с. 20].

Під емотивним текстом розуміємо висловлювання, сформульоване одним чи декількома реченнями, яке «передає поряд з факультативною і емоційну інформацію (чи тільки лише її) за допомогою як мінімум одного емотивного засобу – лінгвістичного чи паралінгвістичного, який виражає визначену емоцію, більш чи менш адекватно усвідомлювану усіма комунікантами у певній ситуації» [41, с. 68].

Емотивність тексту є «однією з основних властивостей тексту, яка

співвідноситься із опредметненими в ньому емоціогенними знаннями та актуалізується через емотивно навантажені текстові компоненти, які втілюють авторські емоційні наміри й моделюють ймовірні емоції адресата, пов'язані зі сприйняттям та інтерпретацією текстової дійсності» [42].

Відповідно до емоцій та емотивності, розрізняють емоційну та емотивну комунікацію.

Л. Чибирова [38] розуміє під емотивною комунікацією «свідому демонстрацію емоцій, яка використовується для інформаційного впливу, виконує соціальну функцію – вплинути на сприйняття ситуації співрозмовником та на її розуміння, тоді як емоційна комунікація – це спонтанна незапланована природна демонстрація емоцій людиною».

Яка ж лексика несе емоційно-оцінне навантаження у тексті? А. Стадній дає таке визначення цьому поняттю: «Емоційно-оцінна (конотативна) лексика – це лінгвістичний феномен, особливий компонент, за допомогою якого виражається психічний стан мовця, його ставлення до предмета, об'єкта, адресата мовлення, тобто все те, що складає емотивну функцію слова» [33].

Питання про емотивні компоненти й емотивну структуру тексту сьогодні, звичайно, не є вирішеним, однак певні результати проведених досліджень вже можуть бути використані в аналізі тексту при встановленні його семантичної структури.

Джерела емотивності тексту виражаються різноманітними засобами і не усіма дослідниками розуміються однаково. З одного боку, основним джерелом емотивності тексту є емотивні мовні засоби. Способи відображення емотивних ситуацій у тексті різноманітні: «від згорнутих (семний конкретизатор, слово) і мінімально розгорнутих (словосполучення, речення) до максимально розгорнутих (фрагмент тексту, текст)» [16, с.16].

Тональністю тексту, або сентиментом тексту (від англ. sentiment — почуття, настрої), називається емоційна оцінка, виражена в тексті різними імпліцитними та експліцитними засобами:

- ✓ лексикою;
- ✓ пунктуацією;
- ✓ інверсією в реченні (зміною порядку слів в реченні);
- ✓ звертаннями;
- ✓ графічними невербальними засобами.

Дотримуючись комунікативного підходу у вивченні емотивності, В. Маслова вважає, що найважливішим джерелом емотивності тексту є його зміст. На думку дослідниці, «зміст тексту є потенційно емоціогенним, тому що завжди знайдеться реципієнт, для якого він виявиться особистісно значущим. Емоціогенний зміст тексту – це, у підсумку, емоціогенні фрагменти світу, відображені у тексті» [20, с.21].

Тональність тексту в цілому визначається лексичною тональністю (емоційний складник, виражений на рівні лексеми або комунікативного фрагмента). Емоційний складник тексту, виражений на рівні слововживань та словосполучень (правил сполучуваності тональної лексики), називається лексичною тональністю або лексичним сентиментом. Автоматичне визначення тональності тексту спрямоване на виділення тих фрагментів тексту, які висловлюють позитивну або негативну емоційність відносно об'єкта емоційної оцінки (об'єкта тональності). Таким об'єктом може бути власна назва, назва організації, послуги, професії і т.ін., стосовно якого аналізується текст.

Таким чином, тональність тексту визначається трьома факторами:

- 1) суб'єктом тональності (той, хто виражає свою думку з приводу того чи іншого об'єкта комунікації, автор);
- 2) власне тональною оцінкою (позитивна / нейтральна / негативна) - емоційним ставленням автора до об'єкта;
- 3) об'єктом тональності(про кого або що говорить автор).

Л. Бабенко зазначає, що «Бінарність емотивного змісту обумовлена не тільки текстовими властивостями носіїв (автор-персонаж), а й функціонально. Цілісний емотивний зміст передбачає обов'язкову інтерпретацію світу людських емоцій (рівень персонажа) та оцінку цього світу з позиції автора з метою впливу на цей

світ, його перетворення» [3,с.123].

1.2. Сентимент-аналіз: історія розвитку та сучасний стан

Мова слугує не лише для обміну думками, але й для вираження емоцій автора та впливу на емоційну сферу читача. Проблеми виявлення експресивного потенціалу мовних одиниць стали предметом особливої уваги відомих мовознавців: В. Виноградова [7], Г. Колшанського [15], А. Леонтьєва [18], А. Уфимцевої [37] та ін.

У “Словнику лінгвістичних термінів” О. Ахманової [2] категорію оцінки визначено як «судження мовця, його відношення – позитивне або негативне, бажання, заохочення; одна з основних частин стилістичної конотації».

У Енциклопедії «Українська мова» [36] категорію оцінки визначають як «ставлення носіїв мови до об’єкта, зумовлене визнанням чи невизнанням його цінності з погляду відповідності чи невідповідності його якостей певним ціннісним критеріям».

За В. Шаховським у сучасній лінгвістиці визначають три основні напрями досліджень: «комунікацію емоцій, категоризація емоцій та емотивний семантичний простір мови» [40, с. 40].

Лінгвістика емоцій як наука сформувалась у ХХ ст. на основі психології та традиційного мовознавства. Проте до середини 70-х років проблема вираження емоцій не була актуальною в лінгвістиці.

Із розвитком антропоцентричної лінгвістичної парадигми проблема емотивності мови стає однією з провідних. Висуваються різні концепції та постають питання «мовної вербалізації та концептуалізації, категоризації емоцій, виникнення емоції, чи виходить вона зі слів і зворотів, чи йде від особистості, яка промовляє фрази, чи вона є поняттям мовним, чи когнітивним» [40, с. 21].

Сьогодні, після десятків років досліджень у галузі лінгвістики, вивчення емоцій, на думку В. Шаховського, вийшло на новий рівень [40, с. 42]. Дослідник зауважує, що на відміну від французьких лінгвістів, які дотримувалися думки,

що «лише деякі слова мають емоційний компонент, а більшість словникового запасу мови має нейтральне забарвлення, російські психолінгвісти вважають такий поділ неправильним і переконані, що кожне слово може мати емоційний компонент» [42, с. 81–92].

У сучасній комп'ютерній лінгвістиці створено велику кількість систем для автоматичного емотивного аналізу тексту, зокрема для англійської та російської мов. Для англійської мови було створено семантичні тезауруси, які враховують емоційну складову понять. Одним з найвідоміших є WordNet-Affect. Це семантичний тезаурус, у якому поняття пов'язані з емоціями («емоційними концептами»), представлені вербально. WordNet-Affect складається з такої підмножини синсетів WordNet, де кожен синсет, відповідно до «емоційного концепту», може бути представлений за допомогою «емоційних слів» [44, с. 1083–1086.] Таким чином, WordNet-Affect був створений на основі WordNet для англійської мови, шляхом вибору і віднесення наборів синонімів (синсетів) до різних емоційних понять. Зокрема, синсети дієслів, іменників, прикметників, прислівників, які описують емоції, були вручну розмічені за допомогою спеціальних емоційних міток (affective labels, A-labels). Ці емоційні мітки характеризують різні стани, що виражають настрої, емоційне ставлення або ситуації, які викликають емоції. Приклади таких емоційних міток наведені в Табл. 1.1.

ЕМОЦІЙНА МІТКА	ПРИКЛАД
Емоція(emotion)	ім. гнів#1, дієсл. боятися#1 (fear)
Настрій (mood)	ім. ворожість#1 (animosity), прикм. люб'язність#1J (amiable)
Особливість (trait)	ім. агресивність#1 (aggressiveness), прикм. який_змагається#1 (competitive)
Когнітивний стан (cognitivestate)	ім. збентеження, зніяковіння#2 (confusion), прикм. вражений#2 (dazed)

Фізичний стан (physicalstate)	ім.хвороба#1 (illness), прикм. знесилений#1 (allin)
Гедонічний сигнал (gedonicsignal)	ім. біль#3(hurt), ім. страждання#4 (suffering)
Ситуації, що викликають емоції (emotion- eliciting situation)	ім. незручність#3 (awkwardness), ім. безпека#1 (outofdanger)
Емоційні відгуки (emotionalresponses)	ім. холодний піт#1 (coldsweat), дієсл. тремтіти#2 (tremble)
Вчинки (behaviour)	ім. злочин#1 (offense), прикм. загальмований#1 (inhibited)
Ставлення, позиція (attitude)	ім. нетерплячість#1 (intolerance), ім. оборонна позиція#1 (defensive)
Відчуття (sensation)	ім. холод#1 (coldness), дієсл. відчувати#3 (feel)

Таблиця 1.1 Приклади емоційних міток WordNet-Affect

Також у WordNet-Affect використовуються додаткові емоційні мітки для того, щоб розділяти синсети відповідно до їх емоційної валентності. Для цього визначаються чотири додаткові емоційні мітки: позитивна, негативна, неоднозначна і нейтральна. Перша відповідає позитивним емоціям, які визначаються як емоційні стани, що характеризуються наявністю позитивних гедоністичних сигналів (або задоволення). Вона містить такі синсети, як радість#1 або захоплення#1. Аналогічно негативна мітка ідентифікує негативні емоції, які характеризуються негативними гедоністичними сигналами (або болем), наприклад, гнів#1 або печаль#1. Синсети, що представляють емоційні стани, валентність яких залежить від семантичного контексту (наприклад, здивування#1), позначаються як неоднозначні. Синсети, що визначають психологічні стани, і завжди розглядаються як неоднозначні, але при цьому не актуалізуються через валентність, є нейтральними.

Синсети, помічені емоційними мітками, додатково перерозмічуються шістьма емоційними категоріями: радість, страх, гнів, сум, відраза, подив. Таким чином, структура даних WordNet-Affect складається з шести файлів: anger.txt, disgust.txt, fear.txt, joy.txt, sadness.txt, surprise.txt, кожен із яких описує одну категорію [14].

Іншою відомою системою, розробленою на основі WordNet є SentiWordNet. Це лексичний семантичний тезаурус, перша версія якого була розроблена в 2006 році. Ця система є результатом процесу автоматичного анотування кожного WordNet синсета за його ступенем позитивності, негативності і об'єктивності [32]. Таким чином, кожному синонімічному ряду з WordNet присвоюється три чисельні оцінки, кожна з яких визначає об'єктивний, позитивний або негативний складник синсета [52, с. 2200–2204.]. Кожна з оцінок набуває значення в інтервалі від 0 до 1, ці показники додаються та обраховується їх середнє арифметичне значення, можливе максимальне значення якого 1. [52, с. 2200–2204].

Ще однією системою емотивного аналізу тексту є SenticNet. Це семантичний тезаурус для роботи з наборами емоційних понять. SenticNet є проектом, медіалабораторії Масачусетського технологічного інституту в 2010 році. Головним призначенням SenticNet є спрощення процедури машинного розпізнавання концептуальної та емоційної інформації, переданої природною мовою. Якщо порівняти лексичні тезауруси, SentiWordNet і WordNet-Affect з SenticNet, то їх основною відмінністю буде те, що SentiWordNet і WordNet-Affect забезпечують визначення сполучуваності емоційно-маркованих слів на синтаксичному рівні, тоді як SenticNet пов'язує поняття на семантичному рівні. SenticNet дозволяє виявляти смисловий складник тексту, наприклад, «досягнення мети», «погане почуття», «відсвяткувати особливу подію» або «бути на сьомому небі від щастя». SenticNet забезпечує визначення сполучуваності слів із більш ніж 14000 поняттями та дозволяє проводити глибший і багатогранний аналіз тексту природною мовою. SenticNet побудований за допомогою «sentic-обчислень», парадигми, яка використовує

методи штучного інтелекту і семантичної мережі для поліпшення розпізнавання, інтерпретації та оброблення висловлювань природною мовою.

Для російської мови було розроблено систему ВААЛ [6]. Програмна система ВААЛ розроблена з метою автоматизованого прогнозування ефекту неусвідомлюваного впливу текстів на масову аудиторію. Для кожного виду психоемоційного впливу в системі передбачена своя категорія. Ступінь належності тексту до певної категорії подано за шкалою оцінки певного виду впливу. Кожна шкала задається двома протилежними за змістом поняттями (наприклад, хороший – поганий, легкий – важкий і т.п.). Такий функціонал не повністю відповідає вирішенню задачі визначення емоційного забарвлення природномовних текстових даних, але є достатньо близьким до неї. Алгоритмічну основу в системі ВААЛ представляє алгоритм фоносемантичної оцінки текстових даних. В основі оцінювання емоційного впливу на підсвідомість людини тексту російською мовою лежить алгоритм Н.М. Журавльова. Цей алгоритм базується на аналізі емоційного впливу звуків в окремих словах. Тобто основною одиницею, що формує вплив тексту на підсвідомість людини є звук. Для розв'язання задачі автоматизованого визначення емоційного забарвлення такий підхід не є коректним. А. Бегунов стверджує, що «емоційне забарвлення тексту формується за допомогою емоційно забарвлених слів, а також зв'язків між ними та нейтрально забарвленими словами» [4].

1.3. Методи тонального аналізу тексту

Існує п'ять основних методів визначення тональності тексту.

1. Аналіз тексту методами векторного аналізу (часто із застосуванням n-грамних моделей), порівняння з раніше розміченим еталонним корпусом за обраною мірою близькості і віднесення тексту до негативу або позитиву на підставі отриманого результату порівняння. Детально цей метод описано у роботі Б. Панг «Використання машинного навчання для сентимент-аналізу тексту» [43]. Цей метод дуже ефективний, але вимагає наявності попередньо розміченого еталонного корпусу, на основі якого відбувається навчання алгоритму

порівняння. Істотними недоліками такого підходу є велика трудомісткість і обмеження різномірності корпусу (тобто неповнота лексичного покриття), що призводить до втрати точності. До того ж цей метод не дозволяє провести глибокий аналіз тексту, тобто виявити і показати емотивність на рівні пропозиції.

2. Пошук емотивної лексики (лексичної тональності) в тексті

за задалегідь складеним тональним словником (списками патернів) із застосуванням лінгвістичного аналізу. За сукупністю знайденої емотивної лексики текст може бути оцінений за шкалою, в основі якої лежить кількість негативної та позитивної лексики. Цей метод може використовувати як списки патернів, що підставляються в регулярні вирази, так і правила з'єднання тональної лексики у пропозиції [47, с. 70-77]. Цей метод не менш трудомісткий у частині складання тональних словників (або отримання списку тональних ланцюгів), але в поєднанні зі синтаксичним і морфологічним аналізом більш гнучкий: він дозволяє не тільки показати ланцюг тональної лексики, а й отримати синтаксично коректні емоційні вирази. За умови якісного наповнення тональних словникових списків цей метод дозволяє досягти повноти аналізу (покриття емотивної лексики). Недолік цього методу в тому, що за допомогою нього складно дати кількісну оцінку негативності-позитивності тексту.

3. Змішаний метод (комбінація першого і другого підходів) [23].

З метою уникнення недоліків першого і другого методів, використовують змішаний підхід, який частково містить два перших. Детально цей метод описав у своїй роботі Р. Прабово «Сентимент аналіз: змішаний підхід» [48].

4. Метод на основі правил із використанням шаблонів

Цей метод заснований на пошуку емотивної лексики (лексичної тональності) у тексті за укладеним тональним словником і правилами із застосуванням лінгвістичного аналізу. За сукупністю знайденої емотивної лексики текст може бути оцінений за шкалою, що містить кількість негативної та позитивної лексики. Цей метод може використовувати як правила, з використанням

регулярних виразів, так і спеціальні правила поєднання тональної лексики всередині висловлення. Щоб проаналізувати текст, можна скористатися таким алгоритмом: спочатку кожному слову в тексті приписати його значення тональності зі словника (якщо воно присутнє в словнику), а потім обчислити загальну тональність всього тексту шляхом підсумовування значення тональностей кожного окремого висловлення [24]. І. Четв'оркін стверджує, що «основною проблемою методів, заснованих на словниках і правилах, вважається трудомісткість процесу складання словника» [45]. Для того, щоб отримати метод, що класифікує документ з високою точністю, терміни словника повинні мати значення відповідно до теми досліджуваного тексту. Наприклад, слово *величезний* по відношенню до обсягу пам'яті жорсткого диска є позитивною характеристикою, але негативною по відношенню до розміру мобільного телефону. Тому даний метод вимагає значних трудовитрат, тому що для хорошої роботи системи необхідно скласти велику кількість правил.

5. Метод графів

В основі цього методу використовується припущення про те, що не всі слова в текстовому корпусі документа рівнозначні. Якісь слова мають більшу вагу і сильніше впливають на тональність тексту. При використанні цього методу аналіз тональності розбивається на кілька етапів:

- побудова графа на основі досліджуваного тексту;
- ранжування його вершин;
- класифікація знайдених слів;
- обчислення результату.

Для класифікації слів використовується тональний словник, у якому кожне слово співвідноситься з оцінкою, наприклад «позитивна», «негативна» або «нейтральна». Для отримання кінцевого результату потрібно обчислити значення двох оцінок: позитивного складника тексту і негативної. Для того, щоб знайти позитивний складник тексту необхідно знайти суму тональностей всіх позитивних слів тексту. Значення негативного складника тексту знаходиться

аналогічним чином. Для підсумкової оцінки тональності всього тексту потрібно обчислити відношення цих складників за формулою:

$$T=P/N$$

де T - підсумкова оцінка тональності, P - оцінка позитивного складника тексту і N – оцінка негативного складника тексту. Відповідно до статті І. Меньшикова [21] текст, в якому значення T близьке до одиниці, буде вважатися нейтральним, якщо трохи перевищує 1 - позитивним. За таким самим принципом визначається і негативна складова тексту. Для обрахунку оцінки тональності тексту за негативним складником відповідно використовується формула:

$$T= N/P$$

Тоді текст, в якому значення T близьке до одиниці, також буде вважатися нейтральним, якщо ж перевищує 1 - негативним.

Найоптимальнішим методом є пошук емотивної лексики за заздалегідь складеним тональним словником із застосуванням лінгвістичного аналізу. Тональний словник - це список слів і словосполучень зі значенням тональності для кожного слова. А. Романюк говорить, що «це може бути як числове значення в межах певної визначеної шкали (наприклад, від 1 до 10, де 1 – негативне слово, а 10 – позитивне), так і категорія (наприклад, позитивне чи негативне слово). Кожному слову приписується його частиномовна характеристика та початкова форма у випадку, коли словник містить різні словоформи одного слова». [28, с. 64]

Часто списки слів у словнику обмежені лише самостійними частинами мови. Наприклад, А. Пазельська та А. Соловйов у роботі над повідомленнями російською мовою [23] використовували тональний словник із найчастотніших іменників, прикметників, прислівників та дієслів, зібраних зі статей ЗМІ. Для кожного слововживання було визначено частину мови, тональність і силу тональності за шкалою від 1 до 3. Прикметники і прислівники поділені на позитивні, негативні та ті, що підсилюють тональність. Іменники поділені на

позитивні, негативні, потенційно позитивні та потенційно негативні (залежно від оточення: позитивні в позитивному оточенні, а негативні – в негативному). Дієслова поділені на 8 класів залежно від оточення і синтаксичної ролі в реченні, окремо подані дієслова-зв'язки. Є також слова-інвертори, що змінюють тональність сусідніх слів на протилежну. Таким чином, бачимо, що слова поділені за частинами мови, і відповідно кожна група слів має свою класифікацію.

Набір слів у тональному словнику переважно має стосуватися однієї тематики, яку опрацьовуватиме аналізатор. Оскільки складання такого списку слів власноруч потребує дуже багато часу, застосовують автоматичні способи вилучення слів певної тематики із сентимент-анотованих корпусів та онтологій, що стосуються певної тематики. Якщо ж тематика словника не передбачена, використовують загальнотематичні семантичні мережі. Наприклад, у деяких системах послуговуються тональним словником, укладеним на основі WordNet 2.1. Прикладом багатомовного тонального словника є SentiWordNet, який часто використовується для здійснення різноманітних завдань у сфері емоційно-сміслового аналізу. Доволі часто користуються також GeneralInquirerLexicon.

Проаналізувавши різноманітні методи автоматичного визначення емотивності тексту, можна стверджувати, що у кожного методу є свої переваги і недоліки. Крім того, особливу складність вивчення феномену текстової емотивності в лінгвістиці обумовлює існування різноманітних підходів до нього. У рамках кожного із підходів розробляються власні концепції текстової емотивності, використовується своя термінологія і методи дослідження, що ускладнює розуміння досліджуваного явища [30, с. 275].

1.4 Тональні словники української мови

В українській комп'ютерній лінгвістиці ще не створено автоматичної системи емотивного аналізу українськомовного тексту, але існують різноманітні інформаційні продукти автоматичного оброблення українськомовного тексту, які можуть бути використані в автоматичному емотивному аналізі.

Однією з перших українських систем автоматичного оброблення природної мови є система автоматичного граматичного аналізу тексту – АГАТ [1]. Система АГАТ призначена для автоматичного опрацювання українськомовного тексту і є інструментом аналізу Корпусу української мови, що нараховує більше ніж 80 млн слововживань.

До словникового комплексу АГАТ належить низка лексичних словників, які описують лексичні відношення між словами (словник синонімів, антонімів, тезаурус), морфемний словник, в якому словам приписано інформацію про морфну будову слова, аломорфію та омонімію коренів. Усі вони об'єднані у базу даних (БД) і становлять гіперсловник.

З лінгвістичного погляду до кожної лексеми подається різноаспектна (зональна) інформація: зона морфологічна (частина мови, для змінюваних – номер парадигматичного класу, певні граматичні характеристики для дієслів та лексико-семантичні розряди для іменників); зона синтаксична (містить припис про синтаксичну валентність і сполучуваність іменників, дієслів, ад'єктивів в українській мові); зона семантична (кожній лексемі приписується семантичний код за таксономічною класифікацією).

Семантичний модуль системи АГАТ створювався за таксономічною методологією національного корпусу російської мови.

Лінгвістична таксономія – сукупність принципів і правил класифікації об'єктів, а також сама класифікація. Таксономія передбачає систематизацію як онтологічний результат, що частково відображає ієрархічну організацію, а для кожної частини мови розроблено свою таксономію зі своїм набором таксонів. У своїй роботі «Задачи и принципы семантической разметки лексики» Є. Рахіліна [27, с. 226] визначає основні вимоги до семантичних класів у корпусній таксономічній розмітці російської мови. До них належать : незалежність таксонів; базовість ознак; максимальне укрупнення класів; породження мінімального шуму на запит користувача; оптимальність результату пошуку.

Таксономічна класифікація лексики всіх частин мови складається з трьох частин: синоптичної, аналогічної та алфавітної.

Синоптична частина поняттєва класифікація змістовного аспекту лексики української мови.

В аналогічній частині таксономії кожна з поняттєвих груп наповнюється словами, причому поняттєві групи розташовані в алфавітному порядку заголовкових слів-понять.

Алфавітна частина є не що інше, як звичайний тлумачний словник сучасної української мови, в якому слова в кожному із своїх значень не тільки супроводжується вербальним визначенням, але й вказівкою на їх місце в таксономічній класифікації.

На рис. 1.1 наведено приклад таксономічної класифікації іменника, де зліва синоптична частина, справа – перша колонка – аналогічна частина, тобто лексеми в алфавітному порядку, третя колонка – тлумачення слова із тлумачного словника.

The screenshot shows the MOVA.info website interface. At the top, there is a navigation bar with the logo 'MOVA.info ЛІНГВІСТИЧНИЙ ПОРТАЛ' and search options. Below the navigation bar, there is a breadcrumb trail: 'Головна / Словники / Граматичний словник української мови (словозміна)'. The main content area is divided into two parts. On the left, there is a tree view of grammatical categories. The 'Іменник' (Noun) category is expanded, showing sub-categories like 'ВЛАСНІ ІМЕНА' and 'НЕПРЕДМЕТНІ ІМЕНА'. Under 'НЕПРЕДМЕТНІ ІМЕНА', the 'влада' (power) category is selected. On the right, there is a table of words with their meanings and a 'Вилучити' (Remove) button for each. The table contains the following entries:

1	<i>абсолютизм</i>	необмежена верховна влада в руках однієї людини	Вилучити
1	<i>автократія</i>	необмежена верховна влада в руках однієї людини	Вилучити
51	<i>автократія</i>	Форма правління, за якої одній особі належить необмежена верховна влада; абсолютизм.	Вилучити
51	<i>ареопаг</i>	У Стародавніх Афінах - найвищий орган судової та політичної влади, що засідав на пагорбі Ареса - бога війни, звідки й дістав назву	Вилучити
1	<i>власі</i>		Вилучити
1	<i>всевладдя</i>	необмежена влада	Вилучити

Рис.1.1 Приклад таксономічної класифікації іменника у системі АГАТ

Іншою системою автоматичного оброблення українськомовного тексту є Великий електронний словник української мови (ВЕСУМ). Система ВЕСУМ включає:

- Мікросервіси: токенизацію, лематизацію, NER-розмітку тексту;
- Корпус NER-анотацій. Містить 229 текстів з українського браунівського корпусу на 217 381 токенів з 6 751 розмічених іменованих сутностей);
- Корпус UberText – корпус текстів українських періодичних видань;
- тональний словник.

Тональний словник цієї системи містить 3442 слова з тональною характеристикою, охарактеризованою за шкалою від -2 до 2.

У словнику по можливості всі слова приведені до базової граматичної форми, а також прислівники замінені на спільнокореневі прикметники.

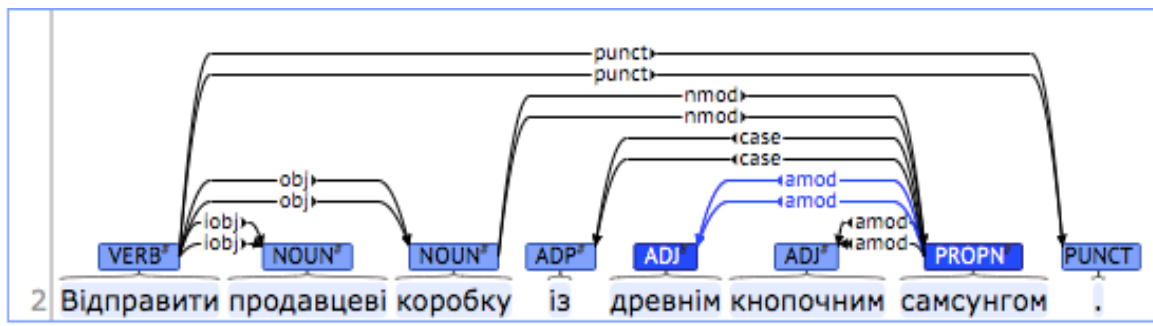
З системи АГАТ було відібрано таксони «оцінка позитивна» та «оцінка негативна» для дієслів, іменників, прикметників та прислівників. На основі цих таксонів було укладено тональний словник української мови для створення автоматичної системи сентимент аналізу.

1.5. Синтаксичні зв'язки в реченні для визначення емотивності тексту

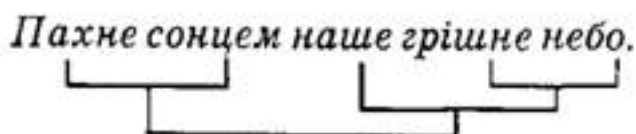
Для визначення сентимент аналізу недостатньо працювати лише на базі словника, тобто визначати сентимент на лексичному рівні. Ми можемо поглиблювати аналіз на рівень словосполучень. Для автоматичного визначення синтаксичних зв'язків слів у реченні існує багато алгоритмів, за допомогою яких можна будувати синтаксичні дерева.

Існує два підходи до побудови синтаксичних дерев: граматики дерев залежностей та граматики безпосередніх складників.

Деревом залежностей називається синтаксична структура, яка використовується для зображення структури підпорядкування в реченнях природної мови. Будь-яке дерево відображає своєрідний граф залежностей.



Грамматика ж безпосередніх складників характеризує синтаксичну структуру речення як ієрархічну систему множин словоформ та їхніх груп. Система безпосередніх складників, утворена граматиною складників, містить інформацію про морфологічні категорії словоформ у вузлах синтаксичної структури [10, с. 42-45].



Для визначення емотивності тексту краще використовувати граматику залежностей, оскільки за її допомогою ми можемо зрозуміти зв'язки між членами речення і таким чином визначити як слова змінюють емотивне забарвлення один одного.

Для англійської мови існує багато інструментів для побудови дерев залежностей, зокрема:

- Stanford parser [50]
- nltk
- Spacy [49]
- Textrazor [53]

Для української мови у 2020 році командою lang_uk [26] було створено застосунок для побудови дерев залежностей за допомогою мови програмування python:

- Stanza [51]

Система працює на базі UniversalDependencies.

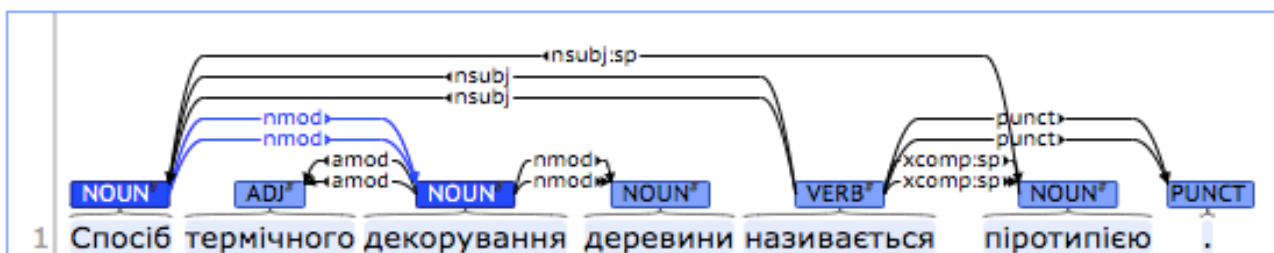
Таким чином, для подальшої роботи над визначенням емотивності тексту, було обрано синтаксичний аналіз на основі побудови дерев залежностей та виявлення основних зв'язків між членами речення.

Оскільки в резидентному словнику містяться такі частини мови, як дієслово, іменник, прикметник, прислівник, було проаналізовано основні характерні зв'язки для української мови та обрано ті, в яких фігурують ці частини мови.

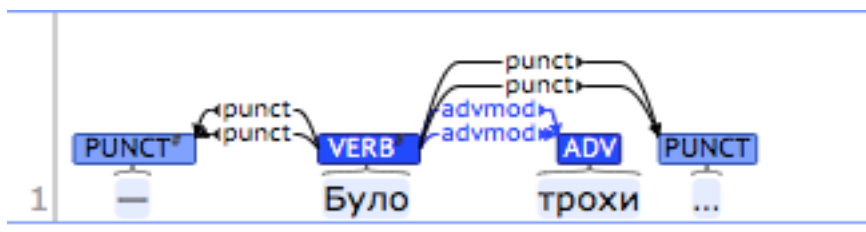
amod - тип підрядного зв'язку між головним і залежним словами, коли форма залежного слова відповідає формі головного, тобто узгоджується з ним у роді, числі, відмінку. Відповідає поєднанню іменника з прикметником.



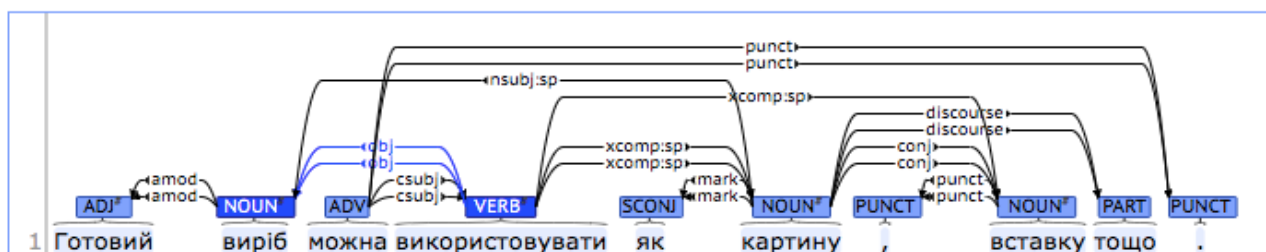
nmod - тип підрядного зв'язку між головним і залежним словами, при якому залежне слово ставиться у відмінку, якого вимагає головне слово. Цей зв'язок відповідає у словосполученні зв'язку керування, а саме іменникове керування.



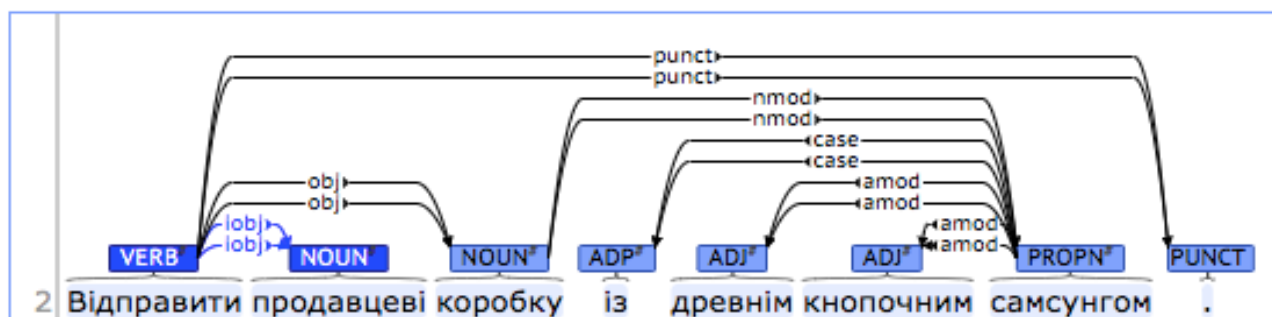
advmod – поєднання дієслова з прислівником.



obj – прямий додаток.



iobj – непрямий додаток.



Для визначення сентименту нам важливо знайти в текстах ці зв'язки, знайти члени словосполучень у словнику і надалі працювати лише з тими, у яких обидва слова є емотивно маркованими.

Емпіричним методом було встановлено закономірність впливу одного слова на інше в межах словосполучень з такими зв'язками: емотивність сполуки визначає залежне слово. Якщо залежне слово має позитивне значення, то і все словосполучення набуває позитивного смислу, і навпаки.

1.6. Висновки до розділу 1

У розділі 1 описані основні теоретичні поняття, пов'язані з категорією емотивності тексту, проаналізована історія розвитку та методи сентимент-аналізу, розглянуто різні автоматичні системи визначення емотивності тексту.

У роботі ми оперуємо поняттями емоція та емотивність. Різниця між цими поняттями полягає в тому, що емоція – це психологічна категорія, тоді як емотивність – мовна категорія. В. Шаховський стверджував, що «емоція постає способом оцінки предметів і явищ реальної дійсності, виникнення оцінного ставлення є умовою виникнення емоції» [41]. Емотивність тексту є однією з основних властивостей тексту, яка актуалізується через емотивно навантажені текстові компоненти, що втілюють авторські емоційні наміри й моделюють ймовірні емоції адресата.

З емотивністю тексту пов'язано поняття тональності тексту, яка може виражатися різними способами: лексикою, словосполученнями, фразеологізмами, порядком слів у реченні, пунктуаційними засобами, графічними засобами.

Для визначення емотивності тексту нам необхідно виділити ті фрагменти тексту, що виражають позитивну або негативну емоційність стосовно об'єкта емоційної оцінки. Тому тональність визначається 3-ма факторами: суб'єктом, власне тональною оцінкою та об'єктом тональності.

На основі аналізу тональності текстів можна виявляти які емоції переважають у тексті: позитивні чи негативні, інтенсивність емоцій у тексті за певною шкалою.

Аналіз методів та існуючих баз даних дозволили визначити основним підходом до створення власної системи сентимент-аналізу – метод на основі тонального словника і правил. Словник укладено на основі семантичного словника таксонів системи АГАТ, а правила за якими визначаються зв'язки за допомогою застосунку для побудови дерев залежностей Stanza мовою програмування python.

РОЗДІЛ 2

СИСТЕМА АВТОМАТИЧНОГО ВИЗНАЧЕННЯ ТОНАЛЬНОСТІ УКРАЇНСЬКОМОВНИХ ТЕКСТІВ

У створенні системи автоматичного визначення тональності українськомовних текстів робота виконувалася у послідовності таких завдань:

1. Створення словника емотивної лексики: на основі семантичного словника таксонів, укладеного на матеріалі лексики публіцистичного стилю Корпусу української мови було створено резидентний тональний словник, представлений у форматі бази даних.

2. Створення корпусу текстів новин ТСН на політичну тематику за двома зонально хронологічними вибірками (2019-2020 років та 2021 року).

3. Розроблення програмного забезпечення спрямованого на виконання таких функцій:

- 1) Зчитування тексту із бази даних текстів;
- 2) Сегментація тексту на речення;
- 3) Сегментація речення на слововживання та їх передобробка: виділення складних слів, скорочень та аббревіатур у текстах, видалення розділових знаків;
- 4) Лематизація слововживань, представлення тексту у лемному записі;
- 5) Зіставлення лем аналізованого тексту із початковою формою тонального резидентного словника (ідентифікація лексики);
- 6) Визначення зв'язків між емоційно-маркованими членами речення за допомогою дерев залежностей;
- 7) Тональне маркування слів-конверторів та слів-інтенсифікаторів у контактній дистрибуції;
- 8) Тональне маркування ідентифікованої лексики;
- 9) Обчислення абсолютної частоти позитивної та негативної лексики – кількість слововживань позитивної та негативної лексики у тексті;
- 10) Обчислення відносної частоти позитивної та негативної лексики;
- 11) Обчислення коефіцієнта тональності;

4. Автоматичне визначення тональності кожного тексту зі сформованого корпусу текстів;
5. Автоматичне укладання частотного словника тональної лексики на базі корпусу текстів;
6. Автоматичне обчислення статистичних характеристик частин мови тонального словника текстів та тонального резидентного словника.

2.1. Укладання тонального резидентного словника української мови

Тональний словник було створено на основі семантичного словника таксонів, укладеного на матеріалі лексики публіцистичного стилю Корпусу української мови, представленого на порталі <http://mova.info>.

Із семантичного словника таксонів було відібрано таксони «оцінка позитивна» та «оцінка негативна» для дієслів, іменників, прикметників та прислівників. На основі цих таксонів було укладено базу даних резидентного тонального словника української мови. База даних містила понад 8 тисяч слів. Додатково були проаналізовані тексти новин ТСН і база була доповнена вручну емотивно-маркованими словами. Загальна кількість слів у базі даних становить понад 10 тисяч.

База даних має таку структуру (див. рис 2.1):

- порядковий номер слова у таблиці(id);
- слово(word);
- частина мови(POS);
- визначення(definition);
- тональність (Sentiment) «+» – позитив, «-» – негатив.

id	word	POS	Definition	Sentiment
Filter	Filter	Filter	Filter	Filter
1	анексіоністський	adj	Прикм. до анек...	-
2	анексований	adj	Дієприкм. пас. ...	-
3	антагоністичний	adj	про людей, угр...	-
4	антагоністичний	adj	Пройнятий анта...	-
5	апатичний	adj	який не виявля...	-
6	апатичний	adj	про людину, по...	-
7	апатичний	adj	який виражає б...	-
8	астенічний	adj	який виснажив...	-
9	астматичний	adj	характерний дл...	-
10	аховий	adj	Поганий, нікуди...	-
11	базарний	adj	Грубий, вульга...	-
12	байдужний	adj	який виражає б...	-
13	байдужний	adj	про людину, як...	-

Рис. 2.1 Структура бази даних тональної лексики

База даних тональної лексики налічує 10193 слова:

- Іменники – 3193;
- Прикметники – 4904;
- Дієслова – 258;
- Прислівники – 1809.

2.2 Укладання корпусу текстів

Для аналізу було обрано тексти новин, тому що сьогодні жанр новин перетворився із інформаційного у інформаційно-аналітичний, маніпулятивний та емоційно-оцінний текст.

Сучасні концепції науковців (Т. Шмельова, О. Тертичний [35], Л. Шевченко, Г. Бобровська [12], О. Ляшенко[19], Т. Добросклонська [11], Л. Дускаєва [13] та ін.) відзначають функціональне розширення жанру новин від інформаційної функції до:

- аналітичної;
- маніпулятивної;

- сугестивної.

Реалізація маніпулятивної функції здійснюється через експлікований або прихований вплив на респондента для формування певного емоційно-психологічного настрою отримувача інформації. Такий вплив на респондента відбувається через активізацію різних елементів психологічного механізму сприймання повідомлення адресатом, зокрема одним з експлікованих засобів емоціогенної маніпулятивної комунікації є емоційно-оцінна лексика тексту.

Для дослідження було обрано тексти новин ТСН на політичну тематику. Відповідно було сформовано дві вибірки, які були сформовані за текстами за період 2019 - 2020 років та 2021 року. Тексти були автоматично відібрані із сайту новин за допомогою бібліотеки scrapy [5].

Вибірка 2019-2020 років налічує 232 010 слововживань, 1000 текстів, кожен з яких є окремою підвибіркою. Середня довжина тексту складає 129 слововживань.

Вибірка 2021 року налічує 181 275 слововживань, 629 текстів, кожен з яких є окремою підвибіркою. Середня довжина тексту складає 288 слововживань.

База даних текстів новин була укладена за допомогою реляційної системи керування базами даних SQLite (див. рис. 2.2).

База даних відображає метатекстову інформацію:

- посилання на текст (url);
- дата публікації статті (date);
- тематика (topic);
- заголовок (title);
- текст статті.

	url	date	topic	title	text
	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр
1	https://tsn.ua/...	2018-04-03	Політика	Адвокат Савче...	Адвокат Савче...
2	https://tsn.ua/...	2018-04-03	Політика	Нардепи запро...	Нардепи запро...
3	https://tsn.ua/...	2018-04-02	Політика	Справа про "ви...	Справа про "ви...
4	https://tsn.ua/...	2018-04-03	Політика	Стало відомо, к...	Стало відомо, к...
5	https://tsn.ua/...	2018-04-03	Політика	Холодницького ...	Холодницького ...
6	https://tsn.ua/...	2018-04-03	Політика	Е-декларації по...	Е-декларації по...
7	https://tsn.ua/...	2018-04-03	Політика	"Не в цьому жи...	"Не в цьому жи...
8	https://tsn.ua/...	2018-04-03	Політика	Холодницький ...	Холодницький ...
9	https://tsn.ua/...	2018-04-03	Політика	Верховна Рада ...	Верховна Рада ...
10	https://tsn.ua/...	2018-04-03	Політика	Реакція на аре...	Реакція на аре...
11	https://tsn.ua/...	2017-11-22	Політика	До України не п...	До України не п...
12	https://tsn.ua/...	2017-08-29	Політика	Вторгнення РФ ...	Вторгнення РФ ...
13	https://tsn.ua/...	2017-07-22	Політика	Грицак назвав і...	Грицак назвав і...
14	https://tsn.ua/...	2017-11-11	Політика	Клімкін: Польщ...	Клімкін: Польщ...
15	https://tsn.ua/...	2017-11-11	Політика	Клімкін: Польщ...	Клімкін: Польщ...

Рис. 2.2 Структура бази даних текстів новин.

2.3 Розроблення програмного забезпечення системи автоматичного визначення тональності українськомовних текстів новин ТСН

Програма була розроблена для проведення автоматичного сентимент-аналізу текстів новин та дослідження ефективності використання методу сентимент-аналізу за тональним словником.

Сентимент-аналіз відбувається у послідовності виконання завдань сформульованих на початку цього розділу. Ці завдання можна згрупувати у 4 етапи:

- Зчитування тексту;
- Лематизація слововживань;
- Пошук слів за резидентним тональним словником у кожному тексті та їх підрахунок;
- Визначення відсотка позитивних і негативних слів у тексті.

1 етап: На першому етапі ми зчитуємо тексти і готуємо їх до обробки: прибираємо непотрібну для аналізу інформацію (посилання на текст, тему, заголовок).

Ділимо текст на речення:

- 1) знаходження в тексті аббревіатур та скорочень слів і заміна крапки (пр. «с. Бочарівка» → «с\$ Бочарівка»), на інший символ, що не є розділовим знаком. Якщо слово містить дефіс, по аналогії замінюємо його на нерозділовий знак «%».
- 2) Поділ відбувається за спеціальними символами з дотриманням такої послідовності: поділ тексту за розділовими знаками, які складаються з декількох символів («...», «!..», «?...»), врахування прямої мови, поділ тексту на речення за («.», «!», «?»).

Фрагмент коду заміни розділових знаків та поділу тексту на речення відображено на рис. 2.3.

```
def split_on_sentence(self, text) -> list:
    t = self.__abbr(text)
    t9 = self.__quest(t)
    split = self.__split_on_3_dots(t9)

    result_sent=list()

    for sent in split:
        result_sent.extend(self.__simple_split_on_sentence(sent))
    result_sent = [i.replace('$', '.') for i in result_sent]
    result_sent = [i.replace('#', '...') for i in result_sent]
    result_sent = [i.replace('&', '?') for i in result_sent]
    result_sent = [i.replace('@', '!') for i in result_sent]
    result_sent = [i.replace('%', '-') for i in result_sent]
    # for i in result_sent:
    #     print(i)
    return result_sent
```

Рис. 2.3. Фрагмент коду поділу тексту на речення

Мовою python був написаний код, в якому ми за допомогою методу replace() замінюємо розділові знаки на певні символи і повертаємо копію рядка, в якій замінені всі входження зазначеного рядка зазначеним символом. Далі ділимо уже текст за розділовими знаками, використовуючи метод split() (див. Додаток 1).

Після поділу на речення, відбувається поділ на слова, з урахуванням усіх особливостей тексту.

Демонструємо роботу програми спрямованої на токенізацію.

На вхід даємо текст (див. рис. 2.4):

Адвокат Савченко заявив про самовідвід. Договір з підзахисною розірвано офіційно. Адвокат Віктор Чевгуз взяв самовідвід у справі позбавленої недоторканності депутатки Надії Савченко. Про це адвокат повідомив на своїй сторінці у Facebook. "Самовідвід. Не хотів подавати, але..." – таким дописом адвокат прокоментував заяву, яку розмістив на сторінці. Із заяви випливає, що Савченко та Чевгуз вже офіційно розірвали договір про надання правової допомоги. Заяву скеровано до СБУ і адвокат чекає на рішення. Лише напередодні Чевгуз коментував справу та заявляв, що готує матеріали для спростування підозри. Про мотиви свого самовідводу адвокат не повідомляє. Нагадаємо, 22 березня Верховна Рада зняла із Савченко недоторканність і дозволила притягнення до кримінальної відповідальності й арешт. Савченко звинувачують у підготовці масштабного теракту проти керівництва країни з метою державного перевороту. Голосувати проти Савченко депутатів переконало 30-хвилинне відео. На ньому вона разом зі своїм спілником Володимиром Рубаном у деталях планують терористичну атаку із тисячами жертв у Києві. Раніше Шевченківський районний суд Києва обрав для Савченко запобіжний захід у вигляді двох місяців за ґратами, а нардеп оголосила голодування. А 29 березня Апеляційний суд Києва залишив чинним рішення Шевченківського райсуду щодо тримання під вартою народного депутата Надії Савченко.

Рис. 2.4 Текст на вхід програми.

На виході отримуємо список слововживань (див. рис. 2.5).

```
Адвокат
Савченко
заявив
про
самовідвід
Договір
з
підзахисною
розірвано
офіційно
Адвокат
Віктор
Чевгуз
взяв
самовідвід
у
справі
позбавленої
недоторканності
депутатки
Надії
Савченко
Про
це
адвокат
повідомив
на
своїй
сторінці
у
Facebook
```

Рис. 2.5 Результат роботи програми поділу речень на слововживання.

- **2 етап:** Лематизація слововживань тексту. Оскільки у тональному словнику у нас містяться лексеми, то для полегшення пошуку лематизуємо

слововживання. Лематизація реалізована за допомогою бібліотеки stanza на python [51].

На вхід програми подаємо текст (див. рис. 2.6):

Суперечки поміж антикорупційними органами в Україні не справляють хорошого враження на Європу - Вагнер. Попри те, що стільки вже зроблено, набагато більше роботи ще попереду, - Вагнер про реформи. У Брюсселі слідкують інформацією щодо подій навколо НАБУ та Антикорупційної прокуратури і "це не справляє хорошого враження". Про це заявив керівник Групи підтримки України в Єврокомісії Пітер Вагнер, - пише "Радіо Свобода". "Ми знаємо про події у Києві. Ми слідкуємо за повідомленнями, хоча це і не наша справа - аналізувати і вирішувати, що саме відбувається...."

Рис. 2.6 Приклад тексту на вхід програми.

На виході отримуємо список лем, проміжним етапом якого був список слововживань (див. рис. 2.7).

Мовою python було написано код, де ми імпортуємо модуль stanza для лематизації (див. Додаток 1).

```
['Суперечки', 'поміж', 'антикорупційний', 'орган', 'в', 'Україна', 'не', 'справляти', 'хороший', 'враження', 'на', 'Європа', ' ', 'Вагнер', 'Попри', 'той', 'що', 'стільки', 'вже', 'зробити', 'набагато', 'більший', 'робота', 'ще', 'попереду', ' ', 'Вагнер', 'про', 'реформа', 'У', 'Брюссель', 'слідкувати', 'інформація', 'щодо', 'подія', 'навколо', 'НАБУ', 'той', 'Антикорупційної', 'прокуратура', 'і', 'цей', 'не', 'справляти', 'хороший', 'враження', 'Про', 'цей', 'заявити', 'керівник', 'Групи', 'підтримка', 'Україна', 'в', 'Єврокомісія', 'Пітер', 'Вагнер', ' ', 'писати', ' ', 'Радіо', 'Свобода', ' ', 'Ми', 'знати', 'про', 'подія', 'у', 'Кий', 'Ми', 'слідкувати', 'за', 'повідомлення', 'хоча', 'цей', 'і', 'не', 'наш', 'справа', ' ', 'аналізувати', 'і', 'вирішувати', 'що', 'самий', 'відбуватися' ]
```

Рис. 2.7 Результат роботи програми лематизації слововживань

3 етап: Пошук і підрахунок позитивних і негативних слів у межах кожного тексту.

```
def emot_analyze(self, text: str):
    emotional_lemmas = list()
    emotion_words = list()
    lemmas = self.lemmatizer.lemmatize(text)
    len_lemmas = len(lemmas)
    emotional_lemmas = [self.dict_of_emotions.get(x, None) for x in lemmas]
    emotions = self.dict_of_emotions.items()
    for word in emotions:
        for x in lemmas:
            if x in word:
                emotion_words.append(word)
            else:
                continue
    plus_emotions = emotional_lemmas.count('+')
    minus_emotions = emotional_lemmas.count('-')
    print('Позитивні емоції:', plus_emotions, '\n', 'Негативні емоції:', minus_emotions)
    return emotional_lemmas, emotion_words, len_lemmas
```

Рис. 2.8 Фрагмент коду пошуку емотивно маркованої лексики у тексті

Леми лематизованого тексту зіставляються зі словами у початковій формі резидентного тонального словника (див. рис. 2.8). Якщо лема є у резидентному словнику, то їй присвоюється значення «+» або «-». Якщо леми немає, то лема набуває значення «None». У результаті роботи програми (див. рис 2.9) отримуємо тональні маркери, які заміщують лему тексту.

Для цієї процедури був написаний код мовою python, де використано основні методи роботи зі списками та словниками (див. Додаток 1).

Рис.2.9 Тональна структура тексту на лексичному рівні інтерпретації

4 етап: Тональне маркування слів-конверторів та слів-інтенсифікаторів у контактній дистрибуції.

У реченні дуже часто значення слова може змінюватися за допомогою слів-конверторів. В основному, це заперечні частки «не, ні, ані» та прийменники «без, крім, окрім», які стоять перед словом і можуть змінювати сентимент слова. у нашій програмі це враховано, і тому відбувається перевірка наявності слів-конверторів у репозиції до слів з мітками «+» або «-».

Також у тексті можуть зустрічатися слова, які підсилюють сентимент того чи іншого слова. До них належать модальні частки та прислівники «дуже, надзвичайно, безмежно, вкрай, досить, зовсім, так, над, аж, таки». Програма перевіряє наявність таких слів-інтенсифікаторів у тексті, і додає до слова показник «+» або «-», залежно від сентименту слова, знайденого в тексті (див. рис. 2.10).

Рис.2.11 Визначення тональності синтагми зі словами-конверторами та словами-інтенсифікаторами

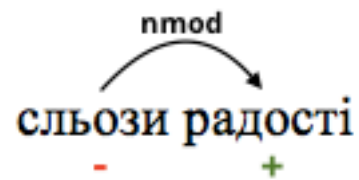
5-етап: Визначення синтаксичних зв'язків між членами речення за допомогою дерев залежностей.

Для цього етапу використовуємо бібліотеку stanza [51] і визначаємо основні синтаксичні зв'язки (amod, advmod, nmod, obj, iobj) тонально маркованих слів в аналізованих текстах.

У результаті роботи синтаксичного аналізатора визначаються слова, які зв'язані з емотивно маркованими словами тексту. Ці слова зіставляються з реєстром резидентного тонального словника для встановлення їхньої тональності (див. рис. 2.12).

Якщо два слова пов'язані між собою та обидва є емотивно маркованими, то визначаємо сентимент сполуки залежно від сентименту підпорядкованого слова.

Наприклад, знаходимо сполуку *сльози радості* марковану у тексті зв'язком nmod. Обидва слова є емотивно-маркованими: *сльози* має негативне забарвлення, *радості* - позитивне. Оскільки залежне слово у сполуці позитивно-марковане, то вся сполука отримує позитивне забарвлення «+».



Оскільки у тональному словнику наявні лише прикметники, іменники, прислівники та дієслова, було проаналізовано та виявлено основні зв'язки, характерні для цих частин мови:

- amod;
- nmod;
- obj;
- iobj;
- advmod.

```

##dependency relations
for token in token_sentence:
    head = next((t for t in token_sentence if t.word_id == token.head_id), None)
    #nmod_relations
    if token.deprel == 'nmod' and head and head.sentiment_label is not None and head.sentiment_label != token.sentiment_label:
        head.sentiment_label = token.sentiment_label

    #amod_relations
    if token.deprel == 'amod' and head and head.sentiment_label is not None and head.sentiment_label != token.sentiment_label:
        head.sentiment_label = token.sentiment_label

    # obj_relations
    if token.deprel == 'obj' and head and head.sentiment_label is not None and head.sentiment_label != token.sentiment_label:
        head.sentiment_label = token.sentiment_label
    if token.deprel == 'iobj' and head and head.sentiment_label is not None and head.sentiment_label != token.sentiment_label:
        head.sentiment_label = token.sentiment_label

    #advmod_relations
    if token.deprel == 'advmod' and head and head.sentiment_label is not None and head.sentiment_label != token.sentiment_label:
        head.sentiment_label = token.sentiment_label

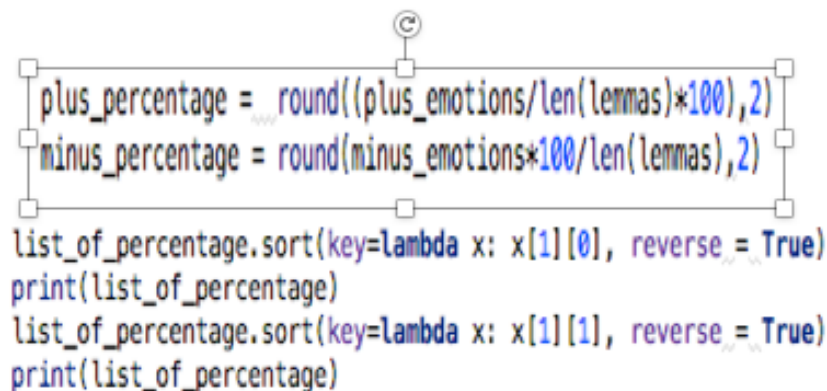
```

Рис. 2.12. Фрагмент коду визначення синтаксичних зв'язків

Для цієї процедури був написаний код мовою python, де визначаються зв'язки та змінюється сенсмент в залежності від підпорядкованого слова.

6 етап: Обрахунок емотивно маркованих слів у тексті

На вхід подаємо текст, на виході отримуємо кількість емотивних слів, та відсоткове співвідношення емотивно забарвлених слів до загальної кількості слововживань тексту (див. рис. 2.13).



```

plus_percentage = round((plus_emotions/len(lemmas)*100),2)
minus_percentage = round(minus_emotions*100/len(lemmas),2)
list_of_percentage.sort(key=lambda x: x[1][0], reverse = True)
print(list_of_percentage)
list_of_percentage.sort(key=lambda x: x[1][1], reverse = True)
print(list_of_percentage)

```

Рис. 2.13 Фрагмент коду обрахунку відсоткового співвідношення емотивно забарвлених слів у тексті

Також обробляємо обраховані дані і вибираємо тексти з найвищими показниками позитивних і негативних емоцій. Кожен текст отримує числову характеристику (див. рис. 2.14), створюється база даних статистичних характеристик тональності текстів (див. Додаток 1).

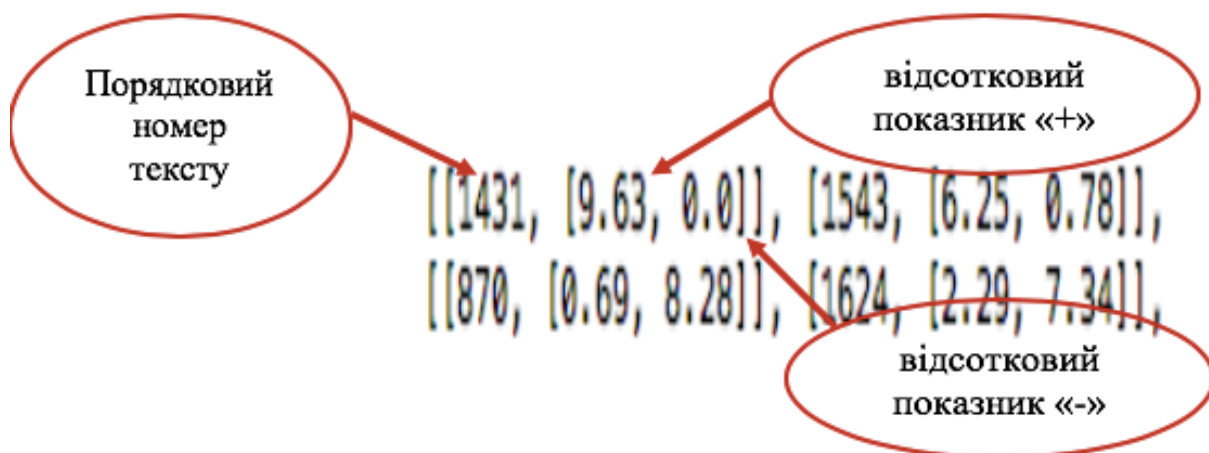


Рис.2.14 Статистична характеристика тональності проаналізованого тексту

У результаті отримуємо два відсортованих списки від найбільшого до найменшого за відсотковим показником «+» або «-». Список складається з номера тексту та відсоткового показника «+» або «-» для кожного тексту. Усі ці показники вносяться автоматично в базу даних емотивних текстів.

Структура бази даних статистичних характеристик тональності текстів (див. рис. 2.15):

- id – номер тексту;
- plus – відсотковий показник «+»;
- minus – відсотковий показник «-».

	id	plus	minus
	Фільтр	Фільтр	Фільтр
1	0	1.07	1.07
2	1	1.08	0
3	2	0	0.96
4	3	0	0
5	4	3.28	0
6	5	2.2	0.22
7	6	0.67	0.34
8	7	1.45	0
9	8	0.62	0.62
10	9	0	3.41
11	10	0	3.77
12	11	1.99	0.57
13	12	0.22	0.43
14	13	1.39	1.39

Рис.2.15 Структура бази даних статистичних характеристик тональності текстів.

2.4 Результати роботи системи автоматичного визначення тональності українськомовних текстів

Було розроблено програму, яка автоматично шукає за тональним словником слова і підраховує скільки “позитивних” та “негативних” слів зустрілося у тексті.

На вхід програми подається будь-який текст (див. рис.2.16).

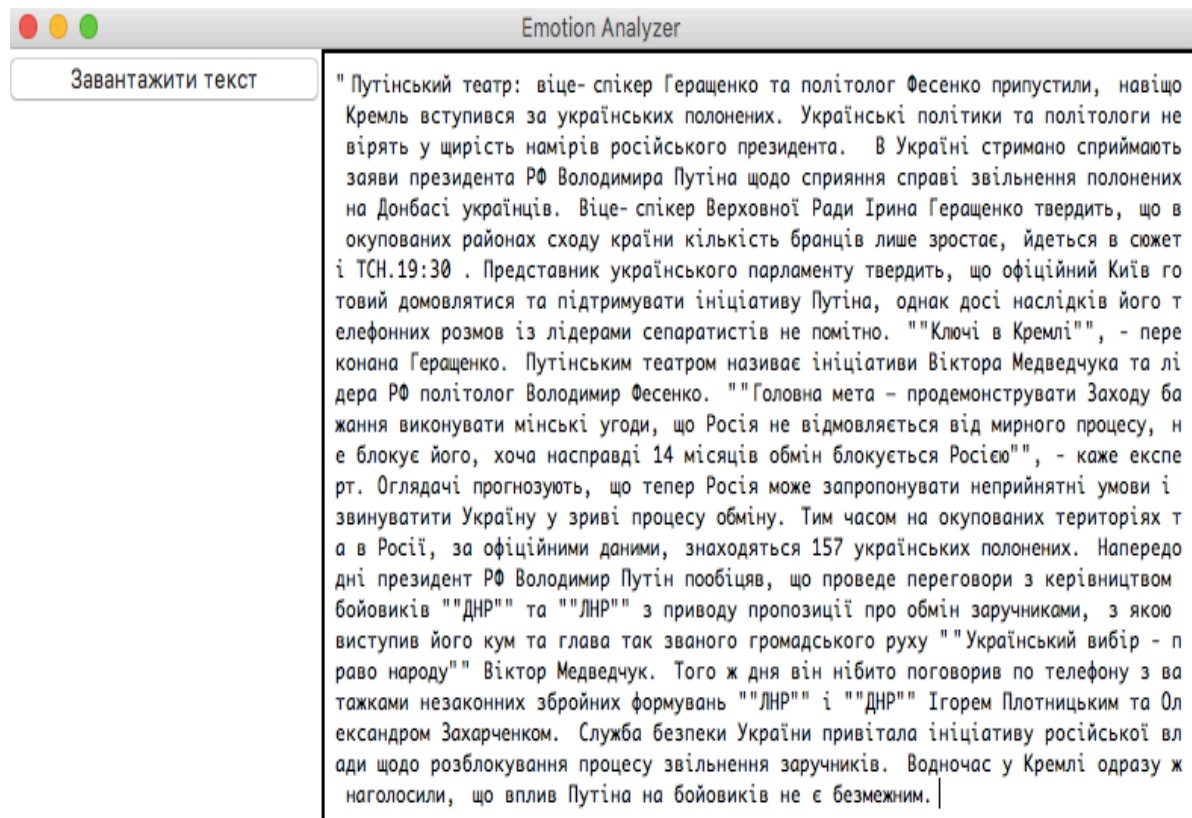


Рис.2.16 Інтерфейс програми

Після натискання на кнопку «Завантажити текст», він автоматично поділяється на слова, здійснюється лематизація словоформ.

Після того, як ми ввели текст, автоматично відбувається пошук і підрахунок позитивних і негативних слів у словнику за текстом (див. рис. 2.17). На виході отримуємо кількість емотивних слів, відсоток емотивних слів у тексті та список емотивних слів, визначених в аналізованому тексті:

- 1) позитивна лексика
- 2) негативна лексика.

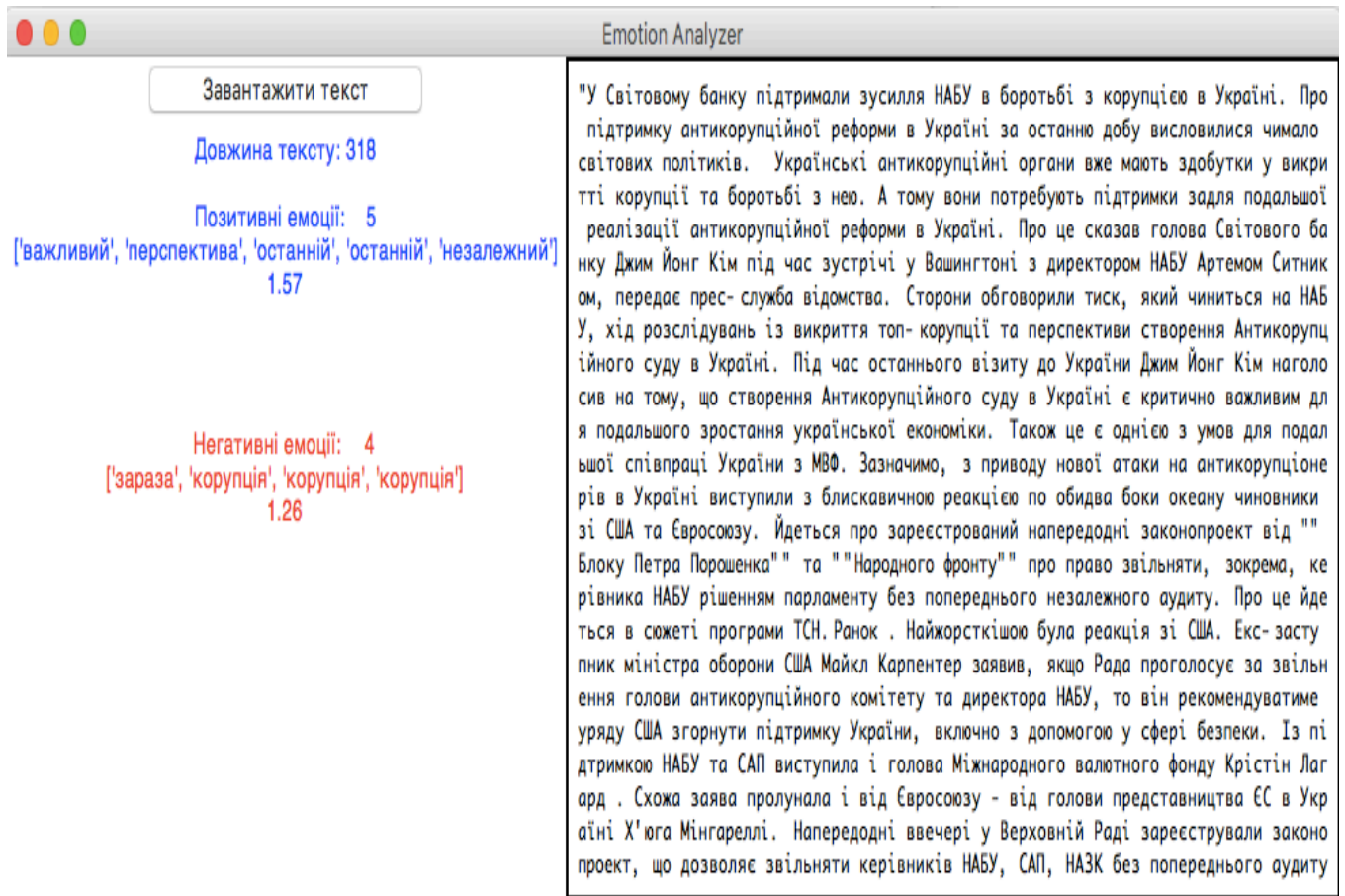


Рис.2.17 Приклад роботи програми

Тепер розглянемо тексти, які мають найбільший відсоток позитивної та негативної лексики.

Суд надав НАБУ доступ до документів лотереї щодо півмільйонних виграшів Ляшка. Нардепа перевіряють, чи не отримав він неправомірну вигоду. Солом'янський районний суд Києва надав детективам Національного антикорупційного бюро України доступ до документів "Української національної лотереї" щодо виграшів нардепа Олега Ляшка. Про це повідомляє "Судовий репортер" із посиланням на ухвалу суду. Детективи протягом місяця матимуть доступ до документів щодо всіх виграшів Ляшка за період з 1 по 24 жовтня 2017 року включно. Політик задекларував, що тричі виграв у лотерею "Українська національна лотерея". Сума виграшів становить 571 тисяч 45 грн. Згодом радикал розповів, що підзаробив завдяки ставкам на футбольні матчі Ліги чемпіонів. Відповідно до листа "Української національної лотереї", протягом жовтня 2017 року виграшів у таких сумах не зареєстровано. У листі підприємства також йдеться, що інформація про виграші у лотерею та персональні дані конкретних гравців є конфіденційною.

Рис. 2.18 Приклад тексту із найбільшим відсотком позитивної лексики
 Майже 10% позитивного тексту становлять слова з позитивною оцінкою.

У тексті взагалі немає слів на позначення негативу. До слів які несуть позитивний сентимент належать «лотерея» (зустрічається у тексті 6 разів), *виграш* (зустрічається у тексті 6 разів), «вигода» (див. рис. 2.19).

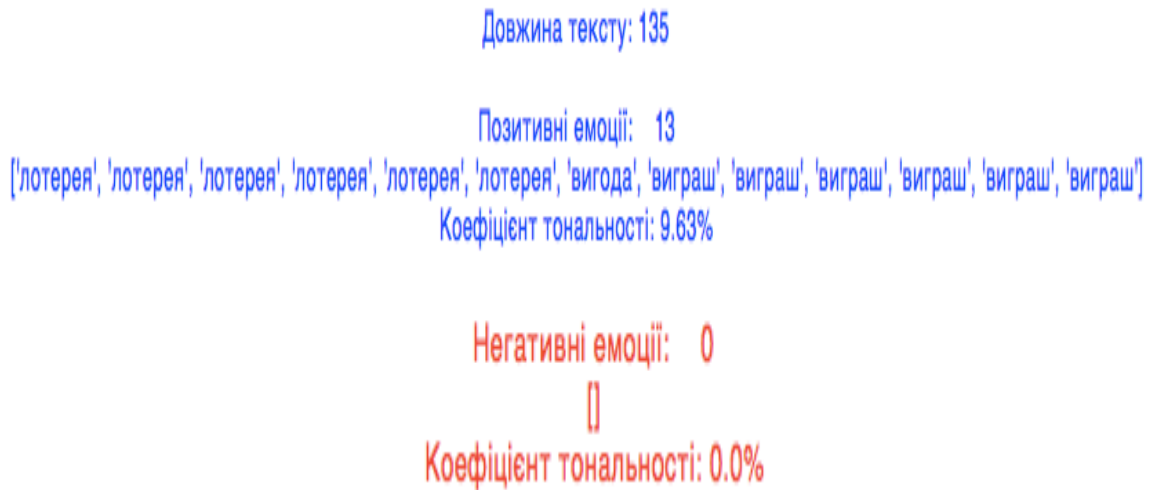


Рис.2.19 Метричні характеристики тональності тексту з найбільшим відсотком позитивної оцінки

Тепер розглянемо текст із найбільшим відсотком негативної оцінки (див. рис. 2.20).

Порошенко про миротворців на Донбасі: Росія буде змушена вивести окупаційні війська. Він додав, що це також вирішить гуманітарну ситуацію. Миротворці на Донбасі зможуть роззброїти терористів на Донбасі, сприяти виведенню російських військ і вирішити проблему гуманітарної катастрофи на окупованих територіях і лінії зіткнення. Про це президент України Петро Порошенко сказав на прес-конференції. "Це не заморожена фаза, це гаряча війна. Миротворці мають діяти на всій окупованій території. Росія вимушена буде вивести свої окупаційні війська. Світ буде бачити", - сказав Порошенко. Президент також додав, що миротворці роззброють терористів, візьмуть під контроль неконтрольовану ділянку кордону між Україною та Росією і зупинити вплив військ і зброї з країни-агресора на окупований Донбас. Також Порошенко наголосив, що у разі появи в ОРДЛО миротворців, Україна зможе відновити гуманітарні місії на окуповані території і ліквідувати гуманітарну катастрофу. Нагадаємо, сьогодні, 28 лютого, у Мистецькому Арсеналі у Києві відбувається прес-конференція президента України Петра Порошенка.

Рис.2.20 Приклад тексту із найбільшим відсотком негативної лексики. Більш ніж 6% тексту становлять слова на позначення негативної оцінки. У тексті взагалі немає слів на позначення позитивної оцінки. До слів, які несуть негативну семантику належать «вимушений», «окупований» (зустрічається у тексті 4 рази), «змушений», «окупаційний», «лютий» (див. рис. 2.21).

Негативні емоції: 9
('вимушений', 'лютий', 'окупований', 'окупований', 'окупований', 'окупований', 'окупаційний', 'окупаційний', 'змушений')
Коефіцієнт тональності: 6.25%

Довжина тексту: 144

Позитивні емоції: 0

□

Коефіцієнт тональності: 0.0%

Рис.2.21 Метричні характеристики тональності тексту з найбільшим відсотком негативної оцінки

До того ж у цьому тексті оцінку виражають лише прикметники.

Наведемо приклад нейтрального тексту (див. рис. 2.22).

Стало відомо, коли Тристороння контактна група обговорить звільнення заручників. Геращенко повідомила, що зустрічалася з президентом України Петром Порошенком для обговорення питань про звільнення заручників. Тристороння контактна група з врегулювання ситуації на Донбасі прагне обговорити звільнення заручників 4 квітня. Про це повідомила перша заступниця голови Верховної Ради, представниця України у підгрупі з гуманітарних питань ТКГ Ірина Геращенко у Facebook. «Ми прагнемо, щоб наступна хвиля звільнення відбулася якомога швидше, і будемо говорити про це на гуманітарній підгрупі в Мінську 4 квітня», – зауважила вона. Геращенко додала, що зустрічалася з президентом України Петром Порошенком для обговорення питань про звільнення заручників та політичних ув'язнених. Також обговорили відсутність доступу до ув'язнених лікарів, представників Міжнародного комітету Червоного Хреста та порушення права ув'язнених на зустріч з консулами. Нагадаємо, 26 березня Тристороння контактна група щодо Донбасу за ініціативою української сторони домовилася про перемир'я у зоні АТО напередодні Великодня.

Рис.2.22 Приклад нейтрального тексту

У цьому тексті немає слів на позначення як позитивної оцінки, так і негативної оцінки (див. рис. 2.23).

Довжина тексту: 139

Позитивні емоції: 0

□

Коефіцієнт тональності: 0.0%

Негативні емоції: 0

□

Коефіцієнт тональності: 0.0%

Рис.2.23 Метричні характеристики тональності нейтрального тексту

Також було проаналізовано заголовки текстів новин 2021 року. Кількість слововживань у заголовках становить 8557.

Порівнюючи коефіцієнт тональності текстів новин з коефіцієнтом тональності заголовків, бачимо, що деякі тексти мають схожі показники. Наприклад, заголовок *У Молдові зафіксовано новий "коронавірусний" рекорд померлих після зараження* має найбільший відсоток негативної лексики. Текст цієї статті також має найбільший відсоток негативної лексики.

Наведемо ще один приклад. Заголовок *Безоплатні курси з самооборони Жіночий десант з'явилися в Рівному та Кам'янському* має один з найбільших відсотків позитивної лексики. Текст цієї статті також має один з найбільших відсотків позитивної лексики.

Оскільки у заголовках зосереджена головна інформація статті, у деяких текстах заголовків відсоток позитивної/негативної лексики є більшим відносно показників за їхніми текстами. Наприклад, *Прем'єр Словаччини особисто попросив вибачення за жарт про Україну* має найвищий відсоток позитивної лексики, тоді як текст цієї новини є не настільки насиченим позитивною лексикою.

Ще одним прикладом є заголовок *Назавжди з Росією: у Москві зробили скандальну заяву щодо анексованого Криму*, який має набагато вищий відсоток негативної лексики порівняно з текстом цієї статті.

Отже, у заголовках зосереджена більша кількість емотивно маркованих слів. На відсоток емотивності впливає також і довжина тексту. Оскільки заголовки складаються з декількох слів, одного речення, відсоток їхньої емотивності буде набагато більшим. Заголовки є настільки насиченими емотивно маркованими словами задля привертання уваги читача до статті.

2.5. Зіставлення автоматичної тональної оцінки з тональною оцінкою людини

Для оцінки ефективності роботи системи автоматичного визначення тональності тексту української мови було проведено опитування. З анкетною детальніше можна ознайомитися за посиланням

[<https://forms.gle/LQkkbrX1VgKS7QTt7>]. Респондентам було надано приклади

речень із текстів новин, які вони мали оцінити за шкалою: позитивний, негативний, нейтральний.

До участі в опитуванні було залучено 97 носіїв української мови, різної статі, професії та віку.

Серед інформантів за віком:

4,2% становлять інформанти 15-19 років, 40%- інформанти 20-24 років, 28,4% - інформанти 25-29 років, 30%- інформанти 30-40 років, 5,3%- інформанти 40-50 років, 2% - інформанти 51-59 років, 1% - інформанти 60+ років.(див. рис. 2.24)

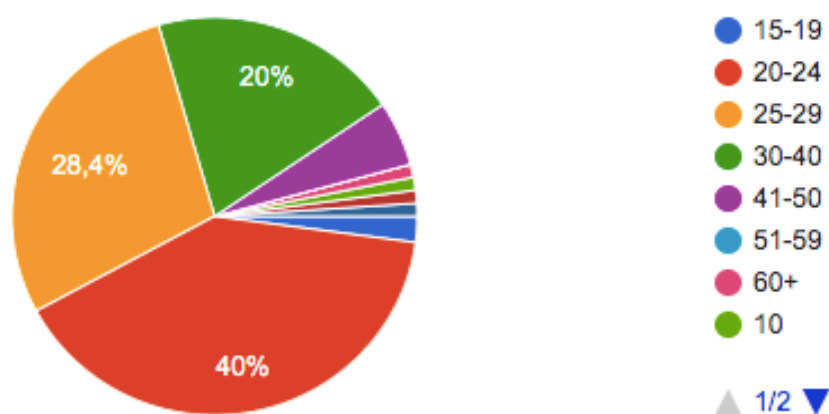


Рис. 2.24 Віковий розподіл інформантів

За гендерним розподілом 79% жінки та 21% чоловіки(див. рис. 2.25):

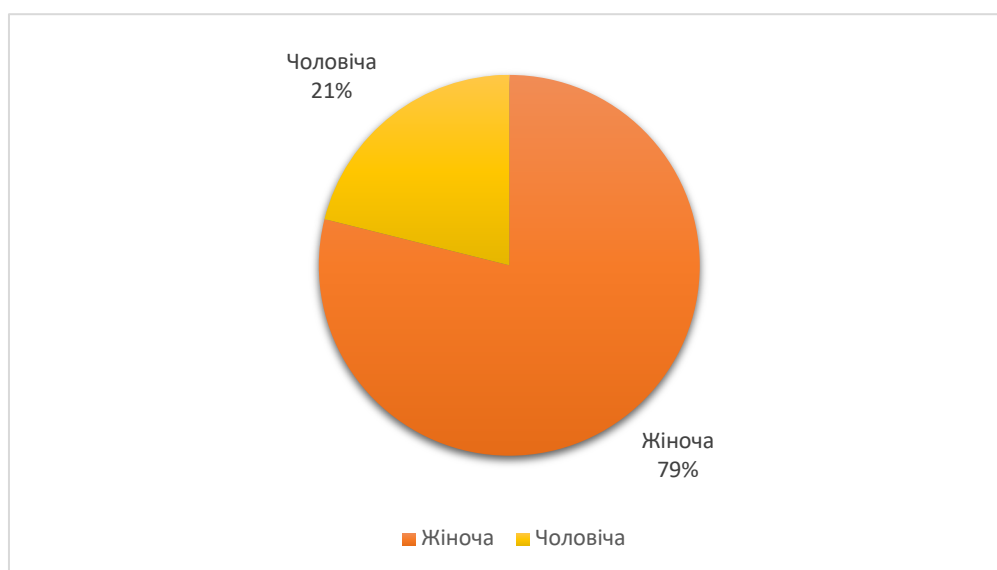


Рис. 2.25 Гендерний розподіл інформантів

За освітою більшість інформантів з вищою освітою(83,2%), 11,6% інформантів з незакінченою вищою освітою, середня освіта - 4% інформантів, 1% - становлять інформанти з незакінченою середньою освітою (див. рис. 2.26).

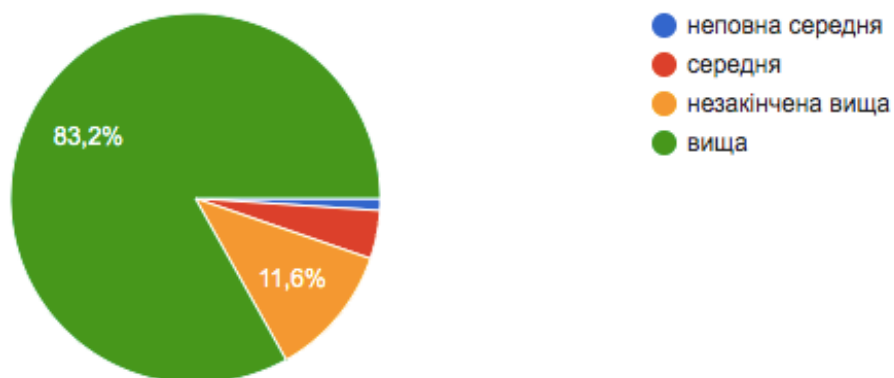


Рис. 2.26 Розподіл інформантів за освітою

За професією (див. рис. 2.27):

Працюють в ІТ сфері – 71%

У сфері викладання / перекладу – 18%

У сфері фітнес-індустрії – 5%

Інші сфери(економіка, маркетинг, менеджмент, туризм, логістика) –4%

Студенти, школярі – 2%

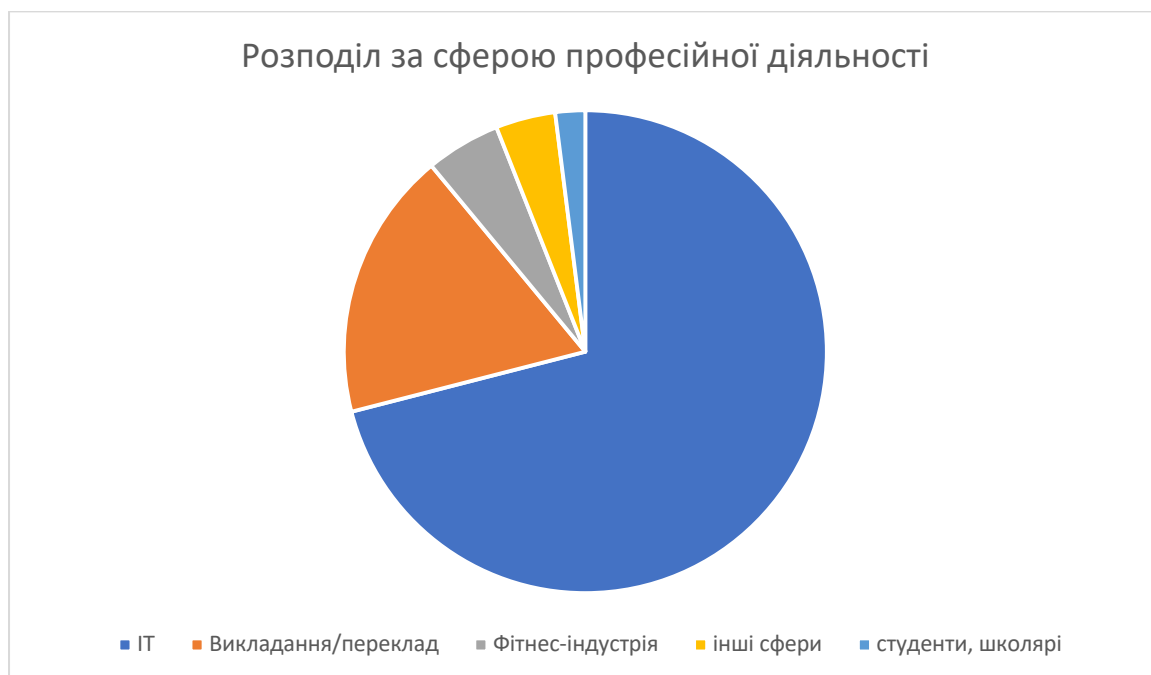


Рис. 2.27 Розподіл інформантів за сферою професійної діяльності

Респондентам було надано 14 речень (див. Додаток 2) для визначення емоційного забарвлення цих речень (позитивне, негативне, нейтральне).
Всі 14 речень були оцінені респондентами і автоматичною системою тонального аналізу, дані сентимент-аналізу представлені у табличному запису у додатку. для зіставлення емоційних оцінок, зроблених людиною і комп'ютером, розглянемо деякі речення.

У ході аналізу відповідей респондентів було виявлено, що деякі речення були визначені як позитивні/негативні за більшістю відповідей.
Наприклад, розглянемо, як респонденти оцінили речення *По її щоках текли сльози радості*. Більшість стверджує що це речення є позитивним (див. рис. 2.28).

По її щоках текли сльози радості.

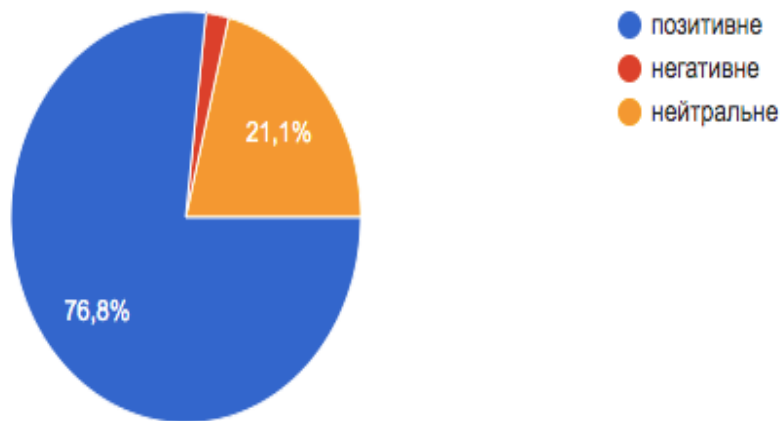


Рис. 2.28 Оцінка речення респондентами

Тепер перевіримо, як система визначає емотивну оцінку цього речення.

Система також визначає це речення позитивним, оскільки за синтаксичними зв'язками словосполучення *сльози радості* визначається як позитивне (див. рис. 2.29).

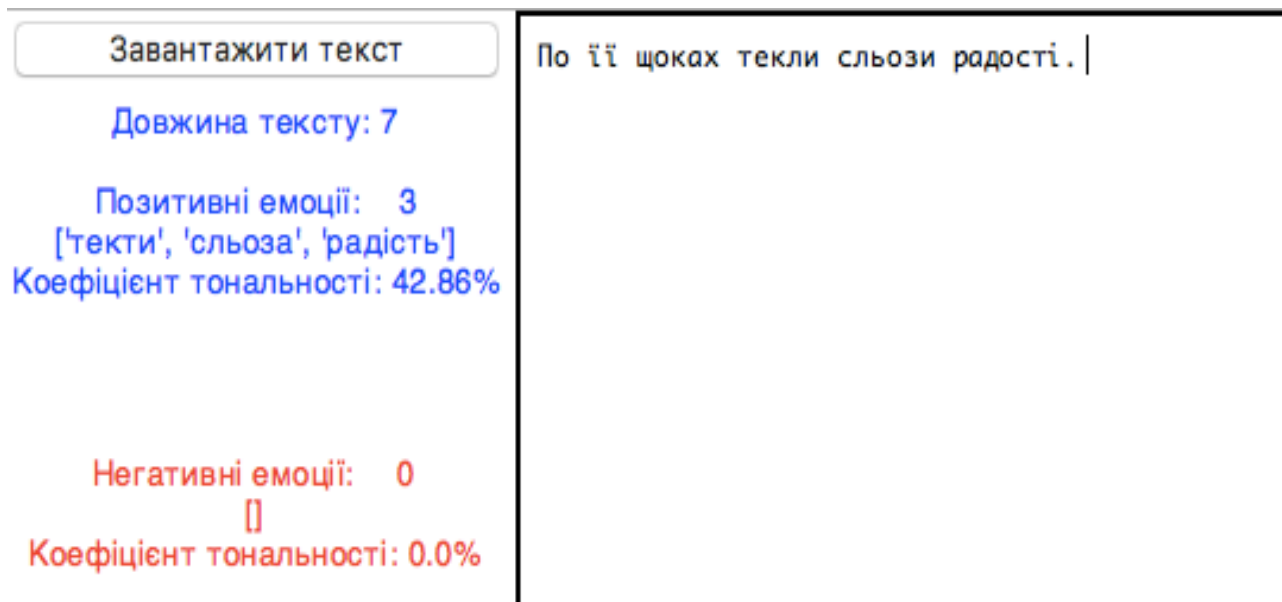


Рис. 2.29 Оцінка речення системою

Розглянемо ще одне речення: *У США молода дівчина померла після невдалої ін'єкції*. З відповідей респондентів, ми бачимо що 87% вважають це речення негативним (див. рис. 2.30).

У США молода дівчина померла після невдалої ін'єкції.

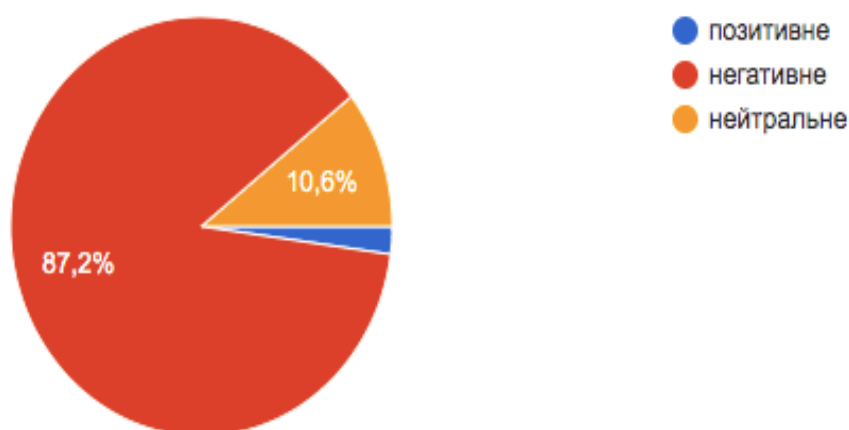


Рис. 2.30 Оцінка речення респондентами

Що ж стосується роботи системи, коефіцієнт тональності негативної лексики у реченні становить 22%, тоді як коефіцієнт тональності позитивної лексики дорівнює 0%. Відповідно, можемо зробити висновок, що це речення система

визначає як негативне (див. рис. 2.31).

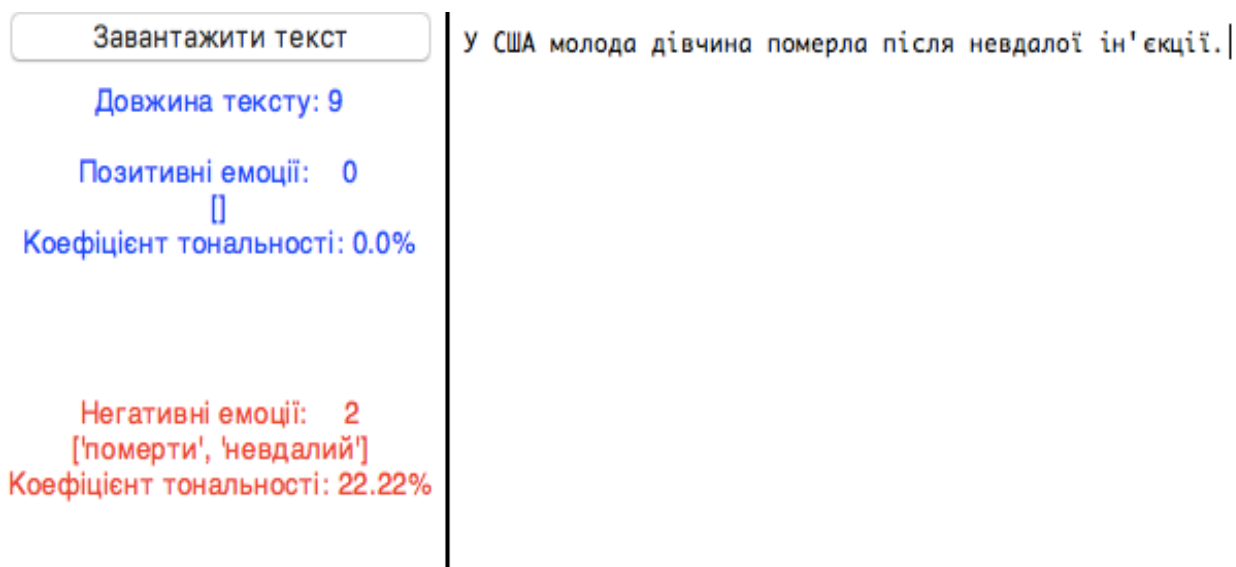


Рис. 2.31 Оцінка речення системою

Також є речення, в яких оцінка респондентів не має одностайності.

Розглянемо як респонденти оцінили речення *Росія вже виробила імунітет до санкцій, тому їх треба посилювати*. З наведеної діаграми (див. рис. 2.32) видно, що думки респондентів розійшлися: 41,1% вважає, що це речення є негативним і такий самий відсоток респондентів вважає, що це речення нейтральне.

Росія вже виробила імунітет до санкцій, тому їх треба посилювати.

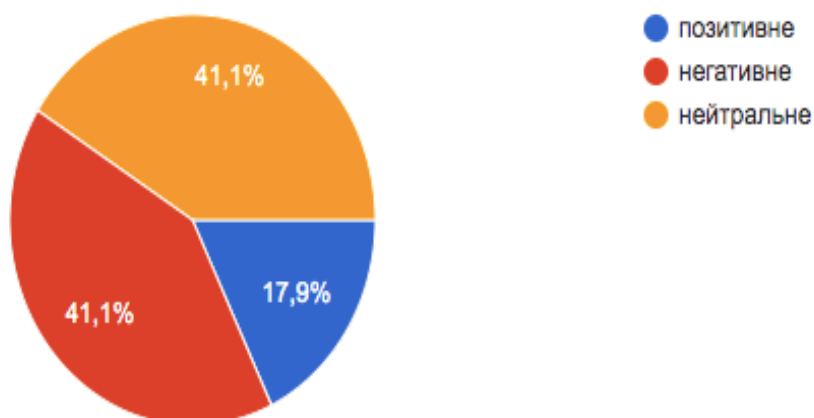


Рис. 2.32 Оцінка речення респондентами

Тепер перевіримо, як система визначає емотивне навантаження цього речення.

За результатами автоматичної оцінки програми (див. рис. 2.33) бачимо, що слів

на позначення негативних емоцій більше, відповідно текст є негативним – відсоток негативності становить 16,67%.

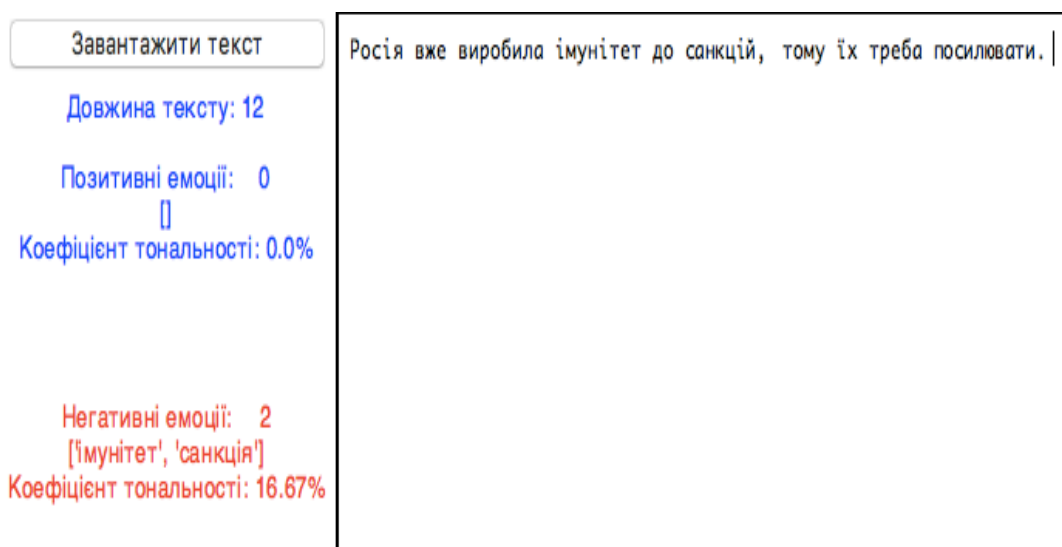


Рис. 2.33 Оцінка речення системою

Розглянемо ще один подібний приклад: *Правда, яку ми почуємо на конференції, може бути болючою, але вона визволяє.* Думки респондентів розійшлися: 48% вважає, що речення є нейтральним, а 37% вважає його позитивним (див. рис. 2.34).

Правда, яку ми почуємо на конференції, може бути болючою, але вона визволяє.

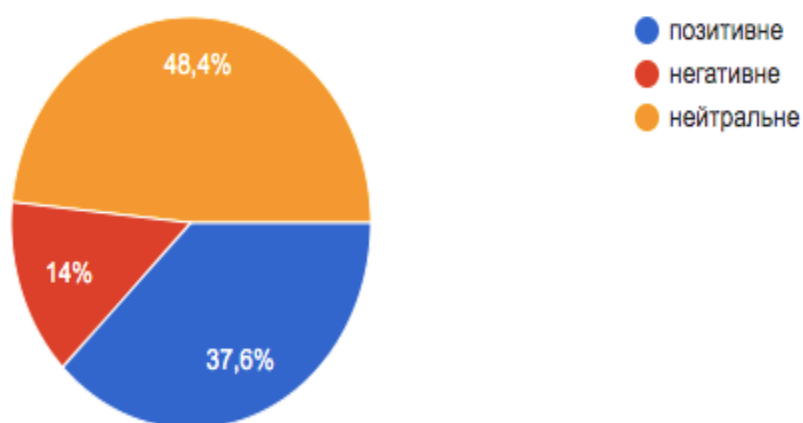


Рис. 2.34 Оцінка речення респондентами

Перевіривши як система визначає тональність цього речення, ми бачимо (див. рис. 2.35), що відсоток позитивності більший за негативний.

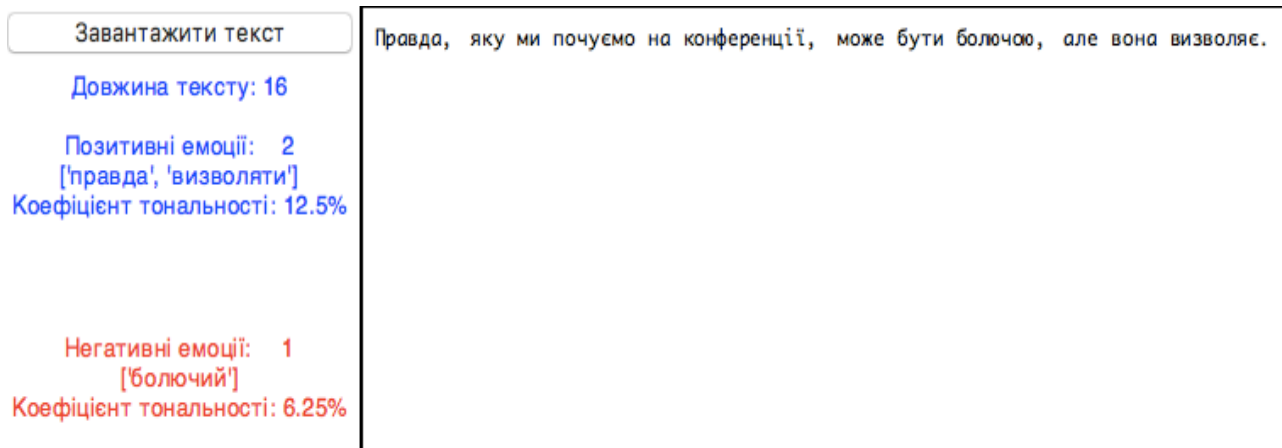


Рис. 2.35 Оцінка речення системою

Також є такі приклади, коли і система, і людина визначає речення як нейтральне. Ось наприклад, у реченні *Люди були засмучені цими звітками, але гірка правда краще за солодку брехню* відповіді респондентів різні - 41% вважає його нейтральним і 34% - позитивним (див. 2.36).

Люди були засмучені цими звітками, але гірка правда краще за солодку брехню.

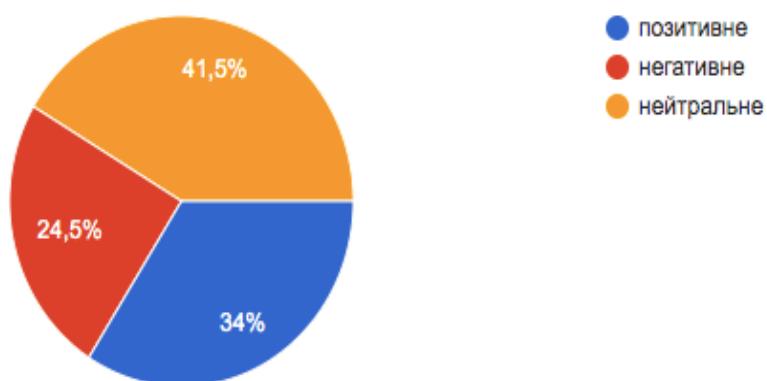


Рис. 2.36 Оцінка речення респондентами

За результатами роботи системи (див. рис. 2.37), можемо сказати, що речення є нейтральним, оскільки відсоткові показники позитивності/негативності тексту є однаковими.

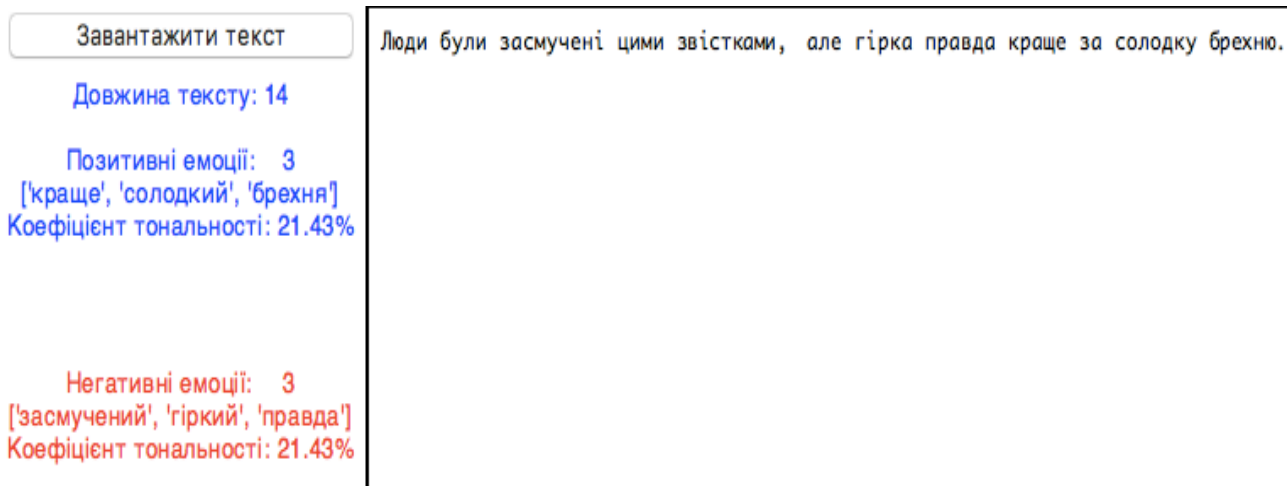


Рис. 2.37 Оцінка речення системою

Були виявлені речення, які система та людина сприймають абсолютно по-різному. Наприклад, речення *Денис Шмигаль доручив міністрам розробити алгоритм відшкодування втрат людям, які були змушені припинити свою діяльність через запровадження червоного рівня епідемічної небезпеки.* більшість респондентів визначило як позитивне(52%) (див. рис. 2.38).

Денис Шмигаль доручив міністрам розробити алгоритм відшкодування втрат людям, які були змушені припинити свою діяльність через запровадження червоного рівня епідемічної небезпеки.

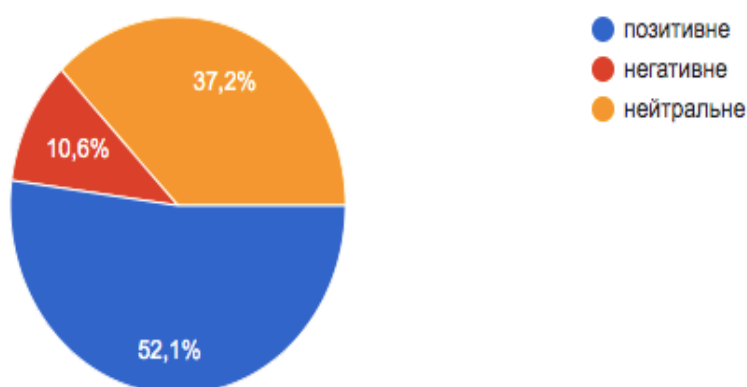


Рис. 2.38 Оцінка речення респондентами

За результатами роботи програми, бачимо, що це речення є негативно-маркованим(див. рис. 2.39).

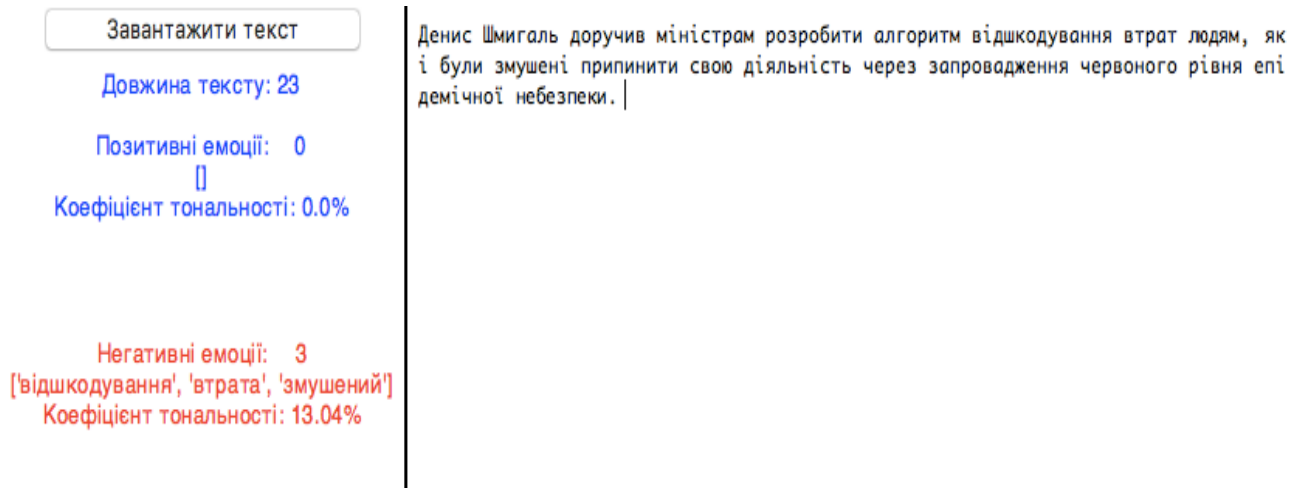


Рис. 2.39 Оцінка речення респондентами

Речення *Уряд запровадив систему підтримки бізнесу, яка дозволить максимально зняти фінансове навантаження на підприємців у період карантинних обмежень*. 73% визначили як позитивне (див. рис. 2.40).

Уряд запровадив систему підтримки бізнесу, яка дозволить максимально зняти фінансове навантаження на підприємців у період карантинних обмежень.

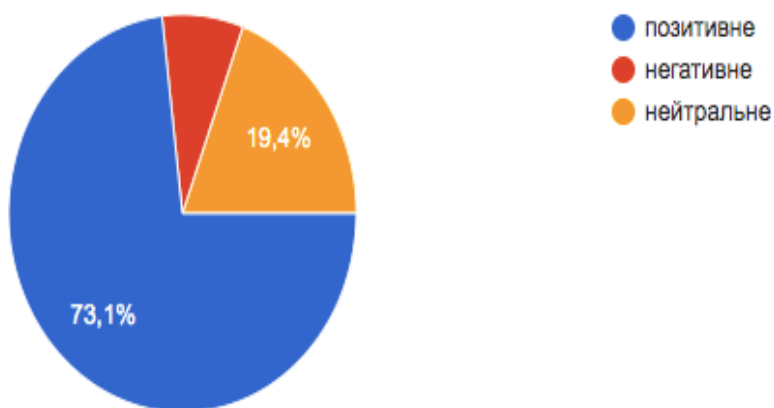


Рис. 2.40 Оцінка речення респондентами

Якщо ж перевірити це речення в системі автоматичного визначення тональності, бачимо, що відсоток негативності більший (див. рис. 2.41).

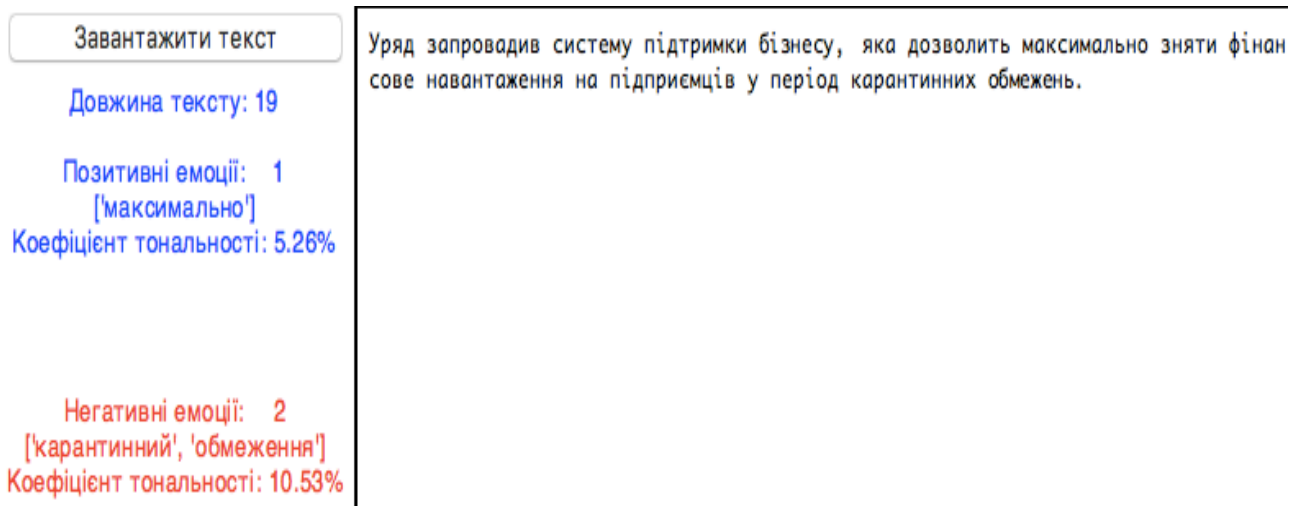


Рис. 2.41 Оцінка речення системою

Проаналізувавши роботу системи автоматичного визначення тональності текстів, а також порівнявши отримані результати з оцінкою людини, можна стверджувати, що за результатами роботи системи ми можемо досить чітко визначити наскільки текст є емотивно-маркованим і віднести його до позитивного чи негативного.

Є такі випадки, коли людина по-іншому визначає емоційне забарвлення тексту.

На це впливають досвід людини, певні соціально-культурні чинники.

Лінгвістичними факторами, які впливають на сприйняття людини, можуть бути будова речення (складносурядні протиставні речення). Наприклад, у першій частині речення подається негативна інформація, а у другій протиставній – позитивна. Таким чином, друга частина речення компенсує першу і людина сприймає це речення, як позитивне.

2.6. Тональна класифікація українськомовних текстів

Визначені метричні тональні характеристики кожного тексту демонструють найрізноманітніші співвідношення негативної та позитивної лексики. Методом суб'єктивного аналізу (лише 2 респонденти) текстів з урахуванням метричних характеристик було встановлено міру коефіцієнта тональності тексту: за умови зіставлення 2-ох коефіцієнтів тональності вони повинні відрізнятися не менше ніж на 3,5. $|K_n - K_p| \geq 3,5$ різниця “по модулю” коефіцієнтів встановлює тональність тексту: якщо $K_n > K_p$ на 3.5 і більше –

негативна тональність; якщо $K_p > K_n$ на 3.5 і більше – позитивна тональність. Усі інші тексти відносимо до нейтральних.

Серед 1000 текстів 2019-2020 років 560 текстів є нейтральними. Серед них 54 текстів з показниками 0/0, тобто в таких текстах взагалі не зустрічаються емотивні слова. Кількість позитивних текстів становить 290, а негативних - 129. Але чи можна стверджувати, що ці тексти ми відносимо до позитивних чи негативних? Проаналізувавши тексти на вміст емотивної лексики, віднести до позитивних або негативних можна, коли Коефіцієнт емотивності відрізняється більше ніж на 3.5-4%. Усі інші тексти відносимо до нейтральних. (див. Табл. 2.1)

Таблиця 2.1 Тональна класифікація текстів новин 2019-2020 років

Нейтральні	560 текстів
Позитивні	290 текстів
Негативні	129 текстів

Також ми проаналізували тексти новин 2021 року і визначили кількість позитивних, негативних і нейтральних текстів. Відповідно до загальної кількості текстів 343 тексти ми відносимо до нейтральних, з яких 42 тексти з показниками 0/0. Кількість позитивних текстів становить 118, а негативних – 168 (див. Таблиця 2.2).

Таблиця 2.2 Тональна класифікація текстів новин 2021 року

Нейтральні	343 текстів
Позитивні	118 текстів
Негативні	168 текстів

Порівнявши результати тональної класифікації текстів 2019-2020 і 2021 років, ми чітко бачимо, що у текстах 2019-2020 років та 2021 року переважають нейтральні тексти. Серед текстів 2019-2020 років кількість негативних текстів, найменша, що ж стосується текстів 2021 року кількість позитивних текстів – найменша. Це пов'язано з динамікою в політичному житті нашої країни.

2.7. Висновки до розділу 2

Отже, у результаті роботи було автоматично укладено тональний словник української мови, створена база даних українськомовних текстів новин.

Цілісна структура баз даних системи виглядає так (див. рис. 2.42):

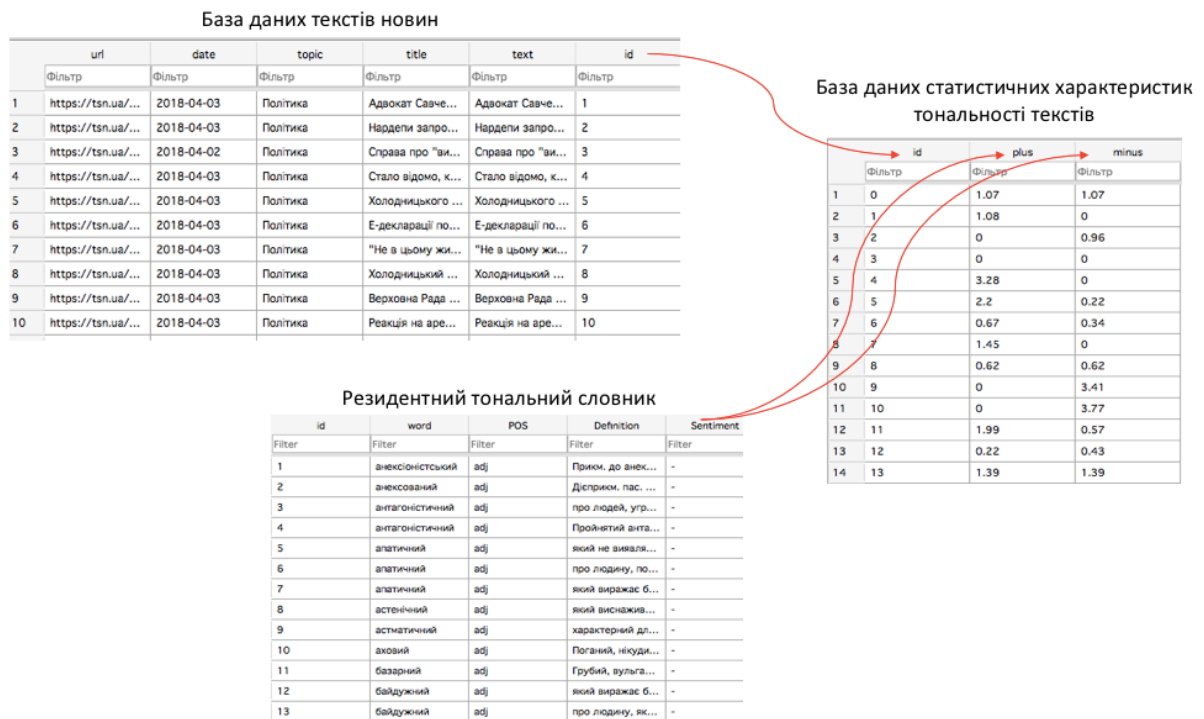


Рис. 2.42 Цілісна структура бази даних системи

Також було створено систему автоматичного визначення тональності українськомовних текстів. Для визначення сентимент аналізу ми працюємо не лише на базі словника, а й поглиблюємо аналіз на рівень словосполучень. Були проаналізовані тексти та визначені метричні тональні характеристики кожного тексту, методом суб'єктивного аналізу текстів встановлено міру коефіцієнта тональності тексту: за умови зіставлення 2-ох коефіцієнтів тональності вони повинні відрізнятись не менше ніж на 3,5. Для оцінки ефективності роботи системи також було проведено порівняльний аналіз роботи системи з оцінкою людиною. Ми можемо стверджувати, що система досить чітко визначає наскільки текст є емотивно-маркованим. Хоча і є такі випадки, коли людина по-іншому визначає емоційне забарвлення тексту, але це пов'язано з певними лінгвістичними, соціально-культурними чинниками.

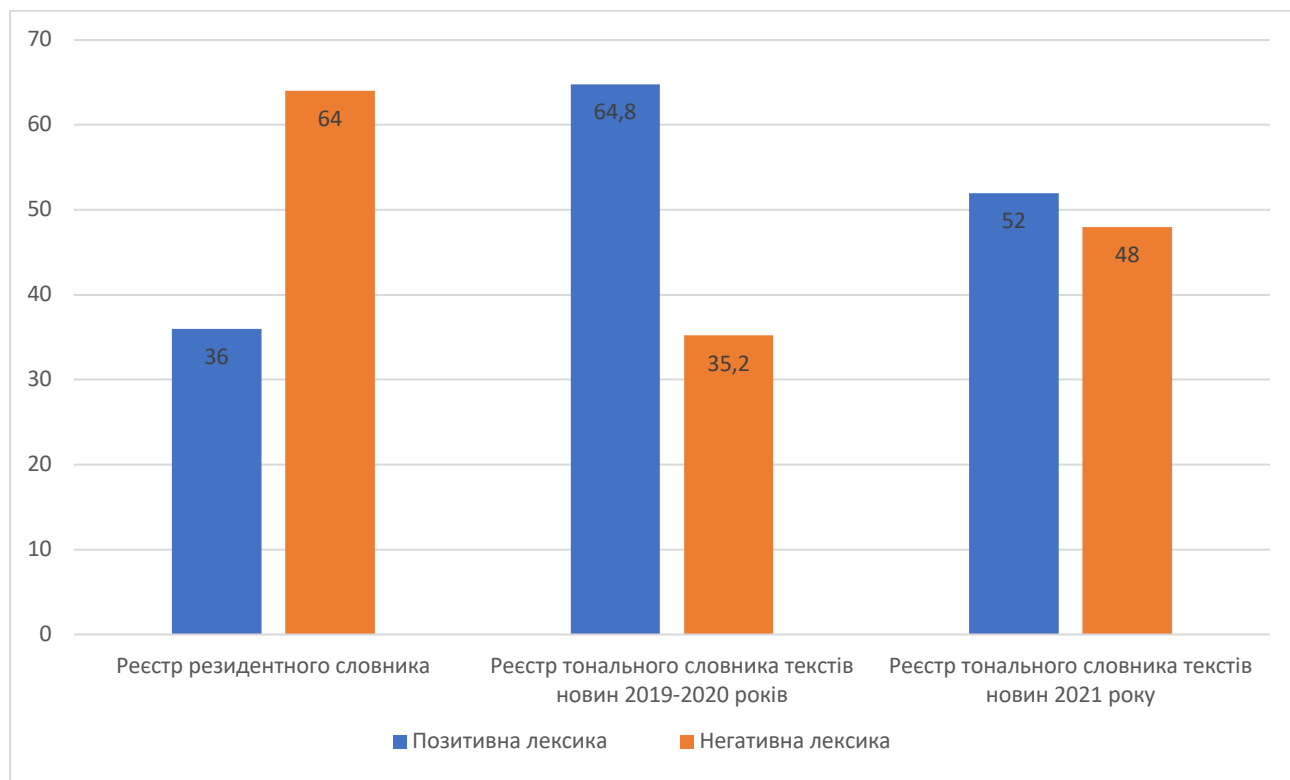
РОЗДІЛ 3

ТОНАЛЬНИЙ СЛОВНИК СУЧАСНОЇ УКРАЇНСЬКОЇ МОВИ: ЛЕКСИЧНА СИСТЕМА ТА ТЕКСТОВА РЕАЛІЗАЦІЯ

3.1. Лінгвостатистична характеристика реєстрів резидентного тонального словника та тонального словника текстів новин

Згенерований тональний словник налічує 10193 лексеми. Серед них негативних лексем 64 % (6576 лексем), позитивних - 36 % (3617 лексем). Проте у текстах новин із цього словника використано набагато менше лексики: за вибіркою 2019 - 2020 років - 1319 лексем; за вибіркою 2021 року - 711 лексем. Емотивні складники позитивної/негативної лексики у тональних словниках (2019-2020 років), згенерованих за досліджуваними текстами, характеризуються таким розподілом: позитивна лексика становить 60,7 % (800 лексем), тоді як негативної лексики - 39,3 % (519 лексем).

Тепер обраховуємо відсоток позитивної/негативної лексики у тональному словнику текстів 2021 року. Відповідно, позитивна лексика становить 52 % (368 лексем), тоді як негативної лексики - 48 % (343 лексем).



Діаграма 3.1

Можемо зробити висновок, що система мови має набагато більше негативних лексем. Це свідчить про вищий відсоток негативності нашого мовлення. Що стосується тональних словників текстів, то більший відсоток навпаки становлять слова на позначення позитивної оцінки у словнику (див. діаграма 3.1). На діаграмі добре видно, що негативні лексеми переважають у реєстрі резидентного словника, але у реєстрах тональних словників тексту новин переважають позитивні лексеми.

Представлені тональні словники характеризуються різним частиномовним (граматичним) розподілом емотивної лексики. До резидентного тонального словника входять прикметники, іменники, дієслова та прислівники (див. табл. 3.1). Прикметників у тональному словнику найбільше (48,2%), що вказує на їхній емоціогенний характер. На другому місці – іменники (31,5%). На третьому місці – прислівники (17,8%). Дієслів у тональному словнику досить мало, оскільки цій частині мови не характерна емотивність.

<i>Частина мови</i>	<i>Кількість</i>	<i>Відсоткове представлення</i>
<i>Прикметник</i>	4911	48,2%
<i>Іменник</i>	3210	31,5%
<i>Дієслово</i>	261	2,5%
<i>Прислівник</i>	1811	17,8%

Таблиця 3.1 Відсотковий розподіл частин мови у резидентному тональному словнику

Також було обраховано відсоткове співвідношення частин мови у розподілі позитивної та негативної лексики (див. табл. 3.2).

	Позитивна лексика	Негативна лексика
Іменник	15,8 %	40,1 %
Прикметник	54,4 %	43,8 %
Дієслово	2,7 %	2,1 %
Прислівник	27,1 %	14 %

Таблиця 3.2 Відсотковий розподіл частин мови у реєстрі резидентного тонального словника

Серед позитивної лексики кількість прикметників значно перевищує кількість інших частин мови(54,4%), кількість прислівників становить 27,1%, а дієслів лише 2,7%. Аналізуючи негативні лексеми словника, можна стверджувати, що кількість прикметників та іменників майже однакова – 43,8% і 40,1% відповідно, дієслів та прислівників значно менше - 2,1% та 14% відповідно.

3.2. Тональна статистична структура тексту

У ході роботи було автоматично укладено частотний словник тональної лексики за текстовими вибірками новин.

Текстова вибірка новин 2019-2020 років (232 010 слововживань)

```

Позитивні: іменник: 463,
            прикметник: 1646,
            дієслово: 179,
            Прислівник: 508
Негативні: іменник:569,
            прикметник: 1160,
            дієслово: 36,
            Прислівник: 152
    
```

Рис. 3.1 Кількісний розподіл частин мови у тональному словнику текстів новин 2019-2020 років

У тональному словнику текстів новин (див. рис. 3.1) на позначення позитивної лексики найбільше зустрічається прикметник (58,9 %). На другому місці – прислівник(18,2 %). На третьому місці – іменник (16,5 %).

На позначення негативної лексики найбільше зустрічається прикметник (60,5 %). На другому місці – іменник(29,7 %). На третьому місці – прислівник(7,9 %). Дієслів у тональному словнику досить мало.

	Позитивна лексика	Негативна лексика
Іменник	16,5 %	29,7 %
Прикметник	58,9 %	60,5 %
Дієслово	6,4 %	1,9 %
Прислівник	18,2 %	7,9 %

Таблиця 3.3 Відсотковий розподіл частин мови у тональному словнику текстів новин 2019-2020 років

Згідно з таблицею 3.3, ми бачимо, що прикметника у текстах найбільше, а дієслова зустрічаються досить рідко. Серед позитивних слів кількість прикметників значно перевищує кількість інших частин мови (58,9%), кількість прислівників становить 18,2%, а дієслів лише 6,4%. Аналізуючи негативні лексеми словника, можна стверджувати, що кількість прикметників перевищує кількість іменників удвічі 60,5% та 29,7% відповідно, дієслів та прислівників значно менше - 1,9% та 7,9% відповідно.

Тепер проаналізуємо розподіл частин мови у текстах новин 2019-2020 років: яка кількість лексем резидентного тонального словника зустрілася у вибірці текстів новин 2019-2020 років(див. рис. 3.2).

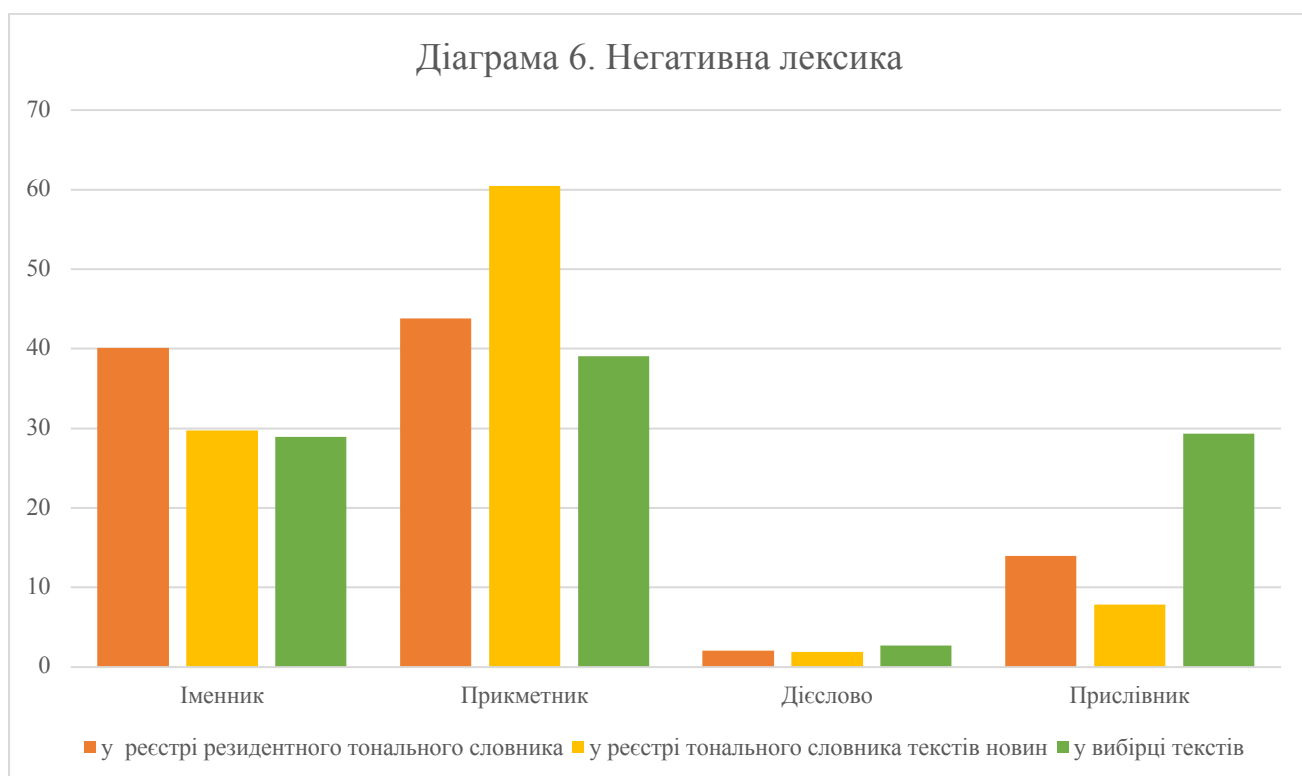
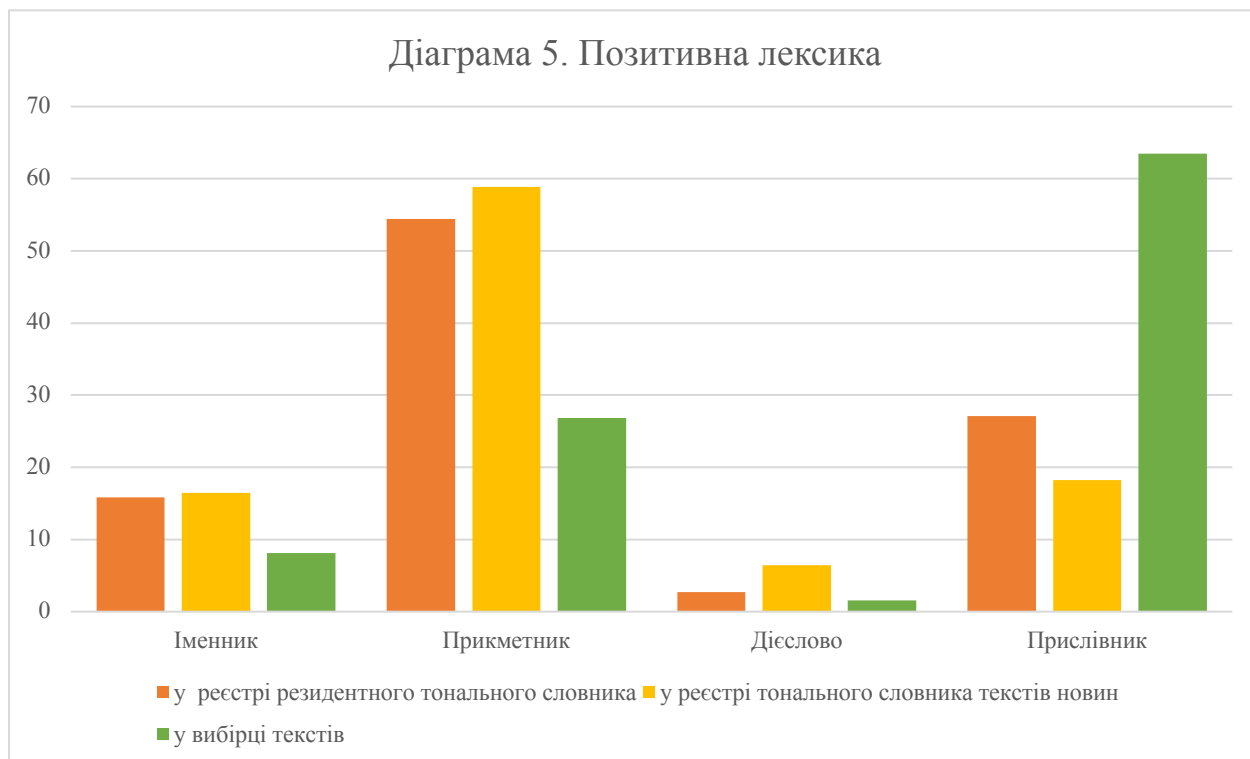
Розподіл частин мови у текстах
 Позитивні: іменник: 66,
 прикметник: 260,
 дієслово: 17,
 Прислівник: 1144
 Негативні: іменник: 200,
 прикметник: 256,
 дієслово: 26,
 Прислівник: 324

Рис.3.2 Кількісний розподіл частин мови у текстах 2019-2020 років

	Позитивна лексика	Негативна лексика
Іменник	8,1 %	28,9 %
Прикметник	26,8 %	39,1 %
Дієслово	1,6 %	2,7 %
Прислівник	63,5 %	29,3 %

Табл. 3.4 Відсотковий розподіл частин мови у текстах новин 2019-2020 років

Проаналізувавши статистичні характеристики частин мови у текстах новин, ми бачимо, що прислівника використано найбільше у текстах, на другому місці – прикметник, а іменник та дієслово займають останні позиції.



Порівнявши відсоткові показники частин мови у реєстрі резидентного тонального словника, тонального словника новин та текстах у загальній вибірці, можна стверджувати, що показники відрізняються не суттєво. Це свідчить про однорідність вживання емотивних слів у тексті.

Текстова вибірка новин 2021 року (181 275 слововживань)

У тональному словнику текстів новин (див. рис. 3.3) на позначення позитивної лексики найбільше зустрічається прикметник (45,5 %). На другому місці – прислівник(31%). На третьому місці – іменник(18,1 %).

На позначення негативної лексики найбільше зустрічається прикметник (46,6 %).

На другому місці – іменник(46,2 %). На третьому місці – прислівник(5 %).

Дієслів у тональному словнику досить мало.

```

Позитивні: іменник: 499,
            прикметник: 1251,
            дієслово: 149,
            Прислівник: 856

Негативні: іменник:953,
            прикметник: 961,
            дієслово: 46,
            Прислівник: 102
    
```

Рис.3.3 Кількісний розподіл частин мови у тональному словнику текстів новин

	Позитивна лексика	Негативна лексика
Іменник	18,1 %	46,2 %
Прикметник	45,5 %	46,6 %
Дієслово	5,4 %	2,2 %
Прислівник	31 %	5 %

Таблиця 3.5 Відсотковий розподіл частин мови у тональному словнику текстів новин

Згідно з таблицею 3.5, ми бачимо, що емотивних прикметників у текстах найбільше, а дієслова зустрічаються досить рідко.

Тепер проаналізуємо розподіл частин мови у текстах новин 2021 року: яка кількість лексем резидентного тонального словника зустрілася у вибірці текстів новин 2021 року (див. рис. 3.4).

```

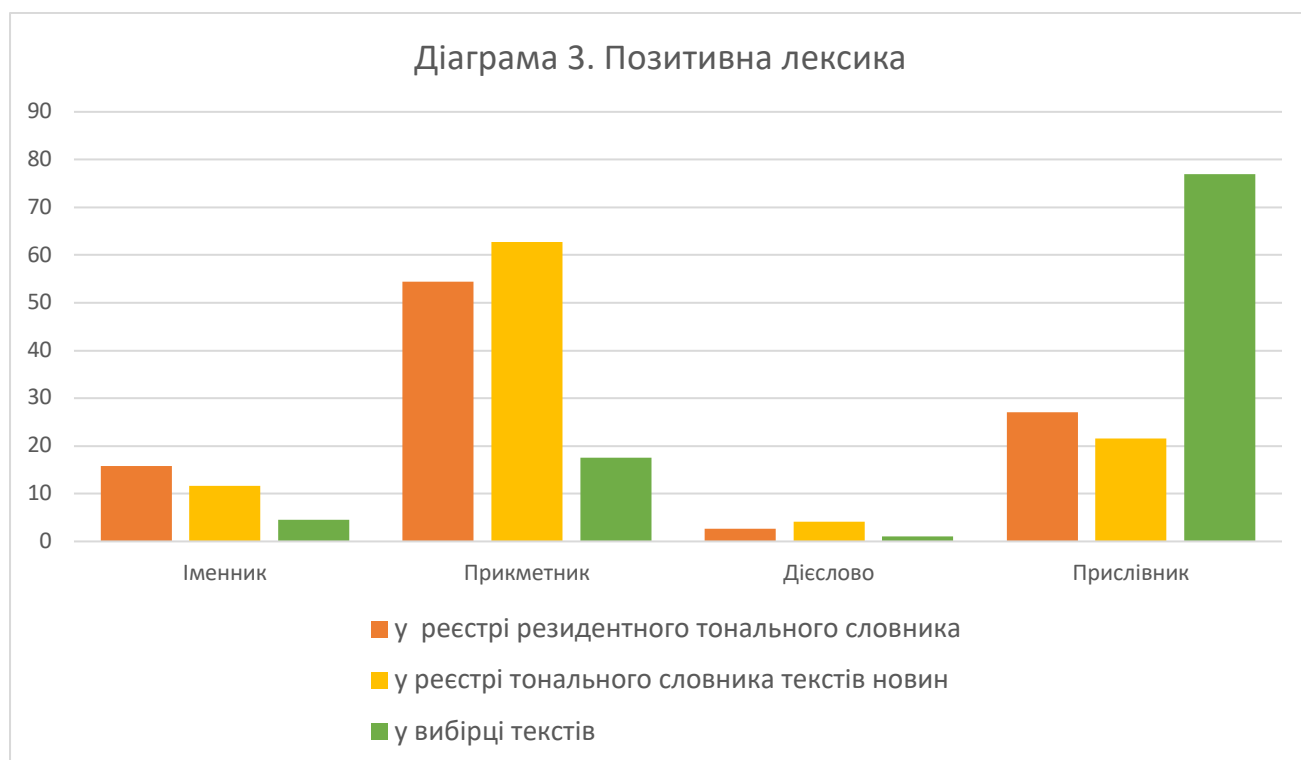
Розподіл частин мови у текстах
Позитивні: іменник: 66,
            прикметник: 260,
            дієслово: 17,
            Прислівник: 1144
Негативні: іменник: 200,
            прикметник: 256,
            дієслово: 26,
            Прислівник: 324
    
```

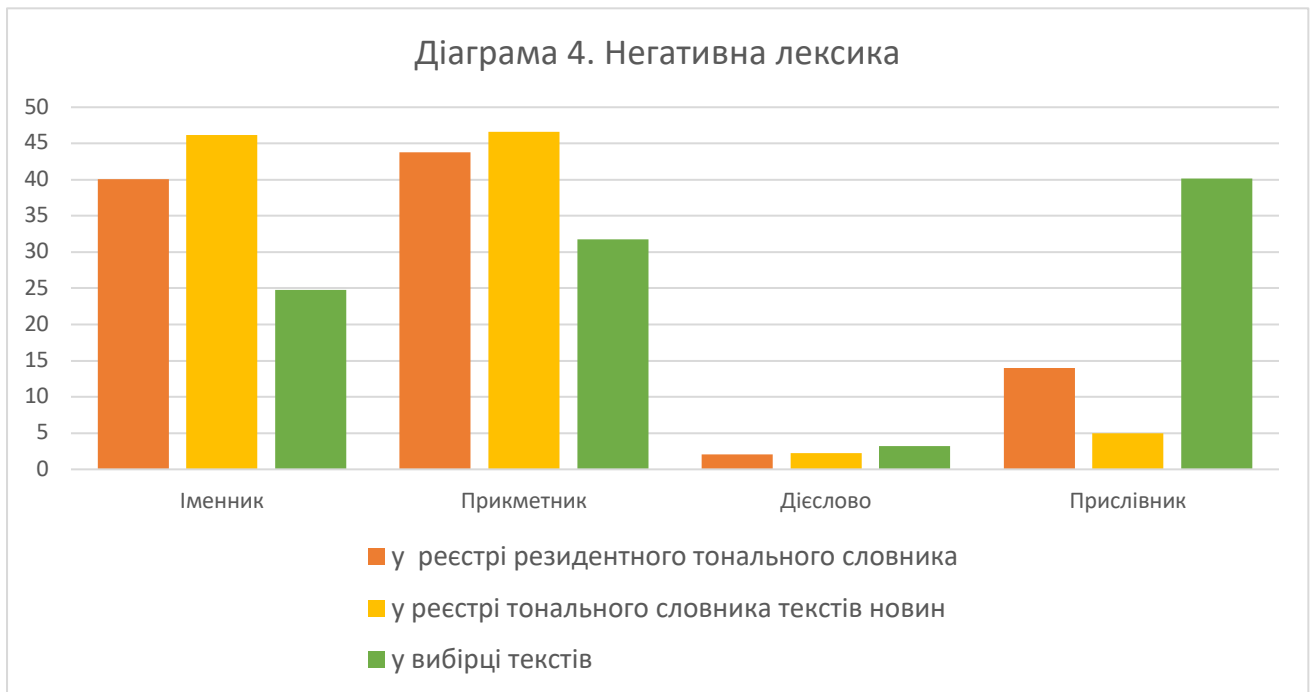
Рис.3.4 Кількісний розподіл частин мови у текстах 2021 року

	Позитивна лексика	Негативна лексика
Іменник	4,5 %	24,8 %
Прикметник	17,5 %	31,8 %
Дієслово	1,1 %	3,2 %
Прислівник	76,9 %	40,2 %

Табл. 3.6 Відсотковий розподіл частин мови у текстах новин 2021 року

Згідно з таблицею 3.6, ми бачимо, що емотивних прислівників резидентного тонального словника покриває тексти новин, а дієслова зустрічаються досить рідко.





Як бачимо, кількість іменників у реєстрі резидентного словника, реєстрі тонального словника текстів новин та текстів у вибірці майже однакові. Та сама ситуація з дієсловами.

Як і у текстах 2019-2020 років, прикметники переважають і в текстах 2021 року. Кількість іменників та дієслів у реєстрі резидентного словника, реєстрі тонального словника текстів новин та текстів у вибірці майже однакові. Можемо зробити висновок, що наші тексти емотивно-марковані.

3.3. Висновки до розділу 3

У розділі 3 було проаналізовано реєстр резидентного тонального словника і виявлено, що система мови має набагато більше негативних лексем. Це свідчить про вищий відсоток негативності нашого мовлення. Що стосується тональних словників текстів, то більший відсоток навпаки становлять слова на позначення позитивної оцінки у словнику.

Також було проаналізовано тексти новин за частиномовними характеристиками тональної лексики. Відповідно, можна зробити такі висновки:

- У текстах найбільше зустрічається прикметників, їхній відсоток становить 50-60%. Тенденція до прикметників вказує на оцінку явища людиною.
- Кількість прислівників у реєстрі резидентного тонального словника та реєстрі тонального словника текстів новин майже однакова, але

прислівника найбільше зустрічається у загальній вибірці текстів, тобто прислівників із резидентного словника зустрілося у текстах більше, ніж інших частин мови. Це говорить про те, що не лише прикметник є емоційною частиною мови, а і прислівник також.

ВИСНОВКИ

У ході роботи було теоретично обґрунтовано та практично реалізовано принципи конструювання системи автоматичного визначення тональності українськомовного тексту.

Автоматична система демонструє алгоритм створення автоматичного сентимент-аналізу. Як свідчать представлені результати, завдання визначення емоційно-оцінної лексики та лексико-емотивної статистичної структури тексту система виконує ефективно.

Для покращення роботи програми було додано:

- Визначення реєстру слів-модифікаторів, які збільшують або зменшують міру тональності емотивної лексики (*дуже, надзвичайно, безмежно, вкрай, досить*);
- Розроблення регулярних правил сполучуваності тональної лексики зі словами-конверторами (заперечними частками *не, ні, ані*, прийменниками *без, крім та ін.*) з метою визначення інвертування позитив-негатив;
- Визначення синтаксичних зв'язків у межах словосполучень;
- Розроблення правил зміни тональності в межах словосполучення, інвертування позитив-негатив за моделями: позитивна оцінка негативного об'єкта, негативна оцінка позитивного об'єкта;
- Словник доповнено новою лексикою.

Після додавання слів-інтенсифікаторів, слів-конверторів та визначення синтаксичних зв'язків, ми отримали точніший відсоток коефіцієнта тональності текстів.

Розроблена програма автоматичного визначення тональності українськомовних текстів дозволила проаналізувати тексти новин ТСН на вміст емотивної лексики. У ході аналізу, було виявлено, що тексти є емотивно-марковані.

Проаналізувавши реєстр резидентного тонального словника, можна стверджувати, система мови має набагато більше негативних лексем. Це свідчить про вищий відсоток негативності нашого мовлення. У реєстрі тонального

словника новин навпаки переважає позитивна лексика. Це пов'язано з динамікою в політичному житті нашої країни.

Проаналізувавши тексти новин за частиномовними характеристиками тональної лексики, можна зробити такі висновки:

Найбільше у текстах зустрічається прикметників, їхній відсоток у текстах становить 50-60%. Тенденція до прикметників вказує на оцінку явища людиною. У підр. “математична лінгвістика” за редакцією В.В. Пасічника: «якщо прикметників багато, текст має високу емоційність»[25].

- Кількість прислівників у реєстрі резидентного тонального словника та реєстрі тонального словника текстів новин майже однакова, але прислівника найбільше зустрічається у загальній вибірці текстів, тобто прислівників із резидентного словника зустрілося у текстах більше, ніж інших частин мови. Це говорить про те, що не лише прикметник є емоціогенною частиною мови, а і прислівник також.

Робота над створенням автоматичної системи визначила завдання, спрямовані на розширення і поглиблення резидентного тонального словника:

- Доповнення словника частками та вигуками;
- Укладання словника тональних словосполучень на матеріалі текстів новин із метою визначення об'єкта тональної оцінки.

Виконання поставлених завдань набагато підвищить ефективність роботи системи емотивного аналізу.

ВИКОРИСТАНІ ДЖЕРЕЛА

1. АГАТ-семантика: семантичне розмічування Корпусу української мови / Н. Дарчук, О. Зубань, М. Лангенбах, Я. Ходаківська // Українське мовознавство. - 2016. - Вип. 1. - С. 92-102. - Режим доступу: http://nbuv.gov.ua/UJRN/Um_2016_1_9
2. Ахманова О.С. Словарь лингвистических терминов / О.С. Ахманова. – 2-е изд., стер. – М : УРСС : Едиториал УРСС, 2004. – 571 с.
3. Бабенко Л.Г. Русская эмотивная лексика как функциональная система: Автореферат дисс... докт. филол. наук. – Свердловск, 1990
4. Бегунов А.Б. Алгоритм Автоматизованого визначення емоційного забарвлення текстових даних
5. Бібліотека scrapy. Режим доступу: https://github.com/12bogdan03/TSN_news_parser
6. ВААЛ [Електронний ресурс]. — Режим доступу :<http://vaal.ru/>
7. Виноградов О. Г. Номенклатура сигнатур емоційного реагування на ситуації публічного виступу / О. Г. Виноградов // Актуальні проблеми соціології, психології, педагогіки. - 2013. - Вип. 18. - С. 119-126. - Режим доступу: http://nbuv.gov.ua/UJRN/apspp_2013_18_17
8. Волек Б. Типология эмотивных знаков // Язык и эмоции: Сб. науч. трудов. – Волгоград: Перемена, 1995
9. Гамзюк М. В. Емотивність фразеологічної системи німецької мови : автореф. дис. на здобуття наук. ступеня доктора філол. наук : спец. 10.02.04 “Германські мови” / М. В. Гамзюк. – К., 2001. – 34 с.
10. Гладкий А. В. Синтаксические структуры естественного языка в автоматизированных системах общения. — М.: Наука, 1985. — 144 с.
11. Добросклонская Т. Г. Вопросы изучения медиатекстов (опыт исследования современной английской медиаречи) / Т.Г. Добросклонская. – М.: Едиториал УРСС, 2005. – 288 с.
12. Добросклонская, Т.Г. Медиалингвистика: системный подход к изучению языка СМИ / Т.Г. Добросклонская. – М.: Едиториал, 2005. – 202с.

13. Дускаева Л.Р. Медиалингвистика в России: становление структуры и векторы развития // Вестник Московского государственного университета. Серия 10. Журналистика. 2018. № 6. С.48-74
14. Канішева О. Розробка словника WordNet. Режим доступу:
http://ena.lp.edu.ua:8080/bitstream/ntb/40817/2/2017_Kanishcheva_O-Rozrobka_slovnykaWordNet_52-56.pdf
15. Колшанский Г. В. Паралингвистика / Г. В. Колшанский. – М. : КомКнига, 2010. – 96 с.
16. Комунікативні одиниці мови : Міжнародна наукова конференція / тези доповідей /. – М. : Освіта, 2006. – 294с.
17. Культура фахового мовлення: Навчальний посібник / За редакцією Н. Д. Бабич. – Чернівці: Книги – XXI, 2006. – 496 с.
18. Леонтьев А. Н. Мотивы и эмоции./ А. Н. Леонтьев – М. : МГУ, 1972. – 32 с.
19. Ляшенко О. І.
Пріоритети розвитку української школи в умовах реформування освіти / О. І. Ляшенко // Збірник наукових праць Кам'янець-Подільського національного університету ім. Івана Огієнка. Серія : Педагогічна. - 2016. - Вип. 22. - С. 39-42. - Режим доступу: http://nbuv.gov.ua/UJRN/znpkr_ped_2016_22_13
20. Маслова В. А. До побудови психолінгвістичної моделі конотації [Текст] / В. А. Маслова // Питання мовознавства. –2009. – № 1. – С. 19-22
21. Меншиков І. Аналіз тональності тексту російською мовою за допомогою графових моделей // УРФУ, Єкатиринбург, Росія: конференція. – 2012.
22. Мягкова Е. Ю. Эмоциональная нагрузка слова и её проявление в материалах ассоциативного эксперимента / Е. Ю. Мягкова // Психолінгвістическіе дослідження в області лексики и фонетики. – Калинин : Калин. гос. ун-т, 1981. – 160 с.
23. Пазельская А. Г., Соловьев А. Н. Метод определения эмоций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии. Сб. научных статей. Вып. 10 (17). – М., 2011. – С. 510–522

24. Пазельская А., Соловьев А. Метод определения эмоций в текстах на русском языке // The international conference on computational linguistics and intellectual technologies «Dialogue 2011» : конференция. – Москва, 2011. – с. 510 – 522
25. Пасічник В. В., Математична лінгвістика. [Книга 1. Квантитативна лінгвістика] : навч. посібник / В. В. Пасічник, Ю. М. Щербина, В. А. Висоцька, Т. В. Шестакевич. — Львів : «Новий світ -2000», 2012. — 359 с.
26. Проекти групи lang_uk. Режим доступу: <https://lang.org.ua/uk/>
27. Рахилина Е.В., Кустова Г.И., Ляшевская О.Н., Резникова Т.И., Шеманаева О.Ю. Задачи и принципы семантической разметки лексики в НКРЯ / Национальный корпус русского языка. Новые результаты и перспективы. – Санкт-Петербург: «НЕСТОР-ИСТОРИЯ» – 2009. С.215 – 239
28. Романюк А., Романишин М., Тональний словник української мови на основі сентимент-анотованого корпусу Міжвідомчий науковий збірник «Українське мовознавство». — К: Київський національний університет імені Тараса Шевченка, 2012. — Випуск 42/1. с. 365-373
29. Сайт новин ТСН. Режим доступу: <https://tsn.ua/politika>
30. Свідер І.А. Основні особливості емотивного тексту / Міжнародна наукова конференція / тези доповідей /. – М. : Освіта, 2012. – 294с., с.275
31. Селіванова О. О. Сучасна лінгвістика : термінологічна енциклопедія / Селіванова О. О. – Полтава: Довкілля – К, 2006. – 716 с.
32. Система SentiWordNet. Режим доступу: <http://sentiwordnet.isti.cnr.it>
33. Стадній А. С. Психолінгвістичний аспект емоційно-оцінної лексики / А. С. Стадній // Вісн. Дніпропетровського ун-ту. – Д. : Вид-во Дніпропетровського нац. ун-ту, 2010. – Вип. 16, No 11, т. 18. – С. 321–325. – (Сер. „Мовознавство”)
34. Телия В. Н. Коннотативный аспект семантики номинативных единиц / В. Н. Телия. – М. : Наука, 1986. – 143 с

35. Тертичний О. О. Особливості використання сучасних рекламних засобів // Зовнішня торгівля: економіка, фінанси, право. - 2014. - № 2. - С. 150-154. - Режим доступу: http://nbuv.gov.ua/UJRN/uazt_2014_2_24
36. Українська мова: Енциклопедія [Електронний ресурс]. – К. : Українська енциклопедія ім. М. П. Бажана, 2000. – 752 с. – Режим доступу: <http://litopys.org.ua/ukrmova/um.htm>
37. Уфимцева А.А. Типы словесных знаков / А.А. Уфимцева . – М., 1974. – С. 153
38. Чибирова Л.Ю., «Эмотивные единицы языка».[Электронный ресурс].
Режим доступу: <http://jarki.ru/wpress/2013/06/12/3257/>
39. Шаховский В. И. Категоризация эмоций в лексико-семантической системе языка. – Воронеж: Изд-во Воронежск. гос. ун-та, 1987. – 192 с.
40. Шаховский В. И. Что такое лингвистика эмоций [электронный ресурс] / В. И. Шаховский/ Волгоградский государственный педагогический университет, г. Волгоград. – Режим доступу: http://tverlingua.by.ru/archive/012/shakhovsky_03_12.htm
41. Шаховский В. И. Эмотивный компонент значения и методы его описания: Учеб. пособие к спецкурсу / В. И. Шаховский. – Волгоград, 1983. – 326 с.
42. Шаховский В. И. Эмоции – мысли в художественной коммуникации / В.И. Шаховский // Языковая личность : социолингвистические и эмотивные аспекты. – Волгоград-Саратов, 1998. – 93 с.
43. V. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd Meeting of the Association for Computational Linguistics, Barcelona, Spain, July 2004
44. Carlo Strapparava, Alessandro Valitutti WordNet-Affect: an Affective Extension of WordNet (англ.) : LREC. — 2004. — Vol. 4.
45. Chetviorkin Iliа, Loukachevitch Natalia. Extraction of Russian Sentiment Lexicon for Product Meta-Domain//COLING: конференция. – 2012.

46. Kasper W. Sentiment Analysis for Hotel Reviews / Walter Kasper // Proceedings of the Computational Linguistics-Applications Conference. – Poland, 10/2011. – P. 45–52.
47. Nasukawa T. & Yi J., Sentiment analysis: capturing favorability using natural language processing // In Proceedings of the 2nd international conference on Knowledge capture, Florida, USA, October 23–25, 2003
48. Prabowo R. and Thelwall M. Sentiment analysis: A combined approach. Journal of Informetrics, 3(2), 2009.
49. Spacy. Режим доступа: <https://explosion.ai/demos/displacy>
50. Stanford parser. Режим доступа: <http://nlp.stanford.edu:8080/parser/index.jsp>
51. Stanza. Режим доступа: <https://stanfordnlp.github.io/stanza/#about>
52. Stefano Vaccianella Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining (англ.) // Proceedings of LREC : конференция. — 2010. — P. 2200–2204
53. Textrazor. Режим доступа: [<https://www.textrazor.com/demo>
54. Werlich, E. A Text Grammar of English / E. A. Werlich. – Heidelberg : Quelle & Meyer, 1976. – 190 p.
55. Yessenov K. Sentiment Analysis of Movie Review Comments / Kuat Yessenov, Sasa Misailovic. – Massachusetts Institute of Technology, Spring 2009. – Режим доступа: <http://people.csail.mit.edu/kuat/courses/6.863/report.pdf>

ДОДАТОК 1

Код програми автоматичного сентимент-аналізу української мови

Textsplitter.py

```
import re
class TextSplitter:

    def split_on_sentence(self, text) -> list:

        t = self.__abbr(text)
        t9 = self.__quest(t)
        split = self.__split_on_3_dots(t9)

        result_sent=list()

        for sent in split:
            result_sent.extend(self.__simple_split_on_sentence(sent))
        result_sent = [i.replace('$', '.') for i in result_sent]
        result_sent = [i.replace('#', '...') for i in result_sent]
        result_sent = [i.replace('&', '?') for i in result_sent]
        result_sent = [i.replace('@', '!') for i in result_sent]
        result_sent = [i.replace('%', '-') for i in result_sent]
        # for i in result_sent:
        #     print(i)
        return result_sent

    def split_on_words(self, text: str) -> list:
        result_sent = self.split_on_sentence(text)
        result=list()
        for i in result_sent:
            i = self.__abbr(i)
            result.extend(self.__split_on_words(i))
            result = [i.replace('$', '.') for i in result]
            result = [i.replace('%', '-') for i in result]
        return result

    def __split_on_words(self, text) -> list:
        text = u''.join([ch for ch in text if ch not in u'—,()/?!.<.;:»'''])
        text = text.split(' ')
        words = [word.strip('.') for word in text]
        return words

    def __abbr(self, text):
        p1 = re.compile('\([а-яюєії]+\.[A-ZА-ЯЇЮЄІ][^\)]+\)')
        m1 = re.findall(p1, text)
        for m in m1:
            text = text.replace(m, m.replace('.', '$'))

        p4 = re.compile('[\"\\«][А-ЯЮЄІІа-яюєіїА-Zа-z ]+\.[А-ЯЮЄІІа-яюєіїА-Zа-z
0-9]+[\"\\»]')
        m4 = re.findall(p4, text)
        for m in m4:
            text = text.replace(m, m.replace('.', '$'))

        p5 = re.compile('[а-яюєіїА-ЯЮЄІІ]+\.[а-яюєії]+[\.]*')
        m5 = re.findall(p5, text)
        for m in m5:
            text = text.replace(m, m.replace('.', '$'))

        p7 = re.compile(' [а-яюєіїА-ЯЮЄІІ]\.[А-ЯЮЄІІ][а-яюєії]*')
        m7 = re.findall(p7, text)
        for m in m7:
```

```

        text = text.replace(m, m.replace('.', '$'))

p10 = re.compile('[А-ЯЮЄїІа-яюєії]+-[А-ЯЮЄїІа-яюєії]+')
m10 = re.findall(p10, text)
for m in m10:
    text = text.replace(m, m.replace('-', '%'))
p14 = re.compile('\«[А-ЯЮЄїІа-яюєії , -?!]+\»')
m14 = re.findall(p14, text)
for m in m14:
    text = text.replace(m, m.replace('.', '$'))
return text

def __quest(self, text: str) -> str:
p910 = re.compile('[\(\\"«][А-ЯЮЄїІа-яюєїіі ]+!\[\]\\"»]')
m910 = re.findall(p910, text)
for m in m910:
    text = text.replace(m, m.replace('!', '@'))

p12 = re.compile('\- [А-ЯЮЄїІа-яюєїі\, ]+? \-')
m12 = re.findall(p12, text)
for m in m12:
    text = text.replace('?', '&')

p13 = re.compile('\"[А-ЯЮЄїІа-яюєїі ]+?\.\\.\\.\'')
m13 = re.findall(p13, text)
for m in m13:
    text = text.replace('?.?.', '&')
return text

def __split_on_3_dots(self, text:str)->list:
text=text.replace('...', '#')
lines=text.split('#')
lines = [sent.strip(' ') for sent in lines]
lines=[sent + '#' for i, sent in enumerate(lines)]
lines[-1] = lines[-1][: -1]
return lines

def __simple_split_on_sentence(self, text:str)->list:
lines = text.split('.')
lines = [sent.strip(' ') for sent in lines]
lines=[sent + '.' for i, sent in enumerate(lines)]
lines[-1] = lines[-1][: -1]
return lines

import sqlite3
import os

# основна база з розміченими словами "+", "-"
class LineSplitter:

    def split_line(self, line: str, sentiment_value: str, pos_value: str) ->
list:
        row = line.split('\t')
        word = row[1]
        if len(row) <= 2:
            row.append('')
        # tail = row[2]
        pos = pos_value
        tail = row[2]
        sentiment = sentiment_value
        # print(word, tail, pos)
        return word, pos, tail, sentiment

```

```

class TextDataReader:

    def __init__(self):
        self.line_splitter = LineSplitter()

    def get_text_data(self, filenames: list, sentiment_values: list, pos_values:
list) -> list:
        res = list()
        for i, filename in enumerate(filenames):
            with open(filename, 'r', errors='ignore') as f:
                text = f.read()
                lines = text.split('\n')
                # print(len(lines),lines)
                for line in lines:
                    row = self.line_splitter.split_line(line,
sentiment_values[i],pos_values[i])
                    print(row)
                    res.append(row)

        return res

```

common_db_creator.py

```

class CommonDBCreator:

    def __init__(self):
        self.text_data_reader = TextDataReader()

    def create_db(self, db_source: str, filenames: list, sentiment_values:
list,pos_values: list):
        text_data = self.text_data_reader.get_text_data(filenames,
sentiment_values, pos_values)
        self.__init_db(db_source)
        self.__fill_db(text_data)

    def __init_db(self, db_source: str):
        if db_source in os.listdir():
            os.remove(db_source)
        self.con = sqlite3.connect(db_source)
        self.c = self.con.cursor()
        self.c.execute("CREATE TABLE `users` ( `id`INTEGER PRIMARY KEY ,"
""word`TEXT,`POS`TEXT,`Definition`TEXT,`Sentiment`TEXT);")
        self.con.commit()

    def __fill_db(self, text_data: list):
        for row in text_data:
            print(row)
            if not row:
                continue
            self.__insert_row(row[0], row[1], row[2], row[3])

    def __insert_row(self, word, pos, definition, sentiment):
        self.c.execute("INSERT INTO users (word,pos, Definition,Sentiment)
VALUES (?,?,?,?)",
                    (word, pos, definition, sentiment))
        self.con.commit()

if __name__ == '__main__':
    common_db_creator = CommonDBCreator()
    common_db_creator.create_db('common.db',
                                ['emotion_words/minus_adj.txt',

```

```

        'emotion_words/plus_adj.txt',
        'emotion_words/minus_noun.txt',
        'emotion_words/plus_noun.txt',
        'emotion_words/noun.txt',
        'emotion_words/minus_verb.txt',
        'emotion_words/verb.txt',
        'emotion_words/minus_adv.txt',
        'emotion_words/plus_adv.txt'],
        ['-','+', '-','+', ',', '-','+', '-','+'],
        ['adj', 'adj', 'noun', 'noun', 'noun', 'verb',
'verb', 'adverb', 'adverb'])
    print('Done!')

```

emotion_dict.py

```

import sqlite3
import os

class Emotion_dict:

    def get_data(self, filename) -> list:
        res = list()
        with open(filename, 'r', errors='ignore') as f:
            text = f.read()
            lines = text.split('\n')
            # print(len(lines), lines)
            for line in lines:
                row=self.split_line(line)
                print(row)
                res.append(row)

        return res

    def split_line(self, line) -> list:
        row = line.split('\t')
        word=row[1]
        if len(row)<=2:
            row.append('')
        # tail = row[2]
        pos = 'adj'
        tail=row[2]
        sentiment='+'
        #print(word, tail, pos)
        return word, pos, tail, sentiment

    def createDB(self, db_source):
        if db_source in os.listdir():
            os.remove(db_source)
        self.con = sqlite3.connect(db_source)
        self.c = self.con.cursor()
        self.c.execute("CREATE TABLE `users` ( `id` INTEGER PRIMARY KEY , "
""`word` TEXT, `POS` TEXT, `Definition` TEXT, `Sentiment` TEXT, `Emotion_value` TEXT);")
        self.con.commit()

    def insert_Row(self, word, pos, definition, sentiment):
        self.c.execute("INSERT INTO users (word, pos, Definition, Sentiment)
VALUES (?, ?, ?, ?)", (word, pos, definition, sentiment))
        self.con.commit()

```

```

if __name__=='__main__':
    Emot=Emotion_dict()
    Db_source="plus.db"
    s = Emot.get_data("plus_adj.txt")

    Emot.createDB(Db_source)

    for row in s:
        print(row)
        if not row:
            continue
        Emot.insert_Row(row[0], row[1],row[2],row[3])

```

emotion_analyzer.py

```

import sqlite3
from lemmatization import Lemmatizer
from get_news import NewsAnalyzer
import stanza

# NEWS_FILENAME = 'political_news.csv'
NEWS_FILENAME = 'news_2021.csv'
DB_SOURCE = 'common.db'

class SentimentToken:
    def __init__(self, stanza_word):
        self.word_id = stanza_word.id
        self.lemma = stanza_word.lemma
        self.head_id = stanza_word.head
        # self.head = stanza_word[stanza_word.head - 1].lemma
        self.deprel = stanza_word.deprel

class EmotionAnalyzer():

    def __init__(self, db_source):
        # self.lemmatizer = Lemmatizer()
        self.news_analyzer = NewsAnalyzer()
        self.con = sqlite3.connect(db_source)
        self.c = self.con.cursor()
        self.dict_of_emotions = self.__read_data_from_db()
        self.dict_of_pos = self.__read_pos_from_db()

    def __read_data_from_db(self) -> dict:
        # self.c.execute("select * from users")
        # data = self.c.fetchall()
        # print(data)
        dict_of_emotions=dict()
        self.c.execute("select word,Sentiment from users")
        emotions = self.c.fetchall()
        # print(emotions)
        for emotion in emotions:
            word=emotion[0]
            sentiment=emotion[1]
            dict_of_emotions[word] = sentiment
        # print(dict_of_emotions)
        return dict_of_emotions

    def __read_pos_from_db(self)-> dict:
        dict_of_pos = dict()
        self.c.execute("select word,POS from users")
        pos = self.c.fetchall()

```

```

# print(emotions)
for p in pos:
    word = p[0]
    pos = p[1]
    dict_of_pos[word] = pos
return dict_of_pos

def syntax_relations(self):
    emot_converters = ['не', 'ні', 'ані', 'без', 'крім', 'окрім']
    intensifiers = ['дуже', 'надзвичайно', 'безмежно', 'вкрай', 'досить',
'зовсім', 'так', 'над', 'аж', 'таки']

    nlp = stanza.Pipeline('uk', processors='tokenize,
mwt,pos,lemma,depparse') # initialize ukr neural pipeline
    list_of_texts = self.news_analyzer.get_text(NEWS_FILENAME)
    list_of_percentage = list()
    docs = list()
    for text in list_of_texts:
        doc = nlp(text)
        docs.append(doc)
    # print(docs)
    sentiment_count = []
    plus_tokens = []
    minus_tokens = []
    for i,doc in enumerate(docs):
        doc_plus_emotions, doc_minus_emotions = 0,0
        for sentence in doc.sentences:
            token_sentence = []
            for j,word in enumerate(sentence.words):
                # print(word.__dir__())
                token = SentimentToken(word)
                next_token = SentimentToken(sentence.words[j+1]) if j <
len(sentence.words)-1 else None
                prev_token = SentimentToken(sentence.words[j - 1]) if j <
len(sentence.words) - 1 else None
                token.sentiment_label =
self.dict_of_emotions.get(token.lemma,None)

                if next_token:
                    next_token.sentiment_label =
self.dict_of_emotions.get(token.lemma, None)

                if prev_token:
                    prev_token.sentiment_label =
self.dict_of_emotions.get(token.lemma, None)

                token_sentence.append(token)

            #change of sentiment by emot_converters
            if token.lemma in emot_converters and next_token and
next_token.sentiment_label == '+':
                token.sentiment_label = '-'
            elif token.lemma in emot_converters and next_token and
next_token.sentiment_label == '-':
                token.sentiment_label = '+'
            #change of sentiment by intensifiers
            if token.lemma in intensifiers and next_token and
next_token.sentiment_label == '+':
                token.sentiment_label = '+'
            elif token.lemma in intensifiers and next_token and
next_token.sentiment_label == '-':
                token.sentiment_label = '-'

```

```

        ##dependency relations
        for token in token_sentence:
            head = next((t for t in token_sentence if t.word_id ==
token.head_id), None)
            #nmod_relations
            if token.deprel == 'nmod' and head and head.sentiment_label
is not None and head.sentiment_label != token.sentiment_label:
                head.sentiment_label = token.sentiment_label

            #amod_relations
            if token.deprel == 'amod' and head and head.sentiment_label
is not None and head.sentiment_label != token.sentiment_label:
                head.sentiment_label = token.sentiment_label

            # obj_relations
            if token.deprel == 'obj' and head and head.sentiment_label
is not None and head.sentiment_label != token.sentiment_label:
                head.sentiment_label = token.sentiment_label
            if token.deprel == 'iobj' and head and head.sentiment_label
is not None and head.sentiment_label != token.sentiment_label:
                head.sentiment_label = token.sentiment_label

            #advmod relations
            if token.deprel == 'advmod' and head and
head.sentiment_label is not None and head.sentiment_label !=
token.sentiment_label:
                head.sentiment_label = token.sentiment_label

        print([(t.word_id,t.lemma,t.head_id, t.deprel,
t.sentiment_label) for t in token_sentence])
        sentiment = [(t.sentiment_label) for t in token_sentence]
        doc_plus_emotions += sentiment.count('+')
        doc_minus_emotions += sentiment.count('-')

        # count_of plus/minus words
        for token in token_sentence:
            if token.sentiment_label == '+':
                # print(token.lemma)
                plus_tokens.append(token.lemma)
            if token.sentiment_label == '-':
                minus_tokens.append(token.lemma)

        #
        # print("Positive: ", plus_percentage)
        # print("Negative: ", minus_percentage)

        #statistics_count
        sentiment_count.append([doc_plus_emotions, doc_minus_emotions,
doc.num_words])
        # print(sentiment_count)
        all_docs_words_count =0
        for i,sentiment in enumerate(sentiment_count):
            # print(sentiment)
            plus_percentage = round((sentiment[0] / sentiment[2] * 100), 2)
            minus_percentage = round((sentiment[1] / sentiment[2] * 100), 2)
            list_of_percentage.append([i, [plus_percentage, minus_percentage]])

        all_docs_words_count += sentiment[2]

        # довжина ГС
        print('ГС:', all_docs_words_count)
        # print(list_of_percentage)

```

```

    all_emotions_count = [sum(row[i] for row in sentiment_count) for i in
range(len(sentiment_count[0]))]
    plus_emot_count = all_emotions_count[0]
    minus_emot_count = all_emotions_count[1]
    print('Усього: Позитивні: {p}, Негативні: {n}'.format(p=plus_emot_count,
n=minus_emot_count))

    # сортує за відсотковими показниками позитивну і негативну лексику

    list_of_percentage.sort(key=lambda x: x[1][0], reverse=True)
    print('\nПозитив: ')
    for i in list_of_percentage:
        print(i)
    list_of_percentage.sort(key=lambda x: x[1][1], reverse=True)
    # print(list_of_percentage)
    print('\nНегатив: ')
    for i in list_of_percentage:
        print(i)

    # # середня довжина вибірки
    # lenlemma = round(all_docs_words_count / 1801) #поменять к-ство текстов
    # print('Середня довжина тексту:', lenlemma)
    unique_plus_tokens = set(plus_tokens)
    print("\nplus_tokens", len(unique_plus_tokens))
    unique_minus_tokens = set(minus_tokens)
    print("minus_tokens", len(unique_minus_tokens))

    print("\nК-сть позитивних слів у текстах: ", len(plus_tokens))
    print("\nКількість негативних слів у текстах: ", len(minus_tokens))

    # визначає з цих слів, загальну к-сть кожної наявної частини мови
    plus_pos = [self.dict_of_pos.get(x, None) for x in plus_tokens]

    plus_adj = plus_pos.count('adj')
    plus_noun = plus_pos.count('noun')
    plus_verb = plus_pos.count('verb')
    plus_adverb = plus_pos.count('adverb')
    print('\nПозитивні:\tіменник: {n},\n\t\tприкметник: {a},\n\t\tдієслово:
{v},\n\t\tПрислівник: {adv}'.format(
        n=plus_noun, a=plus_adj, v=plus_verb, adv=plus_adverb))
    minus_pos = [self.dict_of_pos.get(x, None) for x in minus_tokens]

    minus_adj = minus_pos.count('adj')
    minus_noun = minus_pos.count('noun')
    minus_verb = minus_pos.count('verb')
    minus_adverb = minus_pos.count('adverb')
    print('\nНегативні:\tіменник:{n},\n\t\tприкметник: {a},\n\t\tдієслово:
{v},\n\t\tПрислівник: {adv} '.format(
        n=minus_noun, a=minus_adj,
        v=minus_verb, adv=minus_adverb))

if __name__ == '__main__':
    emotion_analyzer = EmotionAnalyzer(DB_SOURCE)
    # emotion_analyzer.emot_analyze()
    emotion_analyzer.syntax_relations()

```

main_analyzer.py

```

from tkinter import *
import tkinter
from text_emotion_analyzer import TextEmotionAnalyzer

DB_SOURCE='common.db'

```

```

FILENAME = 'input.txt'

class MainSentimentAnalyzer:

    def __init__(self, FILENAME, DB_SOURCE):
        self.emotion_analyzer=TextEmotionAnalyzer(DB_SOURCE)
        self.filename = 'temporal.txt'

    def run_word_analyze(self):
        emotions = self.get_value()
        plus_ts = emotions[0]
        minus_ts = emotions[1]
        len_text = emotions[2]
        plus_emotions = len(plus_ts)
        minus_emotions = len(minus_ts)
        plus_percentage = round((plus_emotions / len_text * 100), 2)#відсоток
        minus_percentage = round(minus_emotions * 100 / len_text, 2)

        str_res_plus = 'Довжина тексту: {0}\n\nПозитивні
емоції:\t{1}\n{2}\nКоефіцієнт тональності: {3}%\n\n\n '.format(len_text,
plus_emotions,
plus_ts,plus_percentage)
        str_res_minus = 'Негативні емоції:\t{0}\n{1}\nКоефіцієнт тональності:
{2}%'.format(minus_emotions,
minus_ts,minus_percentage)

        self.result_label_plus.config(text = str_res_plus, fg='blue', font
='Helvetica')
        self.result_label_minus.config(text=str_res_minus, fg='red', font
='Helvetica')

    def run(self):
        self.main_root = Tk()
        self.main_root.title('Emotion Analyzer')
        self.text = Text(self.main_root, bd=5)
        self.result_label_plus = Label(self.main_root, text='', bd=5)
        self.result_label_minus = Label(self.main_root, text='', bd=5)

        self.text.bind('<Return>', self.write_in_file)
        self.text.pack(side=RIGHT)
        self.text.focus_set()
        self.button = Button(self.main_root, text="Завантажити текст", width=20,
command=self.write_in_file)
        self.button.bind('<Return>', self.write_in_file)
        self.button.pack()
        self.result_label_plus.pack()
        self.result_label_minus.pack()

        self.main_root.mainloop()

    def write_in_file(self):
        text = self.text.get('1.0', END + '-1c')
        with open(self.filename, 'w') as f:
            f.write(text)
        self.run_word_analyze()
        return text

    def get_value(self) -> list:
        with open(self.filename, 'r') as f:
            text = f.read()

```

```
        result = self.emotion_analyzer.emot_analyze(text)

    return result

if __name__ == '__main__':
    main_form = MainSentimentAnalyzer(FILENAME, DB_SOURCE)
    main_form.run()
```

ДОДАТОК 2

Опитування «Визначення емотивності текстів новин»

Опитування було проведено на платформі google_forms.

Посилання на опитування: <https://forms.gle/LQkkbrX1VgKS7Qt7>

1. Ваш вік: *

- 15-19
- 20-24
- 25-29
- 30-40
- 41-50
- 51-59
- 60+

2. Ваша стать: *

- Чоловіча
- Жіноча

3. Ваша освіта: *

- неповна середня
- середня
- незакінчена вища
- вища

4. Ваш рід діяльності/професія/спеціальність/сфера: *

У формі нижче подані речення з текстів новин.

Просимо Вас визначити, яке емоційне забарвлення, на вашу думку, мають ці речення: позитивне, негативне чи нейтральне.

Росія вже виробила імунітет до санкцій, тому їх треба посилювати.

- позитивне
- негативне
- нейтральне

Підтримка міжнародних партнерів говорить про те, що виправдати окупацію Криму ніколи не вдасться.

- позитивне
- негативне
- нейтральне

По її щоках текли сльози радості.

- позитивне
- негативне
- нейтральне

Після успішних випробувань у травні заплановано масове виробництво вакцини.

- позитивне

- негативне
- нейтральне

Оскільки пацієнт себе погано почував, лікар виписав йому цей препарат.

- позитивне
- негативне
- нейтральне

У сніговому полоні опинилася і Німеччина, де за день сталися сотні аварій на дорогах.

- позитивне
- негативне
- нейтральне

Унаслідок ворожого обстрілу біля Авдіївки загинув український борець.

- позитивне
- негативне
- нейтральне

У США молода дівчина померла після невдалої ін'єкції.

- позитивне
- негативне
- нейтральне

У країні спостерігається підвищення колективного імунітету до сезонних захворювань.

- позитивне
- негативне
- нейтральне

Подальша підтримка влади України направлена на забезпечення кращого доступу до правосуддя.

- позитивне
- негативне
- нейтральне

Денис Шмигаль доручив міністрам розробити алгоритм відшкодування втрат людям, які були змушені припинити свою діяльність через запровадження червоного рівня епідемічної небезпеки.

- позитивне
- негативне
- нейтральне

Уряд запровадив систему підтримки бізнесу, яка дозволить максимально зняти фінансове навантаження на підприємців у період карантинних обмежень.

- позитивне
- негативне
- нейтральне

Правда, яку ми почуємо на конференції, може бути болючою, але вона визволяє.

- позитивне
- негативне
- нейтральне

Люди були засмучені цими звітками, але гірка правда краще за солодку брехню.

- позитивне
- негативне
- нейтральне