



## КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

на тему Метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту

Галузь знань 12 – Інформаційні технології  
Шифр і назва галузі знань  
Спеціальність 122 – Комп'ютерні науки  
Шифр і назва спеціальності  
Освітня програма Комп'ютерні науки  
Назва освітньої програми

Виконав: студент 4 курсу, група КН-18-1  
Курс, група виконавця  
 Підпис  
С.С. Клекач  
Ініціали, прізвище


Керівник: викладач кафедри КН  
Науковий ступінь, посада  
 Підпис  
О.В. Собко  
Ініціали, прізвище

Нормоконтроль: к.т.н., доцент кафедри КН  
Науковий ступінь, посада  
 Підпис  
Р.О. Багрій  
Ініціали, прізвище

До захисту допускаю:

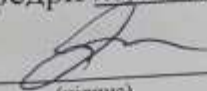
Зав. кафедри КН, д.т.н., професор

13 червня 2022 р.

 Підпис  
О.В. Бармак  
Ініціали, прізвище

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
Факультет інформаційних технологій  
Кафедра комп'ютерних наук  
Освітній ступінь бакалавр  
Галузь знань 12 – Інформаційні технології  
Спеціальність 122 – Комп'ютерні науки


ЗАТВЕРДЖУЮ  
Завідувач кафедри комп'ютерних наук

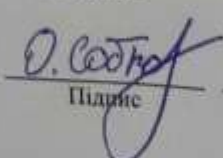
  
(підпис)  
д.т.н., професор О.В. Бармак  
«25» березня 2022 року

### ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА

1. Тема кваліфікаційної роботи бакалавра: «Метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту»
2. Завдання видано студенту Клекачу Степану Сергійовичу  
(прізвище, ім'я, по батькові)
3. Керівник роботи викладач кафедри КН Собко Олена Віталіївна  
(посада, прізвище, ім'я, по батькові)
4. Затверджено наказом університету від «01» березня 2022р. № 18
5. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Мета роботи – розробка методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту, а також розробка відповідної методу інформаційної системи автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту. При формуванні висновків стосовно рівню впевненості автора слід враховувати частоту вживання значущих слів, а також їх оцінку. Також у даній роботі слід сформуванати підхід до визначення оцінки рівню емоційної напруженості тексту.

Виконавець: студент 4 курсу, група КН-18-1  
Курс, група виконавця  С.С. Клекач  
Ініціали, прізвище

Керівник: викладач кафедри КН  
Науковий ступінь, посада  О.В. Собко  
Ініціали, прізвище

## Анотація

Тема кваліфікаційної роботи бакалавра: «Метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту»

Виконавець кваліфікаційної роботи бакалавра: студент групи КН-18-1 Клекач Степан Сергійович

Керівник кваліфікаційної роботи бакалавра: викладач кафедри КН Собко Олена Віталіївна

Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
69	33	8	50	3

Метою кваліфікаційної роботи бакалавра є розробка інформаційної системи автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту. Для розробки запропонованої інформаційної системи було використано мову програмування C#, СКБД MS SQL Server, платформу .NET.

Розробка методів та засобів для автоматизованого визначення психоемоційних характеристик людини як метаданих до створеного нею текстового контенту дозволяє сприяти вирішенню цілої низки безпекових та критичних питань, як на рівні локальних задач, так і на державному рівні.

Розглянуте як приклад автоматизоване визначення рівню впевненості автора за семантичним аналізом текстового контенту є одним із випадків витягу психоемоційних складових автора за наявним цифровим текстом, що дозволяє робити висновки щодо актуальності та достовірності наявних у тексті відомостей.

Ключові слова: семантичний аналіз, частота, текст, рівень емоційної напруженості тексту, рівень впевненості автора.

Виконавець: студент 4 курсу, група КН-18-1  
Курс, група виконавця

  
Підпис

С.С. Клекач  
Ініціали, прізвище

## Зміст

Перелік скорочень .....	3
Вступ.....	4
Розділ 1 Характеристика предметної області: аналіз моделей, методів та реалізацій.....	5
1.1 Аналіз інформаційних моделей.....	5
1.2 Огляд теоретичних підходів до розв’язку подібних задач .....	12
1.3 Аналіз існуючих програмних рішень.....	16
1.4 Аналіз сучасних засобів створення програмного забезпечення .....	20
1.5 Мета, задачі та вимоги до реалізації інформаційної системи .....	23
Розділ 2 Проектування інформаційної системи .....	26
2.1 Метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту .....	26
2.2 Інформаційна структура системи .....	29
2.2.1 Проектна архітектура системи та взаємозв’язок компонентів.....	29
2.2.2 Інформаційна модель.....	31
2.3 Вибір засобів розробки інформаційної системи .....	35
2.3.1. Вибір мови програмування.....	36
2.3.2. Вибір редактора програмного коду .....	36
2.3.3 Вибір СКБД .....	37
Розділ 3 Програмна реалізація інформаційної системи .....	39
3.1 Структура та функціональне призначення програмних складових системи.....	39
3.2 Особливості реалізації програмних складових системи.....	40
3.3 Тестування інформаційної системи .....	46
3.4 Інструкція користувача.....	52
Висновки .....	64
Перелік посилань.....	66
Додатки	

### Перелік скорочень

Скорочення, термін, позначення	Пояснення
БД	База даних
ІС	Інформаційна система
ІТ	Інформаційні технології
КРБ	Кваліфікаційна робота бакалавра
КН	Комп'ютерні науки
ПЗ	Пояснювальна записка
ПП	Програмний продукт
СКБД	Система керування базами даних
ХНУ	Хмельницький національний університет.
CLR	Common Language Runtime
BCL	Base Class Library
DE	Disperce Evaluation
FCL	Framework Class Library
LSA	Latent Semantic Analysis
MS	Microsoft
MFC	Microsoft Foundation Class
TF	Term Frequency
TF-IDF	Term Frequency – Inverse Document Frequency

## Вступ

Розробка методів та засобів для автоматизованого визначення психоемоційних характеристик людини як метаданих до створеного нею текстового контенту дозволяє сприяти вирішенню цілої низки безпекових та критичних питань, як на рівні локальних задач, так і на державному рівні.

Ця задача є актуальна, адже з розвитком мережі Інтернет та різного роду соціальних мереж люди все частіше висловлюють свої думки в текстовій формі. Людям простіше висловлювати своє ставлення до якихось подій не в реальному житті, а у віртуальному, особливо коли є можливість приховати особистість, зробити це анонімно. З метою забезпечення безпеки на державному рівні, такого роду повідомлення можуть перехоплюватися з метою запобігання різного роду злочинів, як проти держави так і проти людяності.

Розглянуте як приклад автоматизоване визначення рівню впевненості автора за семантичним аналізом текстового контенту є одним із випадків витягу психоемоційних складових автора за наявним цифровим текстом, що дозволяє робити висновок щодо актуальності та достовірності наявних у тексті відомостей.

Дане рішення може бути використане для проведення різного роду психологічного аналізу особистості або визначати на скільки вхідний текст є правдивим та може викликати високий рівень довіри до написаного.

**Мета кваліфікаційної роботи бакалавра** – створення й програмна реалізація методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту.

**Об'єкт дослідження** – процес семантичного аналізу цифрового тексту.

**Предмет дослідження** – інформаційні технології, моделі, методи та засоби для автоматизованого визначення рівню впевненості автора у результаті семантичного аналізу цифрового тексту.

## **Розділ 1 Характеристика предметної області: аналіз моделей, методів та реалізацій**

### **1.1 Аналіз інформаційних моделей**

Хоча точні відомості про точне походження суспільства невідомі, все ж є загальноприйнятим фактом, що людина живе в суспільстві з дуже давніх часів. Людині необхідно жити в суспільстві для свого існування і добробуту. Майже в усіх аспектах свого життя вона відчуває потребу суспільства. Біологічно і психологічно він змушений жити в суспільстві [1].

Повна ізолюваність життя для людини нестерпна, і вона не може розвиватися нормально в ізоляції. Суть факту полягає в тому, що людина завжди належала до якогось суспільства, без якого вона взагалі не може існувати. Суспільство задовольняє всі потреби і забезпечує безпеку. Людина народилася, росте, живе і вмирає в суспільстві. Без суспільства життя людини схоже на рибу без води, тобто життя такої людини взагалі під питанням [1].

Звісно між індивідуумами в соціальному середовищі виникають соціальні зв'язки. Два десятиліття досліджень вказують на причинно-наслідкові зв'язки між соціальними відносинами та смертністю, але залишаються важливими питання щодо того, як соціальні відносини впливають на здоров'я, коли з'являються наслідки та як довго вони тривають [2].

Дослідження показують, що міцні соціальні зв'язки пов'язані з тривалим життям. Навпаки, самотність і соціальна ізоляція пов'язані з погіршенням здоров'я, депресією та підвищеним ризиком ранньої смерті. Дослідження показали, що різноманітні соціальні стосунки можуть допомогти зменшити стрес і ризику, пов'язані з серцем [3].

Будь-який стан або поведінка вважалися соціальною проблемою, якщо вони мали негативні наслідки для великої кількості людей. Проте, коли соціальна проблема стає соціальною? На думку деяких соціологів, які дотримуються цієї точки зору, негативні умови та поведінка не є соціальною проблемою, якщо їх не визнають як такі політики або велика кількість громадян

чи інші сегменти нашого суспільства. Інші соціологи стверджують, що негативні умови та поведінку слід вважати соціальною проблемою, навіть якщо їм приділяється мало уваги або зовсім не приділяється жодної уваги [3].

Серед прикладів соціальних проблем можна навести наступні [4].

– Бідність та безпритульність. Майже чверть планети живе в умовах, що загрожують їх безпеці та здоров'ю. Багато хто з цих людей не мають житла, що є однією з основ для їх виживання.

– Зміна клімату. Теплий, мінливий клімат є загрозою для всього світу. Зміна клімату впливає на все населення світу, Союз зацікавлених вчених називає цю соціальну проблему «однією з найбільш руйнівних проблем, з якими коли-небудь стикалося людство».

– Перенаселення. У міру зростання населення світу ресурсів стає все менше. Організація Об'єднаних Націй повідомляє, що поточне населення в 7,7 мільярдів людей, зросте в найближчі десятиліття до 8,5 мільярда людей.

– Громадянські права та расова дискримінація. Права громадян країни на рівне ставлення в суспільному та політичному плані, є одним з найбільш значущих соціальних проблем в усьому світі. Люди, що виділяються расовою приналежністю можуть переслідуватися та принижуватися, їх життя часто у небезпеці у деяких країнах.

– Гендерна нерівність. Жінки все ще відстають від чоловіків у заробітній платі та на керівних посадах. У всьому світі ситуація ще більш критична. За даними ЮНІСЕФ, 12 мільйонів дівчат виходять заміж до повноліття, а 98 мільйонів дівчат старшого шкільного віку не відвідують школу.

– Булінг. Незважаючи на нещодавні ініціативи проти булінгу, багато школярів та студентів визнають, що в своєму житті зазнали знущань. Частина дітей стає жертва так званого кібербулінгу – булінгу у соціальних мережах.

Усі ці соціальні проблеми можуть породжувати ще більшу кількість дрібних, але не менш серйозних проблем, що стають завадою для безпечного життя інших людей. Тому важливим є їх запобігання та активна боротьба. В

даному випадку постає питання забезпечення безпеки соціального життя людини.

Поняття «безпека» доволі складне та є ключовим у системі «природа–людина–суспільство–машина». У центрі цієї системи є людина, яка своєю поведінкою, діяльністю, може завдавати також і шкоди природі, іншій людині, суспільству та машинам [5]. Така система графічно представлена на рисунку 1.1.

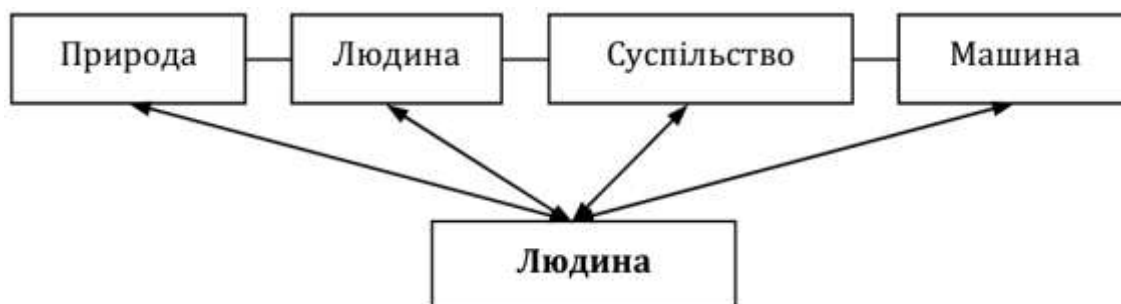


Рисунок 1.1 – Схема взаємодії людини в системі «природа–людина–суспільство–машина» [5]

Термін «безпека» трактується як попередження, усунення та захист від небезпек. Також поняття безпеки можна розділити на дві категорії [5]:

- фізична – захист від загроз, що здатні завдавати шкоди здоров'ю та життю;
- соціальна – захист життєво важливих інтересів людини, як соціальної істоти.

Потреба людини в безпеці є одною із базових потреб, згідно піраміди Маслоу (Рисунок 1.2).

Маючи задоволений нижній рівень потреб у базових потребах існування, людина прагне до збереження та їх убезпечення від втрати. Отже, можна сказати, що почуття безпеки є інстинктивним почуттям людини, тобто тим, яке закладене у неї від природи.

На сьогоднішній день загальні потреби безпеки намагається задовольнити держава. Надаючи доступ до надання медичної допомоги, можливості

застрахувати своє життя та майно, створюючи державні структури, що забезпечують базові потреби безпеки людини, як важливої ланки суспільства.



Рисунок 1.2 – Піраміда потреб Маслоу [6]

Проте часто безпеку людини, як згадувалось вище, може порушувати інша людина. Часто це пов'язано з порушеннями психоемоційного стану людини. Психоемоційний стан можна розділити на дві частини: емоційний та психічний стани. Можна сказати про те, що психічний стан витікає від емоційного стану.

Наш емоційний стан часто диктує те, як ми поведимося. При чому психологи стверджують, що чоловіки і жінки справляються з емоціями зовсім по-різному. Коли жінки засмучені, вони частіше висловлюють свої почуття і шукають підтримки друзів і сім'ї, тоді як чоловіки можуть приховувати свої емоції або віддалятися. Часто те, як ми бачимо поведінку наших батьків, може стати несвідомим шаблоном і нашої поведінки [7].

Психічний стан є мінливим станом людини, виникають під впливом життєвих обставин чи ряду інших факторів. Серед психічних станів можна виділити [8]:

- емоції;

- ейфорія;
- апатія;
- тривожність;
- натхнення;
- тощо.

Можна виділити чотири основних людські емоції:

- смуток;
- злість;
- щастя;
- страх;
- тощо.

З цих чотирьох емоцій, емоція щастя вважається найбільш прийнятною у суспільстві. Проте гнів, страх і смуток повсюдно відчують усі. Ці емоції часто є нормальною реакцією на загрозу та втрату, проте часто ці емоції стають рушієм для загрози безпеки собі та іншим людям, переростаючи в агресію.

Особливим психоемоційним станом людини є відчуття впевненості. Впевненість – це такий стан думки, при якому особистість вважає, що гіпотеза чи прогноз є вірними, або що обраний вектор дій є найкращим та найефективнішим для досягнення цілі [9].

Психологи вважають, що впевненість допомагає людям відчувати себе готовими до життєвого досвіду. Коли люди впевнені, то з більшою ймовірністю рухаються вперед не відступаючи від цілі. І якщо спочатку щось не виходить, впевненість змушує спробувати ще раз. Люди з низьким рівнем впевненості можуть рідше пробувати щось нове. Якщо з першого разу не виходить досягнути цілі, ймовірність того, що вони спробують знову, буде невеликою. Відсутність впевненості може завадити людям повністю реалізувати свій потенціал [10]. Відчуття впевненості може передаватися в тексті стверджувальними реченнями та словами, наприклад, «безперечно», «певна річ», «правда», «сподіваюсь» тощо.

Можна сказати про те, що психічне здоров'я – це не тільки відсутність розладів, а й вміння людини протистояти стресам, створювати навколо себе благополуччя [11].

Існують різні способи оцінки психоемоційного стану. Серед самих прогресивних – це його визначення за зображенням обличчя людини. Варто зазначити, що визначати психоемоційний стан можна як за мімікою всього обличчя, так і за окремими її частинами, наприклад очима. Більш розповсюдженим є визначення психоемоційного стану за так званими опитувальниками. Опитувальники складаються з певного роду запитань для визначення різних станів людини, наприклад, депресивних станів, агресії, ворожості, небезпечності, комфортності, впевненості тощо. Для цього розроблені певні методики. Далі наведено деякі з них.

Наприклад, методика оцінювання агресивності у стосунках А. Ассінгера [12] дає можливість оцінити, чи достатньо людина коректно ставиться до оточення і чи легко спілкуватися з нею. Методика диференціальної діагностики депресивних станів В. Зунга може бути використана для диференціальної діагностики депресивних станів для діагностики з метою долікарської діагностики [13]. Методика визначення стресостійкості та соціальної адаптації Т. Холмса та Р. Раге дозволяє визначити на скільки людина здатна витримувати психофізичні навантаження та переносити стреси без шкоди для організму [14]. Методика дослідження впевненості на усвідомлюваному рівні на основі методики Д. Ковача, яка говорить про те, що впевненість залежить від рішучості так само, як рішучість від впевненості, невпевнена у собі людина недовірлива і має безліч упереджень [15].

Психоемоційний стан людини можна оцінювати за її мовленням. В реальному спілкуванні експерт може опиратися на візуальні емоції людини та, наприклад, тон її голосу. Фонетистами визначено, що впевненість мовця залежить від тону, тривалості висловлювання та стабільності ритму, тембр голосу [16].

Саме тон вказує на істинне значення повідомлення (спадні тони говорять про завершеність та категоричність висловлення, висхідні говорять про незавершеність, невпевненість, спадний-висхідний говорить на наявність прихованого підтексту тощо). Тривалість висловлювання думки вказує на зміни емоційного стану мовця, наприклад збільшення тривалості пауз, розтягнута в часі вимова чи її прискорення свідчать про наявність емотивного компонента [16].

Психоемоційний стан доволі сильно може впливати на мовлення. Наприклад, коли людина невпевнена, то може повторювати одне і те ж слово двічі, у мовленні людини можуть з'являтися так звані «слова-паразити», мовлення затягне з великими паузами. В позитивному настрої, до прикладу, в мовленні людини можуть з'являтися слова-вигуки «ах», «ех», речення набувають окличної форми. При чому складно визначити психоемоційний стан людини лише на одному рівні. Часто засоби вираження емотивності вивчаються на фонологічному, лексичному та фразеологічному, семантичному рівнях [17]

Психоемоційний стан людини, що говорить, накладає певний відбиток і на синтаксичне оформлення мови. Всі випадки вираження емоцій синтаксичної організації висловлювання можна звести до:

- 1) зміни прийнятого у мові порядку слів;
- 2) зростання кількості випадків граматичної незавершеності;
- 3) переривання ланцюга розгортання тексту;
- 4) порушення граматичного узгодження, аж до повної граматичної дезорганізації мови у стані фрустрації [18]

Проте, визначення психоемоційного стану у текстах є складнішою задачею. Як правило, для ідентифікації та аналізу тексту використовується два типи підходу. Один має справу з синтаксисом, а інший – з семантикою. Синтаксис зазвичай визначає граматичну структуру тексту, в той час як семантика має справу з визначенням сенсу тексту. Синтаксична модель заснована на структурі аргументів речень по відношенню до дієслів.

Семантичний підхід же заснований на визначенні значення певних слів в контексті конкретного тексту та виявлення сенсу цього тексту [19]

Виявлення психоемоційного стану автора тексту є важливою задачею, адже це не тільки вирішує задачу аналізу особистості, а й може допомогти в запобіганні порушення безпеки інших особистостей. Такі методи можуть застосовуватись локально, або у більш глобальних задачах забезпечення державної безпеки. У даній роботі розглядається визначення рівня впевненості автора тексту, як приклад вирішення задачі з визначення психоемоційного стану людини.

## **1.2 Огляд теоретичних підходів до розв'язку подібних задач**

Як зазначалось вище, семантичний аналіз тексту є одним із методів визначення психоемоційного стану людини. Семантичний аналіз використовується не тільки для такого роду задач.

Семантика – наука, що займається вивченням значення слів у тексті та їх взаємозв'язків [20].

В семантичному аналізі оперують такими поняттями як «ключове слово», «ключові терміни» та «ключові словосполучення».

Ключове слово – слово, що може бути використане для передавання змісту тексту, воно має смислове навантаження, що може передавати суть тексту в сукупності з іншими ключовими словами. Зазвичай таких слів у тексті може бути багато [21].

Ключове словосполучення – це така лексико-граматична єдність, в якій між словами встановлюються відносини, які доповнюють сенси слів, що поєднуються. Такі словосполучення можуть яскравіше передавати сенс тексту, ніж ключові слова.

Ключові терміни – це слова або словосполучення, що використовуються для означення якогось поняття, що є важливим у даному тексті [22].

Часто сенс тексту визначають пошуком не тільки ключових слів, а і ключових словосполучень, термінів [23].

Існують різні методи пошуку ключових слів. Найпопулярніші з них:

- Text Rank;
- частотний аналіз;
- TF-IDF;
- дисперсна оцінка;

TextRank був запропонований в 2004 на основі алгоритму ранжування веб-сторінок PageRank. Належить до так званих графоорієнтованих алгоритмів. Значимість вершини у графі розраховується через значення суміжних вершин. Ребром графа може бути довільне відношення між лексичними одиницями. Для завдання виділення ключових слів – це відношення спільної появи, яке задається відстанню між словами. Дві вершини вважаються суміжними, якщо відповідні їм лексичні одиниці з'являються всередині діапазону  $2 \leq N \leq 10$ . Перед додаванням вершин до графа може виконуватися очищення лексики, наприклад, за приналежністю до частин мови. Після побудови графа виконується підрахунок значущості вузлів, їх ранжування і перші 5-20 зберігаються для обробки як потенційні ключові слова. Послідовності суміжних ключових слів створюють ключові словосполучення, інші залишаються ключовими словами [24].

Частотний аналіз – найпростіший спосіб визначення ключових слів, що визначається частотою появи слова у тексті. Найчастіше всього перед застосуванням даного методу виконується лематизація, тобто слово приводиться до кореневих лексичних компонентів. Чим частіше слово зустрічається у тексті, тим воно вважається важливіше, проте зачасту верхню і нижню частину слів можуть відкидати, адже вони можуть вважатися «шумом», такими словами, що не несуть смислового навантаження, наприклад частки, сполучники тощо [25]

TF-IDF – це статистичний показник, який оцінює, наскільки релевантне слово документу в колекції документів. Це робиться шляхом множення двох показників: скільки разів слово зустрічається в документі, і оберненої частоти цього слова в наборі документів. Цей метод дуже корисний для оцінки слів в

алгоритмах машинного навчання для обробки природної мови (NLP). Метод працює шляхом збільшення пропорційно кількості зустрічання слова в документі, але компенсується кількістю документів, які містять це слово. Слова, які зустрічаються в кожному документі, наприклад, сполучники, мають низький рейтинг, навіть якщо вони з'являються багато разів, таким чином вони не мають важливого значення для цього документа [26].

Дисперсійна оцінка дозволяє відокремити слова, які у тексті розміщені відносно рівномірно (для рівномірно розподілених слів ця оцінка дорівнює нулю), від слів, розподілених нерівномірно. Тобто. це оцінка розрізняючої, дискримінантної сили слів, зокрема для інформаційного пошуку. Обчислюється за формулою:

$$\sigma_i = \sqrt{\frac{\langle d^2 \rangle - \langle d \rangle^2}{\langle d \rangle}} \quad (1.1)$$

де  $\langle d \rangle$  – середнє значення послідовності  $d_1, d_2, d_n$ ,  $n$  – кількість появи окремо взятого слова в тексті [27].

Для автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту пропонується використовувати психоемоційний вектор слів, який складається із слів та оцінок їх семантичної важливості у психоемоційному розрізі фактора впевненості. Для формування цього вектора пропонується використати наявні у відомих розмічених корпусах текстів української мови і слів відомості, значення оцінок семантичної важливості у психоемоційному розрізі фактора впевненості яких коригується вручну. Зокрема відомі наступні корпуси української мови.

Генеральний регіонально анотований корпус української мови [28]. Цей корпус є досить великим та містить понад 90 тисяч текстів, він є репрезентативним, структурований за категоріями (наприклад, філософія, політологія, медицина, хімія тощо) колекція текстів. Є можливість будувати на базі корпусу власні підкорпуси, шукати слова та їхні граматичні форми, сполучення та сортувати, фільтрувати за потребою.

MOVA.info [29]. Корпус української мови, що призначений для пошуку лексем та словоформ в текстах. Тексти категоризовані за певним стилем (наприклад, фольклор, публіцистика, художні тексти тощо). Загальний розмір корпусу 100 мл слів.

UA-GEC: перший анотований GEC-корпус української мови [30]. Даний корпус містить тексти, написані звичайними людьми: дописи в блогах та соцмережах, відгуки тощо. Ці тексти залишені в тому вигляді, в якому вони були створені в першоджерелі, тому містять граматичні, стилістичні та орфографічні помилки. Це може бути корисним для тренування та оцінки програм, що призначені для виправлення граматичних помилок. Даний корпус є анотованим, тобто люди, що перевіряють тексти, виправляють помилки та вказують що виправили, тобто займаються анотацією корпусу.

Проект lang-uk [31]. Проект представляє декілька корпусів української мови, наприклад корпус законів, корпус УберТекст тощо, що містять різного типу тексти: статті з Вікіпедії, художня література, перелік законів та інше. Також проєкт представляє різного роду словники. Серед них варто відмітити словник тональностей, словник наголосів та слів-гетеронімів (ті, що пишуться однаково, проте мають різне значення).

Серед інших типів словників також варто виділити словники:

- Українського веб-корпусу Лейпцизького університету [32] – частотні словники;
- Словопедія [33] – тлумачні словники;
- Горох [34] – зібрання тлумачних, орфографічних, синонімічних словників, тощо;
- Великий електронний словник української мови [35] цікавий тим, що складається з суржика, слів-покручів, а також різних словоформ.

Отже, для автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту необхідним є використання одного з методів пошуку ключових слів для побудови психоемоційного вектора, а також деякий

вхідний набір текстів, а саме корпус української мови для тестування розробленого методу.

### 1.3 Аналіз існуючих програмних рішень

Визначення психоемоційного стану людини за її мовленням доволі популярна задача над якою працюють довгий час. Проте вона є доволі складною, адже не завжди можна чітко виділити емоцію саме в тексті. Особливо це складним є тоді, коли людина намагається передати декілька емоцій у тексті. Проте є декілька сервісів, які вирішують цю задачу.

Tone Analyzer – це проєкт IBM, який спрямований на розшифровку емоційного тону тексту. Сервіс аналізує текст, а точніше кожне слово в ньому, та асоціює його з певною емоцією [36].

Наприклад, слово *difficult* відноситься до негативних емоцій, можливо, до агресії. Так само і *disappointing*. А слова *we*, *our*, *need* означають відкритість співрозмовника та його бажання вам сподобатися. Tone Analyzer розпізнає наступні емоції:

- злість;
- добросовісність;
- відкритість;
- впевненість;
- радість;
- приємність.

Сервіс має відкритий API, тому здобутками IBM можуть користуватися сторонні розробники. Крім цього, можна не лише розпізнавати текст співрозмовника, а й перевіряти листи перед відправкою. Tone Analyzer може підбирати синоніми, що пом'якшують, або ж навпаки, посилюють емоційний тон повідомлення. Єдиним недоліком сервісу є те, що він працює тільки з англійською мовою. Єдиним варіантом розширення мовного пакету є надія на те, що сторонні розробники займуться мовною локалізацією. Наприклад, на

рисунку 1.3 система розпізнала впевненість у вхідному повідомленні та вказала рівень тону.

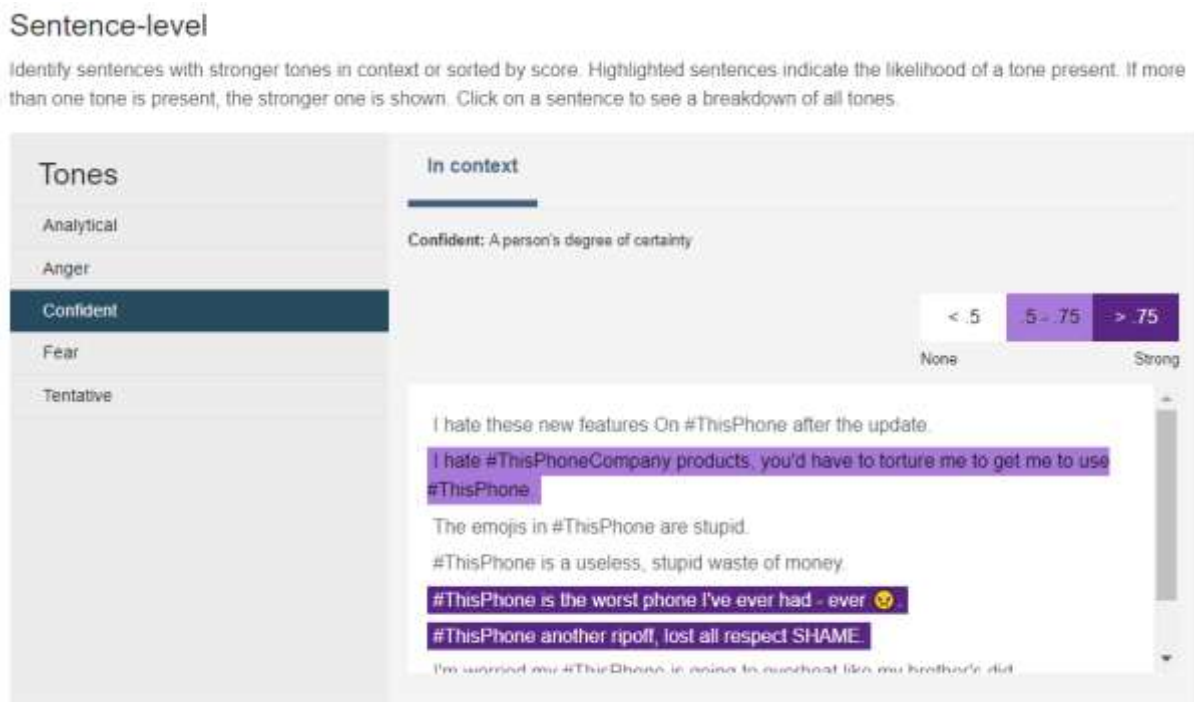


Рисунок 1.3 – Результат аналізу тексту на емоцію [36]

Команда розробників TWIN створила ботів, що також визначають емоційність текстів, а також голосових повідомлень користувачів [37]. Насамперед такі боти використовуються для визначення негативу від клієнтів, щоб або перевести діалог на живу людину, або спробувати за спеціальним сценарієм заспокоїти клієнта. Також модуль можна використовувати для оцінки ефективності скриптів, складання звітів, підрахунку метрик (наприклад, про задоволеність клієнтів спілкуванням з ботом) та побудови різних графіків. Була розроблена нейронна мережа, яка вміє визначати 3 типи емоцій: негативні, нейтральні та позитивні. Нейтральним вважається той текст, який не можна точно віднести до позитивного чи негативного. Отримана точність розпізнавання 75.3% на тексті, довжиною від 1 слова (велику роль відіграє «сміслові навантаження» слів, а не їх кількість). Проте, з визначенням сарказму є ще деякі проблеми, адже це окрема і складна задача.

OneAI Language Skills – це готові для використання вертикально попередньо навчені моделі, упаковані в API. API приймає текст і обробку необхідної мови як вхідні дані, а також відповідає обробленим текстом і витягнутими метаданими як вихідні дані. Дане API дозволяє визначити емоцій тексту та їхню силу, а також містить ще безліч можливостей для роботи з текстами [38]. Приклад визначення емоційної тональності тексту зображений на рисунку 1.4. Єдиним недоліком даного інструменту є те, що він підтримує тільки англійську мову.

The screenshot displays the OneAI Language Skills interface. On the left, there is a 'Text Input' section with a sample text about iPhone updates. Below it is a 'Skill Pipeline' section with a list of skills: 'Sentiments', 'Emotions', and 'Highlights'. The main area shows the 'Original text' with colored overlays indicating sentiment (purple for negative, green for positive) and emotion (yellow for sad, blue for angry, red for happy). A table below the text lists the detected skills, their labels, spans, and values.

SKILL	Label	Span	Value
sentiments	negative	0-1000	1.0000000000000000
emotions	negative	0-1000	1.0000000000000000

Рисунок 1.4 – Результат аналізу тексту на емоцію [38]

Amazon Comprehend – це сервіс обробки природної мови (NLP), в якому для виявлення в тексті аналітичних даних застосовуються технології машинного навчання. Amazon Comprehend надає API для розпізнавання сутностей, у тому числі користувачів, класифікації користувачів, вилучення ключових фраз, аналізу емоційного забарвлення тексту, настрою користувача та багато іншого. API аналізу емоцій повертає загальний емоційний тон тексту (позитивний, негативний, нейтральний або змішаний). Цільові настрої дають змогу отримати більш детальну інформацію про емоційний стан, визначення настрою (позитивне, негативне, нейтральне чи змішане) щодо відношення до об'єкта в

тексті [39]. Загальна схема роботи з даними за допомогою Amazon Comprehend зображено на рисунку 1.5.



Рисунок 1.5 – Загальна схема роботи з даними за допомогою Amazon Comprehend [39]

QM Analyzer – модуль мовленнєвої аналітики, що дозволяє в автоматичному режимі відстежувати події на телефонній лінії, мовленнєву активність дикторів, розпізнавати мовлення та аналізувати емоції. Для аналізу емоційного стану QM Analyzer вимірює фізичні характеристики мовленнєвого сигналу: амплітуда, частотні та тимчасові параметри, шукає ключові слова та терміни, що характеризують відношення мовленнєвого сигналу до теми певної теми. При аналізі голосу перші кілька секунд система накопичує дані та оцінює, який емоційний тон розмови був нормальним, і далі, відмовляючись від нього, фіксує зміни тону в позитивній або негативній стороні [40].

Отже, з вище описаного слідує, що є багато сервісів та інструментів, що дозволяють провести визначення психоемоційного стану людини за її мовленням. При чому деякі інструменти призначені для обробки текстової інформації, інші для аудіозаписів. Також варто зазначити, що існують технології, що вирішують дану задачу за відеозображенням мовця.

## 1.4 Аналіз сучасних засобів створення програмного забезпечення

Розвиток технологій в сучасному світі дав поштовх для створення багатьох типів пристроїв таких, як персональний комп'ютер, планшет, смартфон тощо. Проте усі вони мають свої особливості, з якими стикаються програмісти при створенні програмного забезпечення. Відповідно до типу пристроїв, створюють відповідного типу застосунки, які б забезпечували користувачам доступ до необхідних компонентів програми.

Умовно можна виділити три категорії застосунків:

- застосунки для персональних комп'ютерів;
- застосунки для мобільних телефонів;
- веб-застосунки.

Застосунки для ПК, або як їх ще називають десктопні застосунки – заслужено вважаються найшвидшими та найбільш широко функціональними. Варто зазначити, що ці параметри на пряму залежать і від потужностей ПК [41].

Серед переваг таких варто відзначити [41]:

- функціональність – можливість реалізувати будь-яку ідею для користування, з максимально зручним інтерфейсом користувача;
- безпека та надійність – на відміну від веб-застосунків десктопні застосунки більш захищені від стороннього доступу до даних користувача;
- мобільність – додатки розроблюються для різних операційних систем, що передбачає використання їх на будь яких умовах;
- швидкодія – застосунок працює за принципом використання ресурсів комп'ютера – його пам'яті, процесора тощо, якщо ПК доволі потужний, то проблем із запуском застосунків не виникатиме.

Головним недоліком десктопних застосунків є складність розробки та важкість внесення змін в програму.

Мобільний застосунок – це програмне забезпечення, яке спеціально розробляється на основі можливостей сучасних гаджетів. Мобільний застосунок може застосовуватись не тільки з метою пограти в ігри, послухати музику, це

повноцінний помічник та можливість доступу до різноманітної інформації. Серед переваг виділяють [42]:

- інтерфейс – розроблений спеціально для роботи з мобільною технікою;
- активні сповіщення – людина, яка використовує програму, отримує повідомлення, завдяки чому підвищується якість та тісна взаємодія;
- розширення персоналізації та індивідуальна програма – програма зберігає особисту інформацію користувачів;
- доступність – можливість використання великої кількості ресурсів, наприклад, соціальні мережі або карти;
- адаптивність – можливість підлаштувати розробку під режим користувача, наприклад, налаштувати повідомлення лише у вечірні години.

Головним недоліком є недостатня розповсюдженість смартфонів серед старшого населення, що звужує аудиторію користувачів.

Веб-застосунок — це будь-яка програма, яка використовує веб-сайт як інтерфейс для доступу до інформації. Користувачі можуть легко отримувати доступ до програми з будь-якого комп'ютера, який під'єднано до Інтернет, за допомогою стандартної програми – браузера. Варто звернути увагу на те, що практично будь-яке програму можна створити як веб-застосунок. Веб-застосунки можна відкривати за допомогою спеціальних програм-браузерів, таких як Google Chrome, Opera Microsoft Edge. Програма працює на веб-сервері, а не на ПК, тому потребує мінімум ресурсів зі сторони користувацького пристрою [43]. Серед переваг варто відзначити:

- доступність – на відміну від звичайних програм, веб-застосунки відкриті будь-де, будь-коли та за допомогою будь-якого ПК з доступом до Інтернет;
- швидке налаштування – інтерфейс веб-застосунків простіше змінити, ніж застосунки для ПК та смартфонів. Завдяки цьому простіше оновлювати зовнішній вигляд програми або змінювати введення даних.

- сумісність – користувачу не потрібно думати про те чи взагалі сайт відкриється на його пристрої, про це вже попіклувались розробники, використовуючи стандарти розробки веб-застосунків;

- прямий доступ до останньої версії даних – можливість отримати доступ до даних в реальному часі.

Основним недоліком є недосконалість інтерфейсів та залежність від підключення до Інтернет.

Отже, з вищенаведеного випливає, що десктопні застосунки є найбільш стабільним, безпечним та потужним програмним забезпеченням в порівнянні з іншими типами застосунків.

При розробці десктопних застосунків є можливість використати одну із платформ для розробки. На сьогоднішній день основними платформами для розробки ПЗ є:

- .NET;
- Java.

Платформа .NET – це безкоштовна платформа для створення програмного забезпечення з відкритим вихідним кодом, що була створена Microsoft. Ця платформа пропонує розробникам різноманітні інструменти програмування та принципи створення ПЗ, які можна використовувати для розробки широкого спектру застосунків для Інтернету, мобільних пристроїв, персональних комп'ютерів, розробки ігор. Платформа .NET позиціонується як незалежна від мови платформа, що означає, що розробка може здійснюватися на кількох сумісних мовах, які включають C++, C#, VB.NET, IronPython, Visual COBOL, IronRuby тощо. Також серед переваг даної платформи є те, що вона дозволяє створювати близько 12 типів проєктів. Платформа має потужну підтримку та постійне оновлення та створення бібліотек як самою Microsoft, так і сторонніми розробниками [44].

Платформа Java – це набір програм, які допомагають програмістам ефективно розробляти й запускати програми мовою програмування Java. Вона включає в себе механізм виконання, компілятор і набір бібліотек. Це набір

комп'ютерного програмного забезпечення та специфікацій. Джеймс Гослінг розробив платформу Java в Sun Microsystems, а пізніше її придбала корпорація Oracle [45].

Деякі важливі напрямки, які допомагає реалізувати платформа Java:

- розробка застосунків для Android;
- створення корпоративного програмного забезпечення
- розробка наукових обчислювальних програм;
- використання для аналізу великих даних
- програмування апаратних пристроїв на Java
- використання для серверних технологій Apache, JBoss, GlassFish тощо.

Механізм, який забезпечує середовище виконання для керування кодом Java називається віртуальною машиною Java. Віртуальна машина перетворює байт-код Java на машинну мову. В інших мовах програмування компілятор створює машинний код для певної системи. Однак компілятор Java створює код для віртуальної машини Java [45].

Для програмної реалізації методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту найкраще буде віддати перевагу платформі .NET, адже вона має широкий набір бібліотек та інструментів для створення різного типу програмного забезпечення.

### **1.5 Мета, задачі та вимоги до реалізації інформаційної системи**

Метою кваліфікаційної роботи бакалавра є розробка методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту. Для досягнення мети потрібно вирішити такі задачі:

1. Провести аналіз предметної області, в процесі якого виконати огляд теоретичних підходів до розв'язку задач автоматизованого визначення рівню впевненості за семантичним аналізом тексту.

2. Проаналізувати існуючі програмні рішення для вирішення задачі КРБ.

3. Проаналізувати сучасні засоби для створення програмних рішень та обрати комплект засобів для створення власної програмної реалізації.

4. Розробити метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту, виконати на його основі проектування інформаційної системи.

5. Створити та протестувати відповідну програмну реалізацію, яка буде реалізовувати запропонований метод. Для зручності використання створити інструкцію користувача до розроблюваного програмного продукту.

Розроблена відповідно до методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту інформаційна система має виконувати такі основні функції:

- робота з корпусом слів української мови (перегляд, додавання видалення);

- експорт та імпорт слів між корпусом слів української мови та психоемоційного вектору для категорії впевненості;

- робота з елементами психоемоційного вектору для категорії впевненості (перегляд, привласнення оцінок впливу, коригування, додавання слів);

- завантаження чи введення й збереження дослідного тексту для аналізу;

- первинна обробка дослідного тексту (стоп-символи, регістр);

- перетворення дослідного тексту у масив слів;

- обрахунок значень кількостей слів для кожного унікального слова;

- побудова вектору слів тексту й зіставлення кожному унікальному слову як значень кількостей слів так і відповідних значень оцінок впливу;

- видалення незначущих слів тексту, для яких відсутні значення оцінок впливу;

- обрахунок інтегрованих значень добутку кількостей слів впевненості автора на відповідних значень оцінок впливу для кожного значущого слова тексту;

- обрахунок оцінки рівню впевненості автора;

- обрахунок оцінки рівню емоційної напруженості тексту;
- відображення висновку для користувача стосовно рівню впевненості та рівню емоційної напруженості тексту.

## Розділ 2 Проектування інформаційної системи

### 2.1 Метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту

Для автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту запропоновано метод, схематично проілюстрований на рисунку 2.1.

Вхідними даними пропонованого методу є множина слів української мови та безпосередньо сам текст, який потребує аналізу на визначення рівню впевненості автора.

Метод проходить у декілька кроків, і на першому кроці відбувається формування психоемоційного вектору для категорії впевненості. Також у рамках першого кроку виконується робота з корпусом слів української мови та психоемоційним вектором, а саме експорт та імпорт слів між корпусом слів української мови та психоемоційним вектором для категорії впевненості. Також тут відбувається безпосередньо робота з елементами психоемоційного вектору для категорії впевненості – перегляд, привласнення оцінок впливу  $VW$ , коригування, додавання слів.

Другим кроком є завантаження тексту та його первинна обробка. На цьому кроці можна завантажити чи ввести для збереження дослідний текст для аналізу. Робота з очисткою тексту та переведення до одного регістру, після якої відбувається перетворення дослідного тексту у масив слів та обрахунок значень кількостей слів  $SW$  для кожного унікального слова.

На третьому кроці виконується обрахунок інтегрованих значень впевненості автора для окремих слів тексту. Послідовно будується вектор слів тексту й зіставлення кожного унікального слову як значень кількостей слів  $SW$  так і відповідних значень оцінок впливу  $VW$ . Після чого слідує процес видалення незначущих слів тексту, для яких відсутні значення оцінок впливу  $VW$  та обрахунок інтегрованих значень  $SW \cdot VW$  впевненості автора для кожного

значущого слова тексту на основі значень кількостей слів  $CW$  та відповідних значень оцінок впливу  $VW$ .



Рисунок 2.1 – Схема методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту

На останніх кроках пропонованого методу відбувається обрахунок оцінки рівню впевненості автора та обрахунок оцінки рівню емоційної напруженості тексту. Ці кроки можуть виконуватись незалежно один від одного.

Оцінка рівню впевненості автора тексту  $O$  визначається наступним чином:

$$O = \frac{\sum_{i=1}^n (CW_i \cdot VW_i)}{m}, \quad (2.1)$$

де  $n$  – кількість унікальних значущих слів дослідного тексту,  $m$  – загальна кількість слів у дослідному тексті для аналізу,  $CW_i$  – значення кількості зустрічань слова  $i$  дослідному тексті,  $VW_i$  – значення привласненої оцінки впливу слова  $i$  на рівень впевненості згідно психоемоційного вектору для категорії впевненості.

У таблиці 2.1 наведено шкалу оцінки рівня впевненості згідно психоемоційного вектору для категорії впевненості

Таблиця 2.1 – Шкала оцінки рівня впевненості згідно психоемоційного вектору

<b>Градація</b>	<b>Значення</b>
-1..-0,6	Надвисокий рівень невпевненості
-0,59..-0,3	Високий рівень невпевненості
-0,3..- 0,1	Середній рівень невпевненості
-0,09..0	Низький рівень невпевненості
0,01..0,1	Низький рівень впевненості
0,11..0,3	Середній рівень впевненості
0,31..0,59	Високий рівень впевненості
0,6..1	Надвисокий рівень впевненості

Оцінка рівню емоційної напруженості тексту  $E$  визначається наступним чином:

$$E = \frac{\sum_{i=1}^n |CW_i \cdot VW_i|}{m}, \quad (2.2)$$

де  $n$  – кількість унікальних значущих слів дослідного тексту,  $m$  – загальна кількість слів у дослідному тексті для аналізу,  $CW_i$  – значення кількості зустрічань слова  $i$  дослідному тексті,  $VW_i$  – значення привласненої оцінки впливу слова  $i$  на рівень впевненості згідно психоемоційного вектору для категорії впевненості.

У таблиці 2.2 наведено шкалу оцінки рівню емоційної напруженості тексту.

Таблиця 2.2 – Шкала оцінки рівню емоційної напруженості тексту

Градація	Значення
0,00 – 0,1	Низький рівень емоційної напруженості
0,11 – 0,3	Середній рівень емоційної напруженості
0,31 – 0,59	Високий рівень емоційної напруженості
0,6 – 1	Надвисокий рівень емоційної напруженості

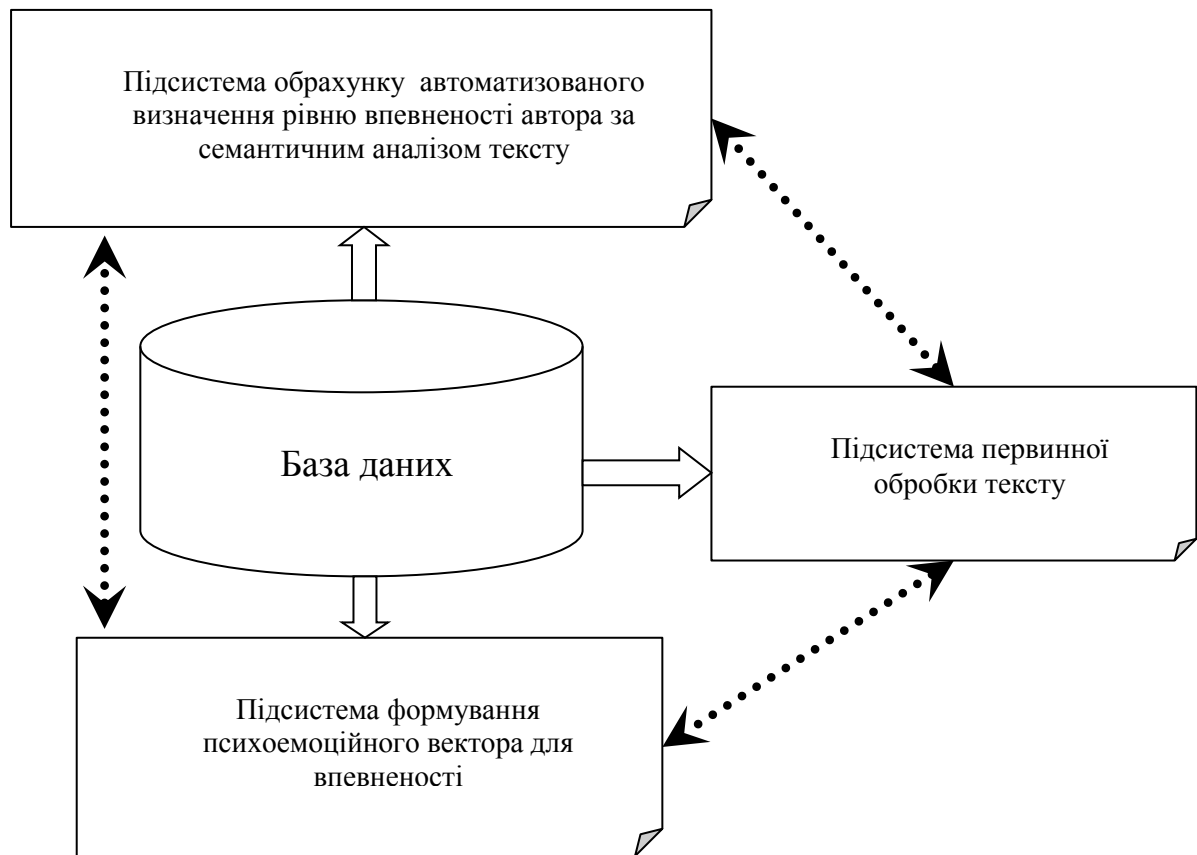
Вихідними даними запропонованого методу визначення рівню впевненості автора за семантичним аналізом є оцінка рівню впевненості автора тексту та оцінка рівню емоційної напруженості тексту.

## 2.2 Інформаційна структура системи

### 2.2.1 Проектна архітектура системи та взаємозв'язок компонентів

Інформаційна структура системи для запропонованого в 2.1 методу автоматизованого визначення рівню впевненості автора за семантичним

аналізом тексту зображено рисунку 2.2. Детальніше про кожну підсистему та її призначення описано нижче.



Рисунком 2.2 – Інформаційна структура системи автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту

На підсистемі первинної обробки тексту передбачається робота з завантаженим текстом або текстом уведеним користувачем, а саме: очистка тексту від стоп-символів (видаляються усі символи окрім літер українського алфавіту) та переводяться до нижнього регістру для подальшого перетворення дослідного тексту у масив слів та обрахунок значень кількостей слів  $SW$  для кожного унікального слова.

Підсистема формування психоемоційного вектора для впевненості передбачає роботу користувача з корпусом слів української мови, де є можливість переглядати, додавати та видаляти обрані слова. Також передбачається можливість експорту та імпорту слів між корпусом слів української мови та психоемоційного вектору для категорії впевненості. У

рамках підсистеми що розглядається ще одним важливим аспектом є робота з елементами психоемоційного вектору для категорії впевненості: перегляд, привласнення оцінок впливу *VW*, коригування, додавання слів.

Підсистема обрахунку автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту виконує обрахунок інтегрованих значень впевненості автора для окремих слів тексту, будуючи послідовно вектор слів тексту й зіставляючи кожне унікальне слово як значень кількостей слів *SW* так і відповідних значень оцінок впливу *VW*. Також у рамках функціонального призначення даної підсистеми є видалення незначущих слів тексту, для яких відсутні значення оцінок впливу *VW* та обрахунок інтегрованих значень *SW·VW* впевненості автора для кожного значущого слова тексту на основі значень кількостей слів *SW* та відповідних значень оцінок впливу *VW*. Як висновок для користувача запропоновано обрахунок оцінки рівню впевненості автора тексту та оцінки рівню емоційної напруженості тексту.

### **2.2.2 Інформаційна модель**

Для роботи методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту було створено базу даних із наступними таблицями: «UnikSlovtextVector», «SlovKorpSliv», «VklSlivUkrMovToPsychVect», «SlovaKorpSliv», «ResTexts» та «UnikSlovKorpSliv». Даталогічну модель БД наведено на рисунку 2.3.

Відповідно до структури БД створили таблиці із вхідними даними в них. Таблиця «ResTexts» (таблиця 2.3) зберігає тексти, що необхідно дослідити та містить наступні поля: ID, Text, ObrobText, DateTimeOfRedag, VpevnenistAuthorTextRating та EmotNaprujTextRating.

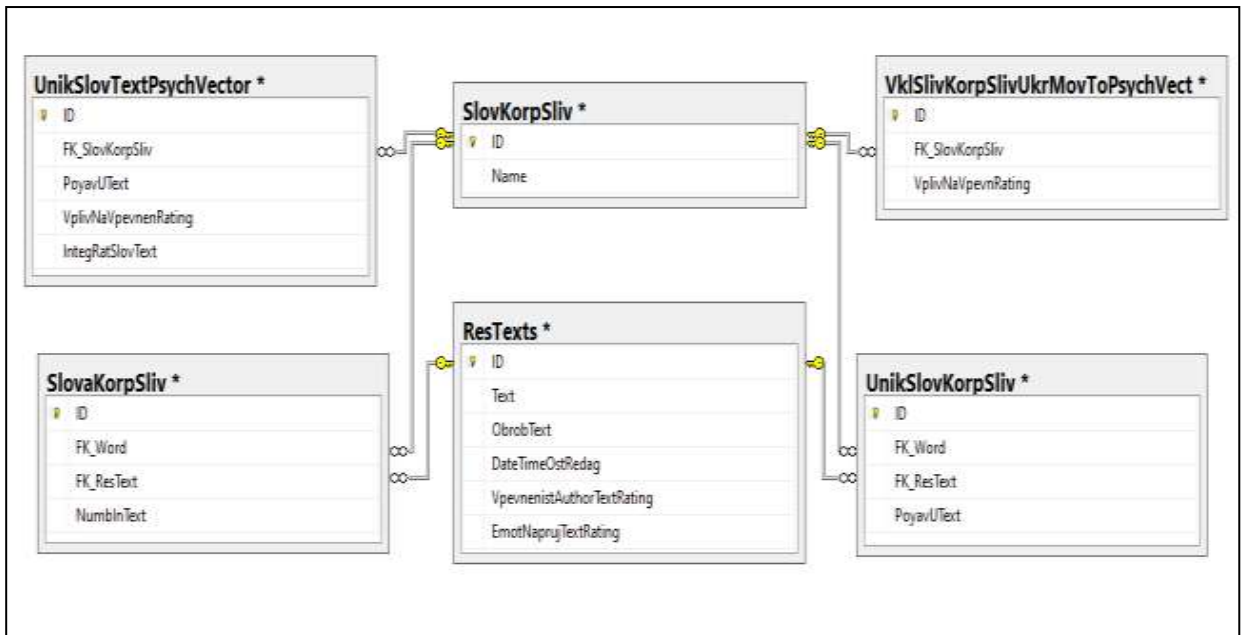


Рисунок 2.3 – Даталогічна модель бази даних системи автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту

Таблиця 2.3 – Атрибути таблиці «ResTexts»

№ п/п	Назва атрибуту	Тип даних	Опис
1.	ID	int	Первинний ключ, числовий ідентифікатор.
2.	Text	text	Зміст текстового матеріалу.
3.	ObrobText	text	Текст, що приведено до нормального вигляду для подальшого аналізу.
4.	DateTimeOstRedag	datetime	Дата й час останнього редагування тексту.
5.	VpevnenistAuthorTextRating	int	Оцінка рівню впевненості автора тексту.
6.	EmotNaprujTextRating	int	Оцінка рівню емоційної напруженості тексту.

Таблиця «SlovKorpSliv» (таблиця 2.4) містить слова корпусу слів. Таблиця має два поля: ID та Name.

Таблиця 2.4 – Атрибути таблиці «SlovKorpSliv»

№ п/п	Назва атрибуту	Тип даних	Опис
1.	ID	int	Первинний ключ, числовий ідентифікатор.
2.	Name	Varchar(50)	Слово корпусу слів.

Наступна таблиця «SlovaKorpSliv» (таблиця 2.5) містить інформацію про слова з корпусу, що знаходяться в тексті та містить наступні поля: ID, FK\_Word, FK\_ResText, NumbInText.

Таблиця 2.5 – Атрибути таблиці «SlovaKorpSliv»

№ п/п	Назва атрибуту	Тип даних	Опис
1.	ID	int	Первинний ключ, числовий ідентифікатор.
2.	FK_Word	int	Вторинний ключ, посилання на таблицю «SlovKorpSliv», на відповідний запис для співставленням із певним словом.
3.	FK_ResText	int	Вторинний ключ, посилання на таблицю «ResText», на запис для співставленням із відповідним текстом, що досліджується.
4.	NumbInText	int	Порядковий номер слова у тексті.

Таблиця «UnikSlovKorpSliv» (таблиця 2.6) містить дані про унікальні слова в тексті з корпусу слів, існують наступні поля: ID, FK\_Word, FK\_ResText, RoyavUText.

Таблиця 2.6 – Атрибути таблиці «UnikSlovKorpSliv»

№ п/п	Назва атрибуту	Тип даних	Опис
1.	ID	int	Первинний ключ, числовий ідентифікатор.
2.	FK_Word	int	Вторинний ключ, посилання на таблицю «SlovKorpSliv», на відповідний запис для співставленням із певним словом.
3.	FK_ResText	int	Вторинний ключ, посилання на таблицю «ResText», на відповідний запис для співставленням із текстом, що досліджується.
4.	PoyavUText	int	Кількість появ слова у тексті.

Таблиця «UnikSlovTextPsychVector» (таблиця 2.7) призначена для збереження даних унікальних слів тексту із психоемоційного вектору. Таблиця містить наступні поля: ID, FK\_SlovKorpSliv, PoyavUText, VplivNaVpevnenRating, IntegRatSlovText.

Таблиця 2.7 – Атрибути таблиці «UnikSlovTextPsychVector»

№ п/п	Назва атрибуту	Тип даних	Опис
1.	ID	int	Первинний ключ, числовий ідентифікатор.
2.	FK_SlovKorpSliv	int	Вторинний ключ, посилання на таблицю «SlovKorpSliv», на запис для співставленням із відповідним словом з корусу.
3.	PoyavUText	int	Кількість появ слова у тексті.
4.	VplivNaVpevnenRating	int	Оцінка впливу на впевненість
5.	IntegRatSlovText	int	Інтегрована оцінка для слова у тексті.

Таблиця «VklSlivKorpSlivUkrMovToPsychVect» (таблиця 2.8) зберігає інформацію щодо включення слів із корпусу слів української мови до психоемоційного вектору, та містить наступні поля: ID, FK\_SlovKorpSliv, VplivNaVpevnRating.

Таблиця 2.8 – Атрибути таблиці «VklSlivKorpSlivUkrMovToPsychVect»

№ п/п	Назва атрибуту	Тип даних	Опис
1.	ID	int	Первинний ключ, числовий ідентифікатор
2.	FK_SlovKorpSliv	int	Вторинний ключ, посилання на таблицю «SlovKorpSliv», на запис для співставлення із відповідним словом з корпусу.
3.	VplivNaVpevnRating	int	Числова оцінка впливу слова на впевненість автора.

Таким чином було створено базу даних методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту, реалізовано усі необхідні для роботи таблиці та заповнено їх початковими даними.

### 2.3 Вибір засобів розробки інформаційної системи

Для створення й програмної реалізації методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту необхідним є вибір мови програмування, на якій буде створюватись застосунок, зручне середовище програмування для обраною мови програмування та вибір

У наступних підрозділах вибір буде розглянуто й обґрунтовано детальніше.

### 2.3.1. Вибір мови програмування

Так як у попередньому розділі було обрано створення віконного додатку за допомогою платформи .NET, доцільним буде зупинити вибір на одній з мов, яка підтримується цією платформою. Платформа .NET може працювати з кількома мовами програмування, такими як C#, VB.NET, C++ і F# та інші [46]. Проте найдоцільнішим буде вибір саме мови програмування C#, адже ця мова було створена спеціально для .NET [[https://uk.wikipedia.org/wiki/C\\_Sharp](https://uk.wikipedia.org/wiki/C_Sharp)].

C# – це мова програмування, що відповідає принципам об'єктно-орієнтованої парадигми. За допомогою C# програмісти можуть створювати різні надійні, довговічні та безпечні програми, які працюють у .NET. Ця структура дозволяє створювати багатоплатформні програми, що дуже корисно, оскільки економить час на розробку, якщо потрібно запускати програми на різних платформах. C# допомагає створювати надійні програми. Однією з переваг цієї мови програмування є те, що збирач сміття .NET автоматично виділяє та звільняє пам'ять програми. Крім того, C# має обробку винятків, а це в свою чергу забезпечує ефективне виявлення та відловлення помилок [47].

В загальному мова програмування просто для застосування, вона типобезпечна, швидко розвивається завдяки відкритому коду [47].

Можна підсумувати, що мова програмування C# є вдалим вибором для створення й програмної реалізації методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту, адже є гнучкою та потужною мовою програмування з постійними оновленнями від Microsoft.

### 2.3.2. Вибір редактора програмного коду

Для створення програм програмісти використовують спеціальні середовища програмування, що спрощують та автоматизують процес розробки. Такі середовища мають ряд корисних функцій – підсвітка синтаксису, пошук помилок у коді, зручна навігація між різними частинами коду тощо. Для мови

програмування C# також є середовище програмування, за допомогою якого зручно і швидко створювати, тестувати застосунки.

Розробники, які бажають створювати програми для хмарних служб, веб-додатків, настільних і мобільних платформ, можуть використовувати для цих цілей Microsoft Visual Studio. Це інтегроване середовище розробки (IDE) спочатку було створено для розробки програм для операційної системи Microsoft Windows, а також веб-застосунків. Сьогодні це середовище дозволяє робити більше, надаючи більше можливостей з контролю застосунків, контролю версій і гнучких планувань [48]. Основні переваги Microsoft Visual Studio включають підтримку рефакторингу коду редактором коду та компонент IntelliSense для завершення коду. Microsoft Visual Studio також має ряд вбудованих інструментів, таких як конструктор класів, конструктор схем бази даних, веб-дизайнер і конструктор форм, який створює програми з графічним інтерфейсом. Також дане середовище програмування підтримує не тільки мову програмування C#, а і Visual Basic, C, C++, F#, додатково є підтримка HTML/XHTML, JavaScript і CSS [48].

Отже, з впевненістю можна вважати, що Microsoft Visual Studio ідеально підходить для реалізації методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту.

### **2.3.3 Вибір СКБД**

Для забезпечення доступу до даних у створеній базі даних використовується спеціальне програмне забезпечення, яке має назву система керування базами даних. СКБД надає можливість створювати, зберігати проводити різні оновлення та шукати потрібну інформацію в базі даних.

Серед різноманітних доступних варіантів програмного забезпечення для керування базами даних Microsoft SQL Server є одним із найпотужніших на сьогоднішній день [49].

Microsoft SQL Server – це система керування реляційною базою даних (RDBMS), вона підтримує велику кількість додатків для обробки транзакцій та аналітики. SQL Server побудований на основі структури таблиць на основі рядків, яка з'єднує відповідні елементи даних у різних таблицях один з одним, при цьому уникає надлишкового зберігання даних у кількох місцях у базі даних. Така модель забезпечує посилальну цілісність для підтримки точності даних [49]. Серед переваг варто відзначити [50]:

- SQL Server підтримує бази даних великого розміру (терабайт), що дає можливість зберігати великі обсяги даних та мати можливість нарощувати базу даних з часом;

- дана СКБД є безпечною, адже адміністратори можуть надавати різного рівня права доступу до бази даних;

- безперервність роботи забезпечується можливістю доступу до бази навіть у разі відключення системи, а також є функції автоматичного відновлення даних за потреби;

- наявність вбудованих інструментів звітності, які підтримують створення діаграм і графіків.

Отже, для реалізації методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту вирішено використати .NET, мову програмування C#, Microsoft Visual Studio та СКБД Microsoft SQL Server. Даний набір інструментів для розробки забезпечить якісну та швидку реалізацію задачі.

## Розділ 3 Програмна реалізація інформаційної системи

### 3.1 Структура та функціональне призначення програмних складових системи

Згідно розробки методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту було сформовано структуру майбутньої програмної реалізації, діаграма класів якої показана на рисунку 3.1.

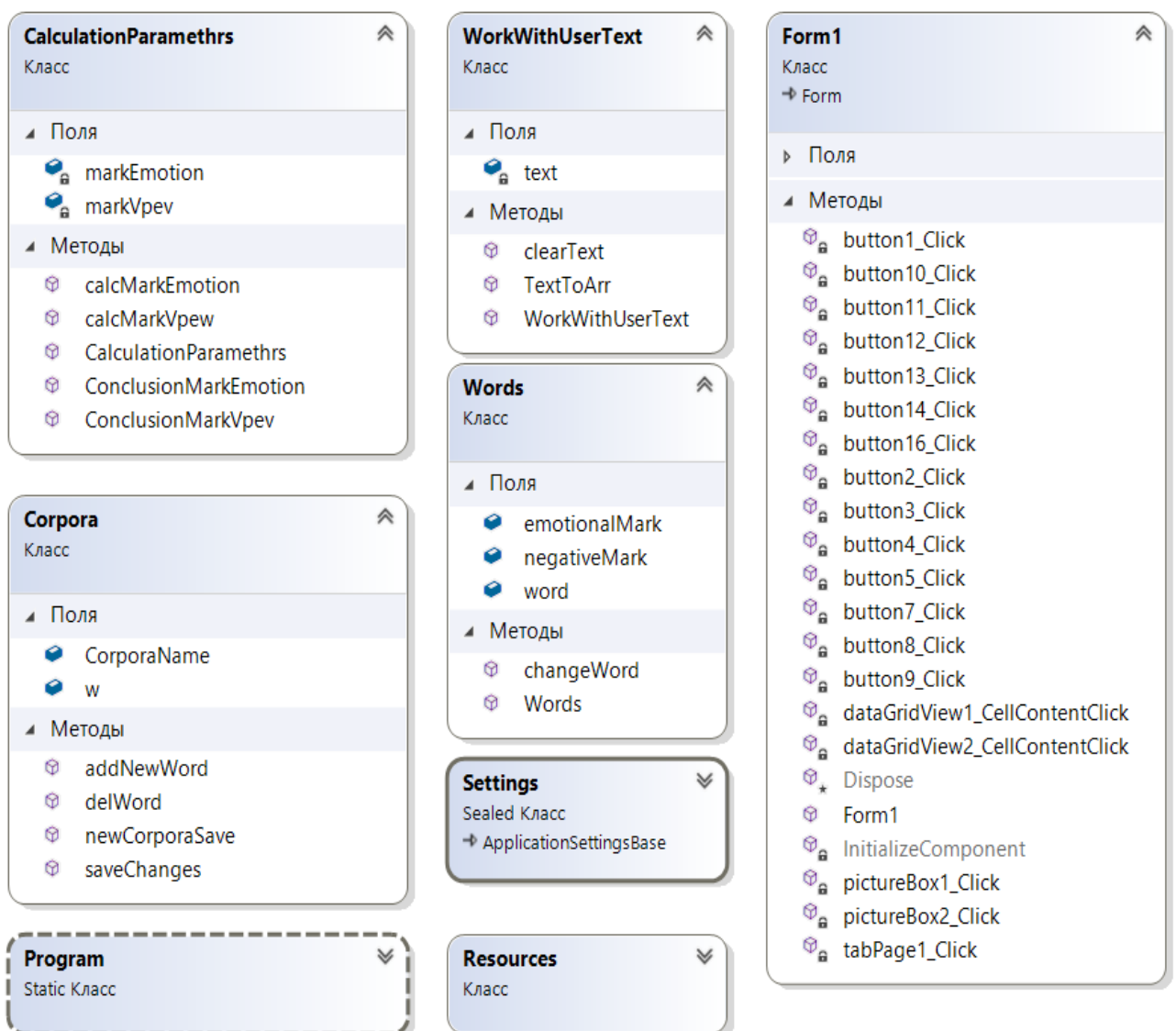


Рисунок 3.1 – Діаграма класів програмної реалізації для методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту

У класі Words зберігається інформація по словам, така як безпосередньо саме слово, оцінка емоційного забарвлення та оцінка впевненості. Також у класі є метод changeWord(), який зберігає зміни у слові.

У класі Corpora зберігається інформація про корпус та масив слів з їх оцінками. З даного класу доступний функціонал додавання нового слова до корпусу (метод addNewWord()), видалення слова з корпусу (метод delWord()), збереження нового корпусу у базі, який користувач зчитає з файлу (метод newCorporaSave()), збереження змін у існуючому корпусі (метод saveChanges()).

Клас CalculationParamethrs реалізовує обрахунок оцінок впевненості автора та оцінку емоційного забарвлення тексту. Метод calcMarkEmotion() рахує оцінку рівня емоційної напруженості тексту, а метод calcMarkVpew() обраховує оцінку рівню впевненості автора. За обрахованими параметрами формується висновок про рівень емоційної напруженості та рівень впевненості автора методами ConclusionMarkEmotion() та ConclusionMarkVpew() відповідно.

Клас WorkWithUserText призначений для попередньої обробки користувацького тексту, який потребує аналізу. Має методи очистки тексту від стоп-символів та переведення тексту у зручну для аналізу та подальшої обробки структуру даних.

Детальніше про функціонування та деталі реалізації будуть розкриті у наступних підрозділах.

### **3.2 Особливості реалізації програмних складових системи**

Під час реалізації методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту було створено ряд функцій для досягнення поставленої в 1.5 мети. Зокрема, для використання нового корпусу слів української мови є можливість завантажити корпус з файлу. Код методу openCorpora() наведено нижче:

```
void openCorpora() {
    dataGridView1.Rows.Clear();
    textBox3.Text = "";
}
```

```

Encoding code = Encoding.UTF8;
IEnumerable<string> s = new string[] { };
OpenFileDialog myFile = new OpenFileDialog();
myFile.Title = "Відкрити корпус слів";
myFile.Filter = "txt files (*.txt)|*.txt|All files (*.*)|*.*";
if (myFile.ShowDialog() == DialogResult.OK)
{
    s = File.ReadLines(myFile.FileName, code);
}
foreach (var s1 in s)
{
    dataGridView1.Rows.Add(s1.ToString());
}
}

```

За допомогою фільтру користувачеві будуть доступні для перегляду лише вказані у фільтрі типи текстових документів, які буде видно через діалогове вікно (рисунок 3.2).

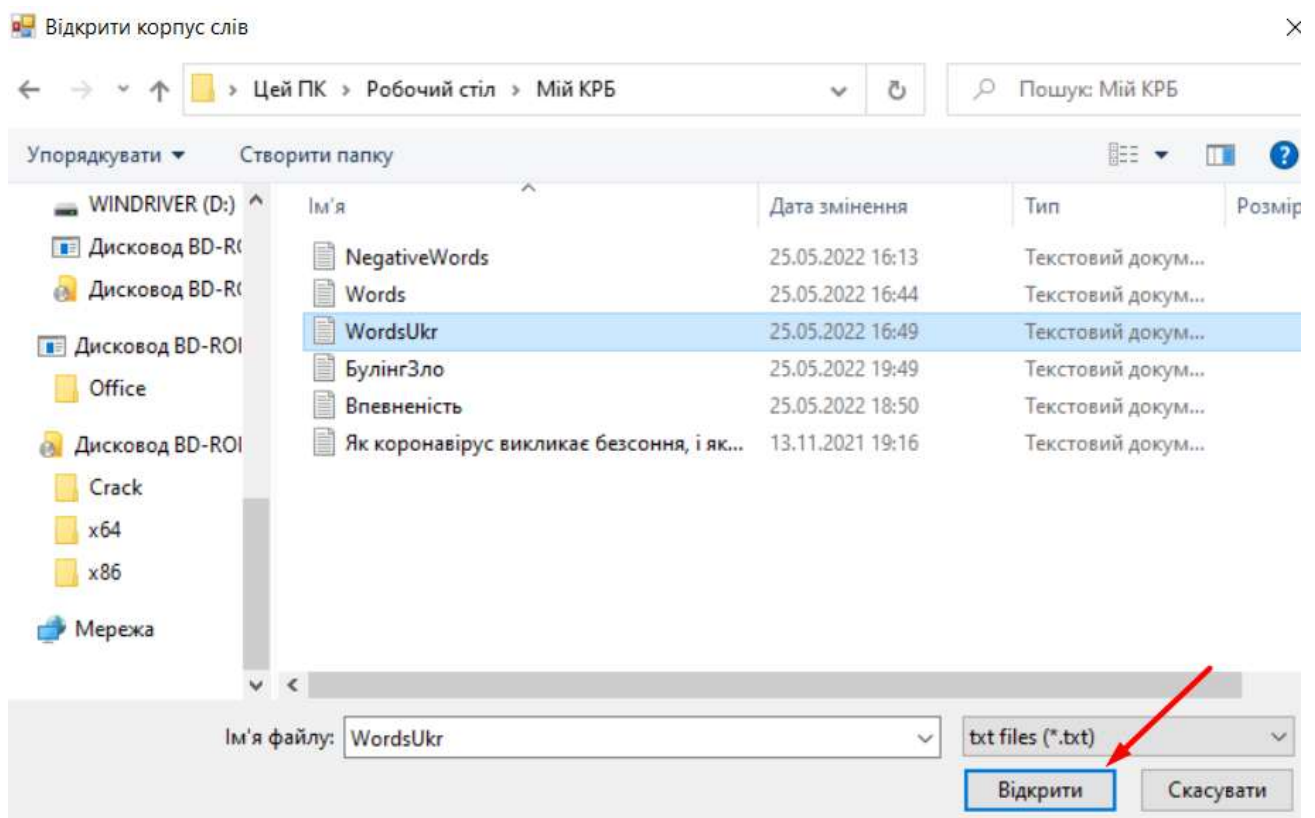


Рисунок 3.2 – Діалогове вікно для відкривання нового корпусу слів української мови

Після відкривання нового корпусу він буде відображений на відповідній вкладці у відповідній таблиці (рисунок 3.3).

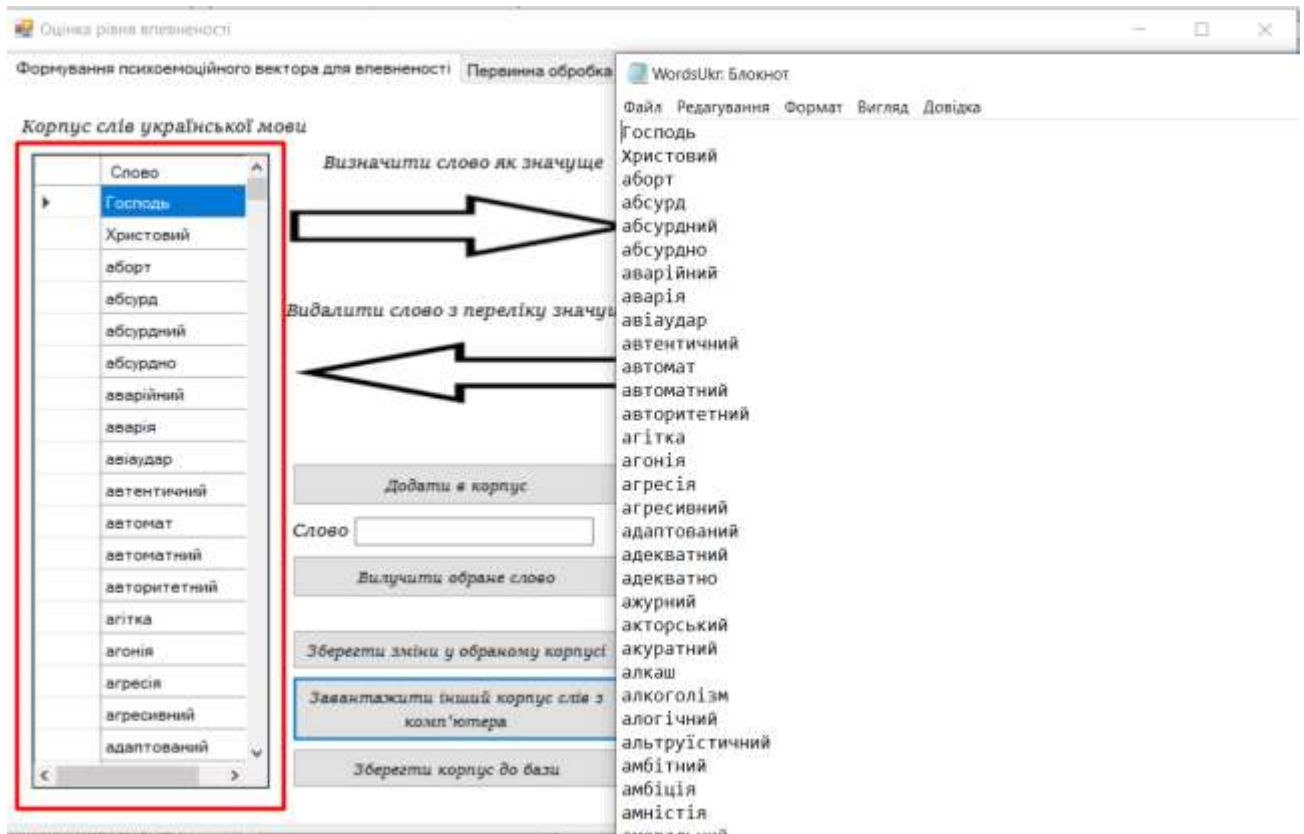


Рисунок 3.3 – Результат виконання функції відкриття нового корпусу

Далі буде розглянуто декілька методів із реалізації вкладки «Первинна обробка тексту». Після завантаження тексту користувачем, з тексту потрібно видаляти стоп-символи, до яких належать перераховані символи: ' ', '.', ',', '!', '?', '/', '-', ')', '(', '[', ']', '—', '—', '»', '«', '#', '=', '"', '"', '@', '+', '0', ':', '1', '2', '3', '4', '5', '6', '7', '8', '9'. Також, окрім перерахованих символів на етапі попередньої обробки всі слова повинні бути переведені до нижнього регістру та потрібно позабирати усі слова з однієї літери. Код методу буде наступним:

```
String text = richTextBox1.Text;

var words = text.Split(new char[] { ' ', '.', ',', '!', '?', '/', '-', ')', '(', '[', ']', '—', '—', '»', '«', '#', '=', '"', '"', '@', '+', '0', ':', '1', '2', '3', '4', '5', '6', '7', '8', '9' })
    .Where(w => !string.IsNullOrEmpty(w))
    .Where(w => w.Length > 2)
    .Select(w => w.ToLower());
richTextBox1.Text = "";

richTextBox1.Text = string.Join(" ", words);
arrOfWords = words;
```

Коректність виконання програмного коду відображена на рисунках 3.4 – 3.5. На рисунку 3.4 показано базовий текст після завантаження, на рисунку 3.5 – текст з видаленими спец-символами та переведений до нижнього регістру.

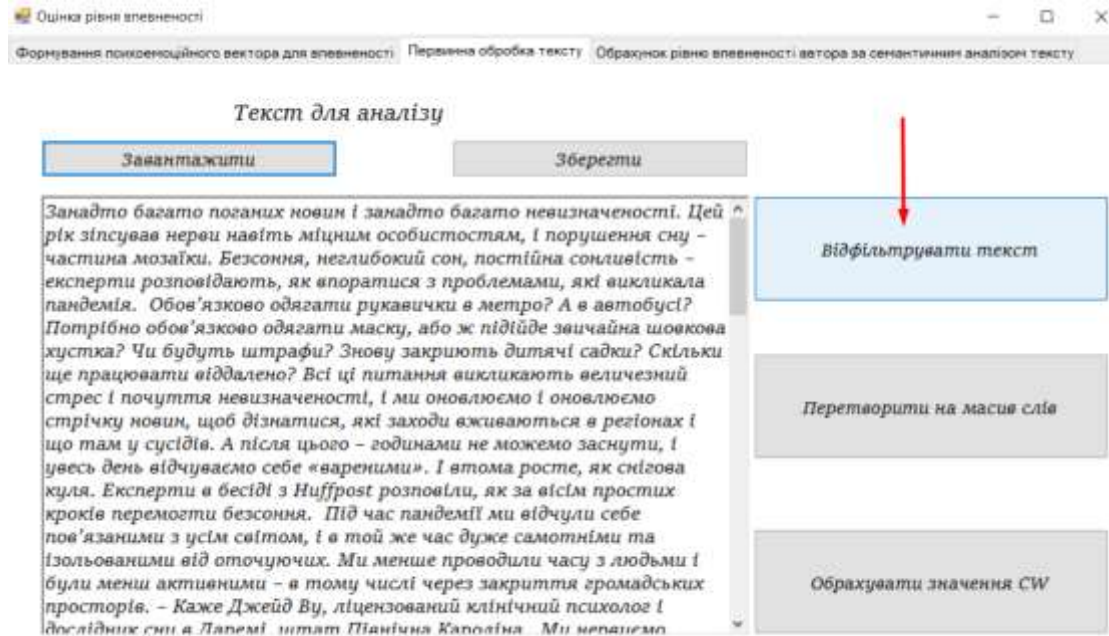


Рисунок 3.4 – Користувацький текст після завантаження

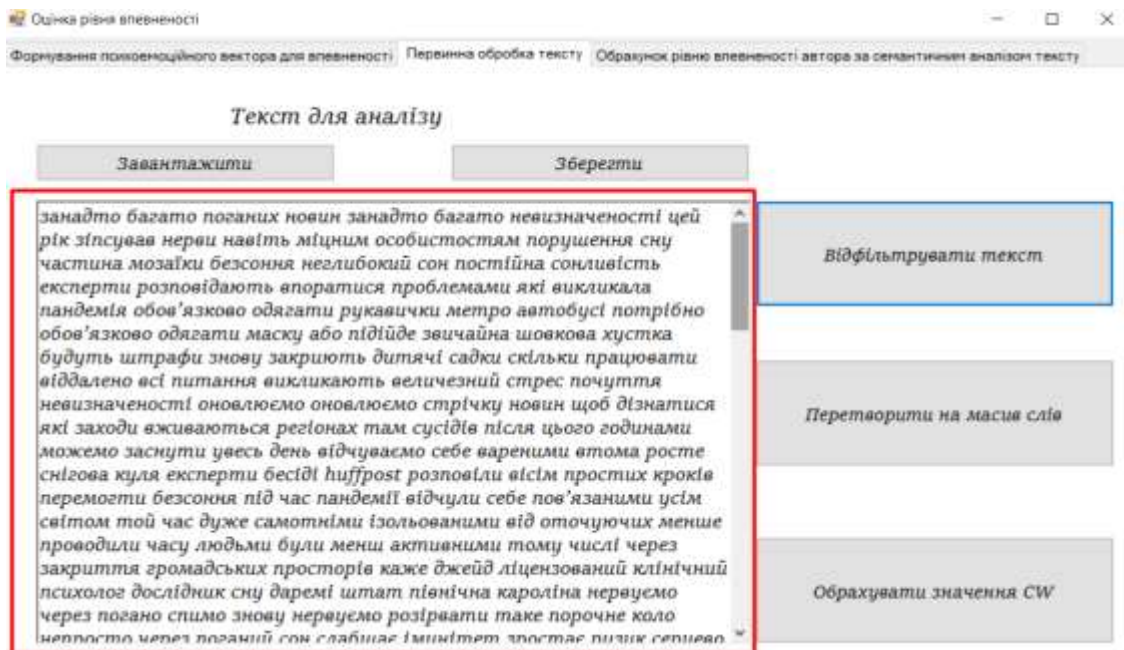


Рисунок 3.5 – Користувацький текст після видалення спец-символів

У подальшому перетворений текст буде переведено у масив та буде проведений обрахунок частоти зустрічання (CW). Фрагмент методу обрахунку частоти зустрічання подано нижче:

```
wordsArr = new Dictionary<string, long>();
    int i = 0;
    foreach (var selectWord in arrOfWords)
    {
        if (!wordsArr.ContainsKey(selectWord))
        {
            wordsArr.Add(selectWord, 1);
        }
        else
        {
            wordsArr[selectWord]++;
        }
        i++;
    }
}
```

Відповідно до методу, формується словник, у якому йде підрахунок частоти зустрічання унікальних слів. Для ілюстрації наведено рисунок 3.6

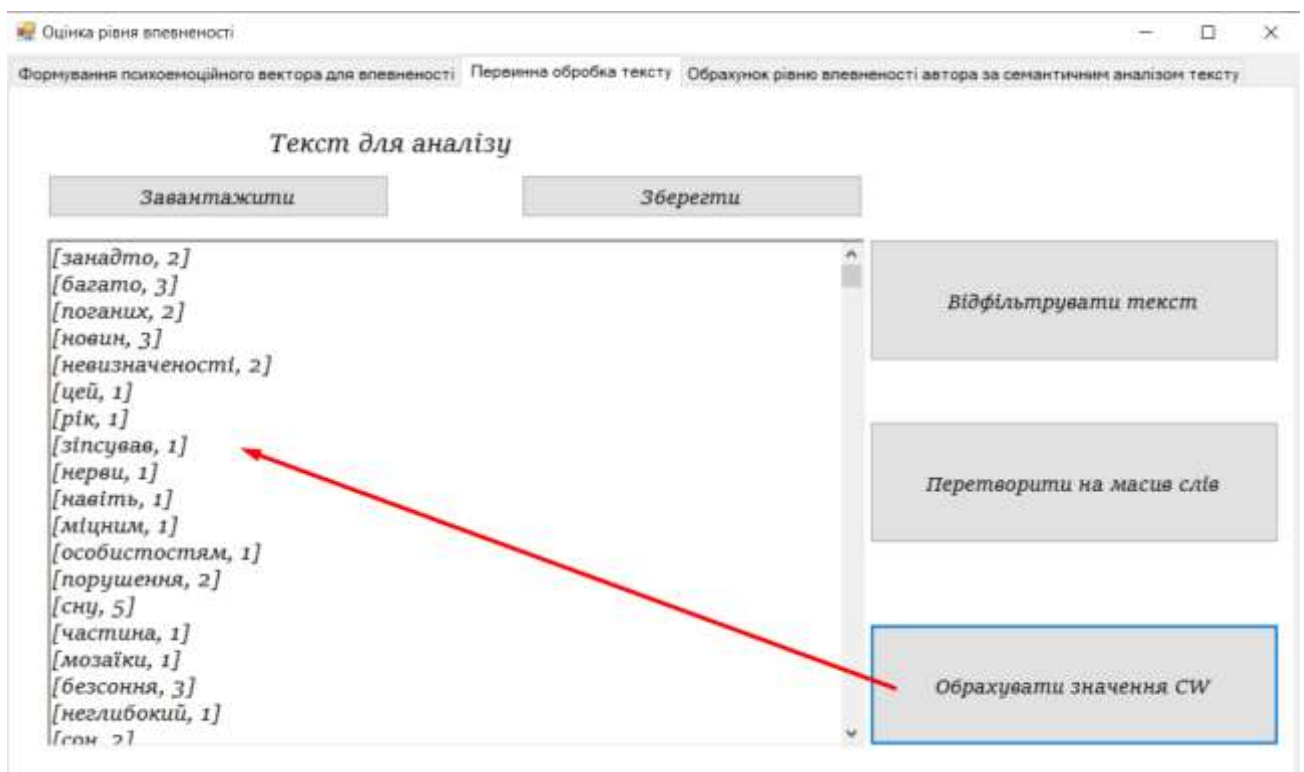


Рисунок 3.6 – Частота зустрічання унікальних слів

Для обрахунку оцінки рівня впевненості автора для кожного значущого слова використовується фрагмент коду:

```
for (int j = 0; j < dataGridView3.RowCount &&
dataGridView3.Rows[j].Cells[0].Value != null; j++)
{
    for (int i = 0; i < pvv.w.Length; i++)
    {
        if
(pvv.w[i].word.Equals(dataGridView3.Rows[j].Cells[0].Value.ToString()))
        {
            dataGridView3.Rows[j].Cells[2].Value = ((pvv.w[i].emotionalMark *
Double.Parse(dataGridView3.Rows[j].Cells[1].Value.ToString()))/
dataGridView3.RowCount).ToString());
        }
    }
}
```

Відображення результату наведено у таблиці (рисунок 3.7).

**Обрахунок рівню впевненості автора за семантичним аналізом тексту**

Слова	CW (кількість зустрічань)	VW (оціночне значення впевненості)
новин	3	
невизначеності	2	-0.0291970802919708
цей	1	
рік	1	
зіпсував	1	-0.0145985401459854
перепи	1	
навіть	1	
міцним	1	

Значення оцінки рівню впевненості автора:

Градуси	Значення
-1..-0.6	Надвисокий рівень непевненості
-0.59..-0.3	Високий рівень непевненості
-0.3..-0.1	Середній рівень непевненості
-0.09..0	Низький рівень непевненості
0.01..0.1	Низький рівень впевненості
0.11..0.3	Середній рівень впевненості
0.31..0.59	Високий рівень непевненості

Значення оцінки рівню емоційної напруженості:

Градуси	Значення
0.00 - 0.1	Низький рівень емоційної напруженості
0.11 - 0.3	Середній рівень емоційної напруженості
0.31 - 0.59	Високий рівень емоційної напруженості
0.6 - 1	Надвисокий рівень емоційної напруженості

Рисунок 3.7 – Оціночне значення впевненості

Таким чином, було проілюстровано деталі реалізації програмних складових інформаційної системи автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту.

### 3.3 Тестування інформаційної системи

Для дослідження працездатності та коректної роботи розроблених функцій було проведено тестування інформаційної системи автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту. Тестування проводилось методом створення тестових таблиць (тест-кейсів) та шляхом функціонального тестування програмного забезпечення.

Для тестування програмного забезпечення було розроблено декілька тест-кейсів. Перший тестовий випадок – перевірка на додавання слова у вектор впевненості автора (таблиця 3.1).

Таблиця 3.1 – Тест-кейс TS0001

<b>Тест-кейс ID:</b> TS 0001	<b>Приоритет:</b> 1	<b>Створено:</b> 6.05.2022, Клекач С.С.
<b>Назва:</b> Перевірка правильності додавання слова до вектору впевненості автора.		
<b>Вхідні дані:</b> Додати слово «колись» із базового корпусу української мови		
<b>Кроки</b>		<b>Очікуваний результат</b>
<ol style="list-style-type: none"> <li>1. Запустити програму</li> <li>2. На першій вкладці прокрутити таблицю зі словами з корпусу української мови до слова «колись» та натиснути на ньому мишкою.</li> <li>3. Натиснути кнопку «Визначити слово як значуще»</li> <li>4. Перевірити наявність внесеного слова у векторі впевненості автора.</li> </ol>		Додане слово відображено вкінці вектору впевненості
<b>Результат виконання тест-кейсу:</b> перевірку пройдено успішно.		

Після запуску програми потрібно виконати кроки, вказані у таблиці 3.1. Після чого у програмі користувач побачить своє додане слово в кінці вектору впевненості (рисунок 3.8).

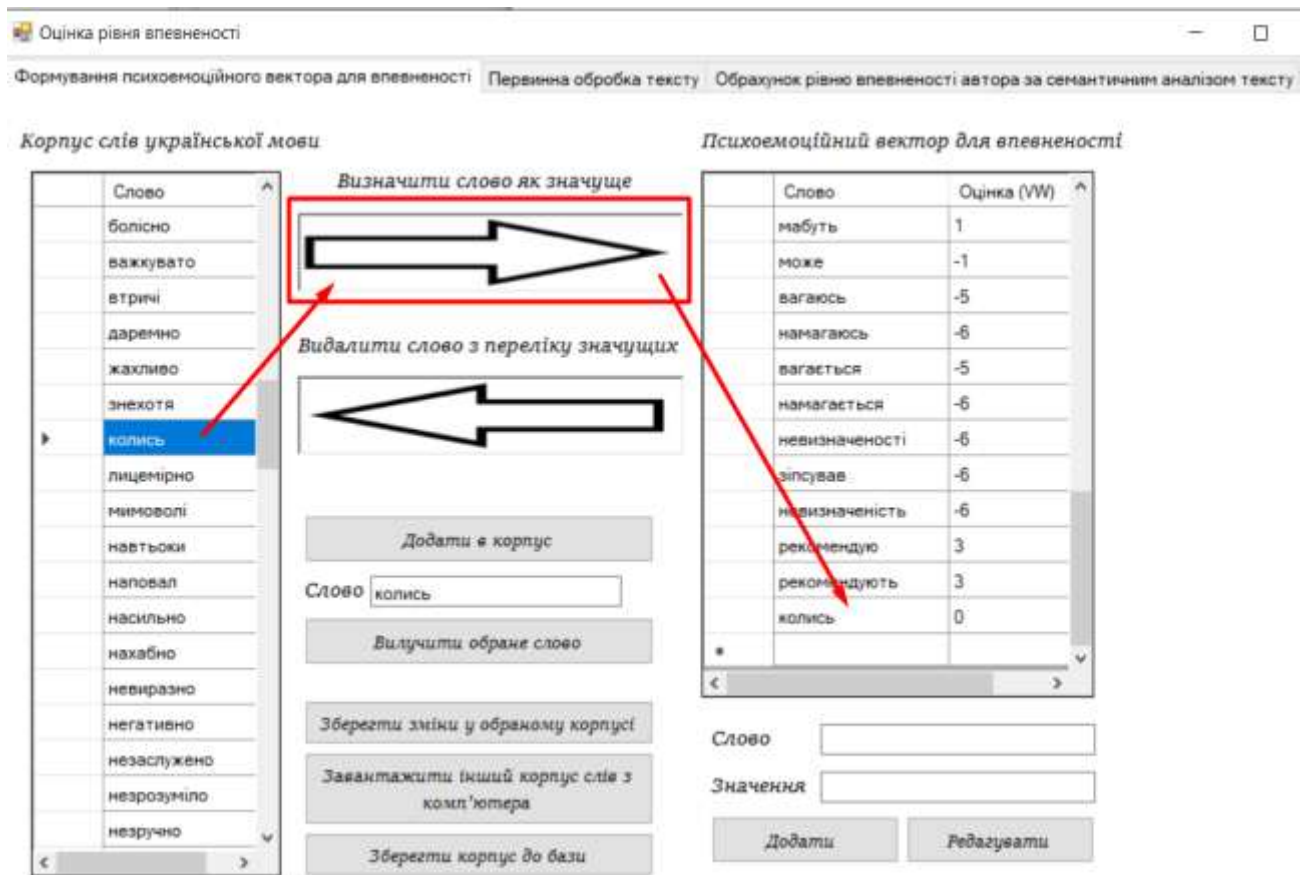


Рисунок 3.8 – Перевірка додавання слова з корпусу слів української мови у психоемоційний вектор впевненості

Як видно з рисунку 3.8, тест виконано успішно. Наступним тестовим випадком досліджується зміна оцінки слова для психоемоційного вектору для впевненості, доданого на попередньому тесті. Алгоритм проведення тестування наведено у таблиці 3.2.

Таблиця 3.2 – Тест-кейс TS0002

<b>Тест-кейс ID:</b> TS 0002	<b>Пріоритет:</b> 1	<b>Створено:</b> 6.05.2022, Клекач С.С.
<b>Назва:</b> Зміна оцінки слова для психоемоційного вектору для впевненості. <b>Вхідні дані:</b> Для обраного слова «колись» із психоемоційного вектору для впевненості задається значення -2		
<b>Кроки</b>		<b>Очікуваний результат</b>
1. Запустити програму 2. На першій вкладці прокрутити таблицю зі словами з психоемоційного вектору для впевненості до слова «колись» та натиснути на ньому мишкою. 3. У полі для редагування оцінки замінити 0 на -2. 4. Перевірити наявність внесених змін у векторі впевненості автора.		У слова «колись» змінена оцінка відображається в таблиці
<b>Результат виконання тест-кейсу:</b> перевірку пройдено успішно.		

Після запуску програми потрібно виконати кроки, вказані у таблиці 3.2. Після чого у програмі користувач побачить зміни у додане раніше слово в кінці вектору впевненості (рисунок 3.9).

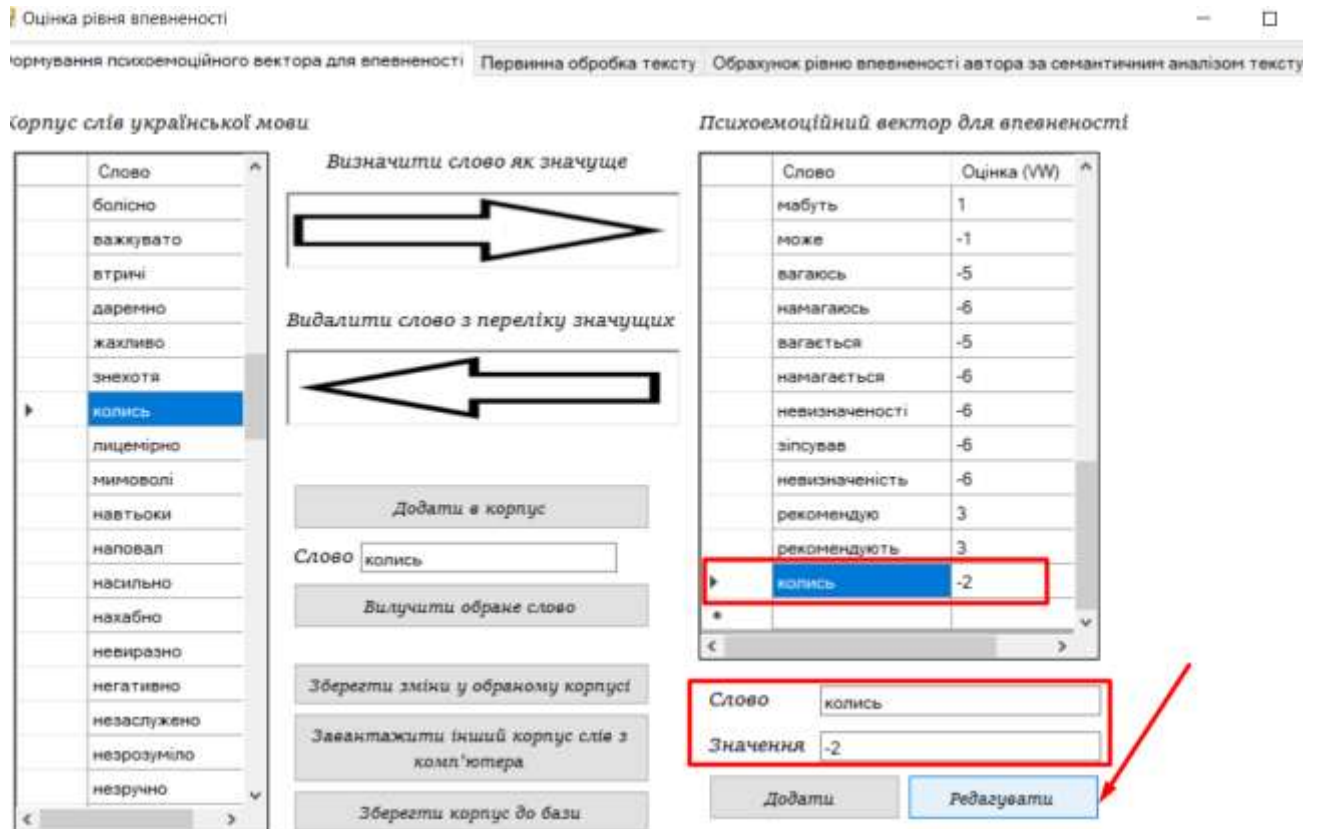


Рисунок 3.9 – Перевірка редагування слова з психоемоційного вектору для впевненості

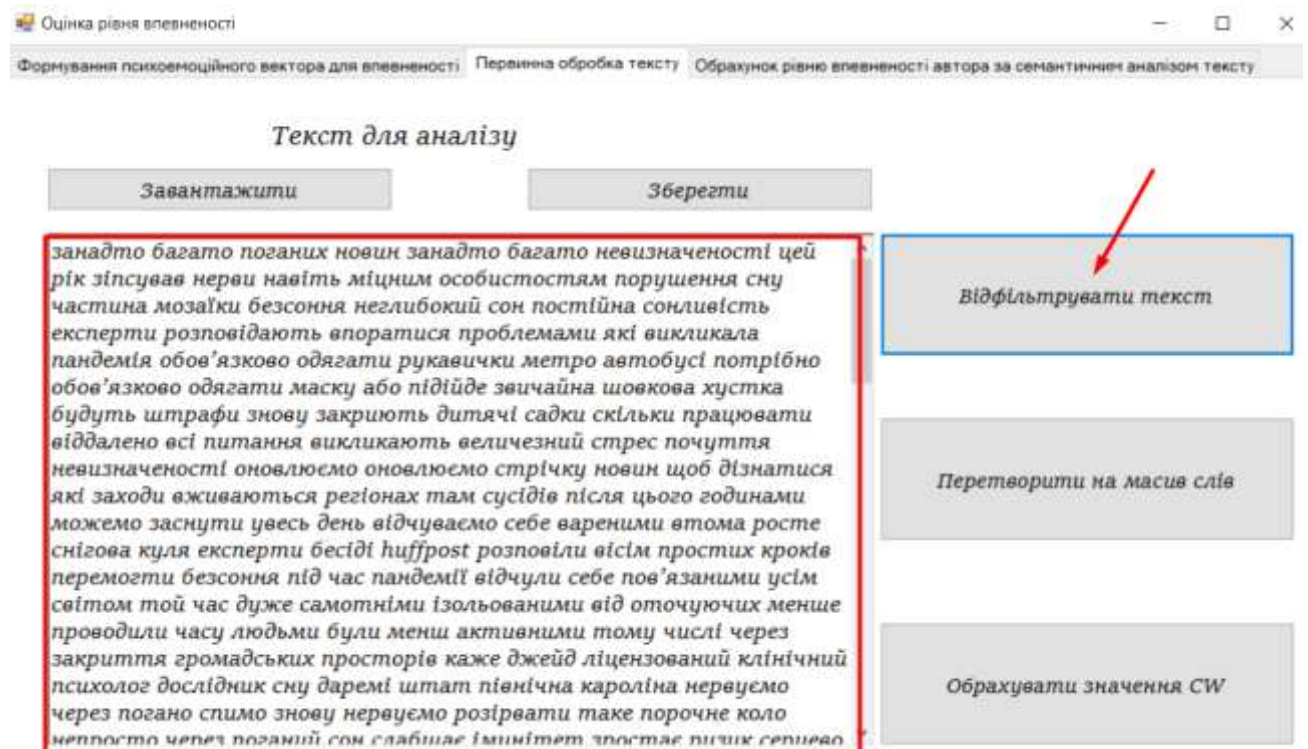


Рисунок 3.10 – Очистка тексту від стоп-символів та переведення слів у нижній регістр

Як показано на рисунку 3.9 – тест пройдено успішно, внесені зміни відображаються у таблиці. Наступним тестовим випадком буде очистка тексту від стоп-символів та переведення слів у нижній регістр. Кроки проходження тесту відображені у таблиці 3.3.

Таблиця 3.3 – Тест-кейс TS0003

<b>Тест-кейс ID:</b> TS 0003	<b>Пріоритет:</b> 1	<b>Створено:</b> 7.05.2022, Клекач С.С.
<b>Назва:</b> Очистка тексту від стоп-символів та переведення слів у нижній регістр.		
<b>Вхідні дані:</b> Завантажити текст для дослідження		
<b>Кроки</b>		<b>Очікуваний результат</b>
<ol style="list-style-type: none"> <li>1. Запустити програму.</li> <li>2. Перейти на вкладку «Первинна обробка тексту».</li> <li>3. Натиснути на кнопку «Завантажити».</li> <li>4. Обрати текст для завантаження, натиснути кнопку «Відкрити».</li> <li>5. Після відображення тексту на екрані користувача натиснути на кнопку «Відфільтрувати текст».</li> </ol>		У текстовому полі відобразиться текст весь у нижньому регістрі без стоп-символів.
<b>Результат виконання тест-кейсу:</b> перевірку пройдено успішно.		

Після запуску програми потрібно виконати кроки, вказані у таблиці 3.3. Після чого у програмі користувач побачить почищений текст у нижньому регістрі (рисунок 3.10).

Наступним тестовим випадком буде обрахунок значень кількості зустрічань унікальних слів (CW). Кроки проходження тесту відображені у таблиці 3.4.

Таблиця 3.4 – Тест-кейс TS0004

<b>Тест-кейс ID:</b> TS 0004	<b>Пріоритет:</b> 1	<b>Створено:</b> 7.05.2022, Клекач С.С.
<b>Назва:</b> Обрахунок значень кількості зустрічань унікальних слів (CW)		
<b>Вхідні дані:</b> Завантажити текст для дослідження		
<b>Кроки</b>		<b>Очікуваний результат</b>
<ol style="list-style-type: none"> <li>1. Запустити програму.</li> <li>2. Перейти на вкладку «Первинна обробка тексту».</li> <li>3. Натиснути на кнопку «Завантажити».</li> <li>4. Обрати текст для завантаження, натиснути кнопку «Відкрити».</li> <li>5. Після відображення тексту на екрані користувача натиснути на кнопку «Відфільтрувати текст».</li> <li>6. Для відфільтрованого тексту натиснути на кнопку «Перетворити на масив слів».</li> <li>7. Натиснути на кнопку «Обрахувати значення CW».</li> </ol>		У текстовому полі відобразиться текст весь у нижньому регістрі без стоп-символів.
<b>Результат виконання тест-кейсу:</b> перевірку пройдено успішно.		

Після запуску програми потрібно виконати кроки, вказані у таблиці 3.4. Після чого у програмі користувач побачить всі унікальні слова у нижньому регістрі з частотою їх зустрічання (рисунок 3.11).

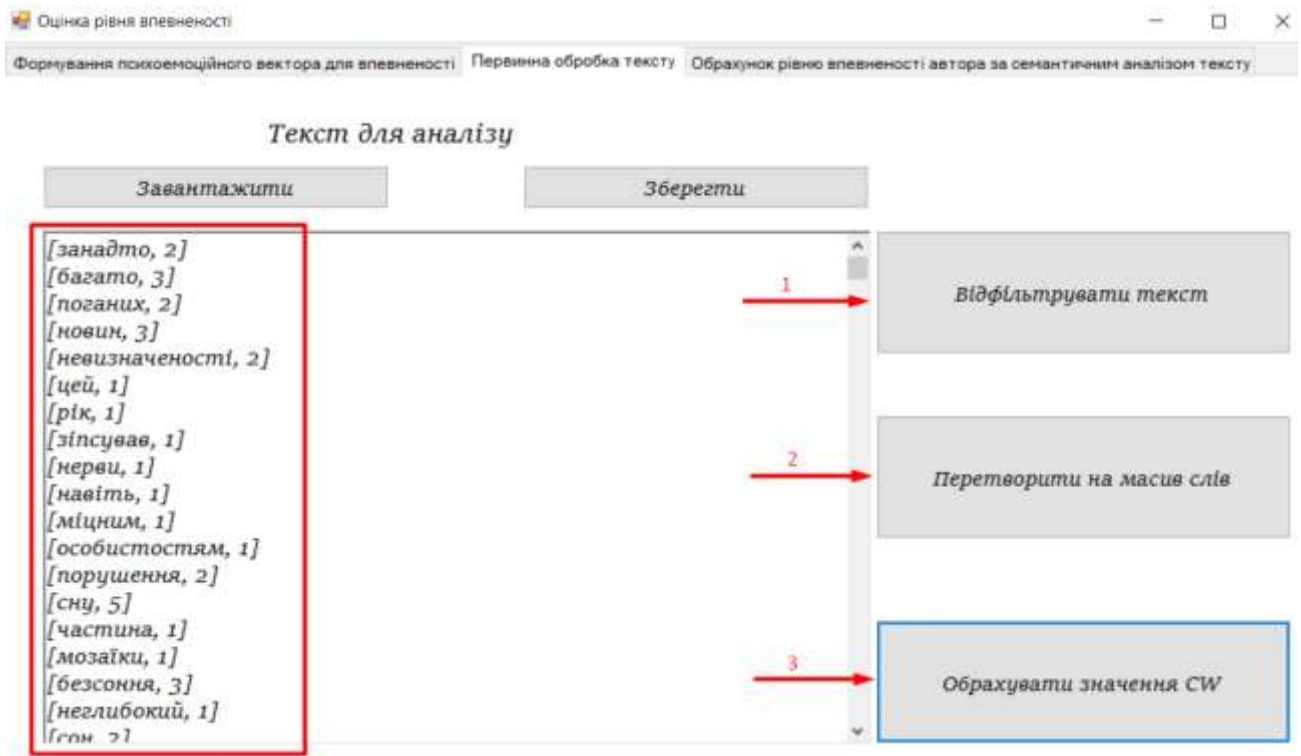


Рисунок 3.11 – Підрахунок унікальних слів тексту

Отже, у процесі виконаного тестування програмного продукту на основі реалізації методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту некоректно працюючих функцій виявлено не було.

### 3.4 Інструкція користувача

Для зручності використання програмного засобу реалізації методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту було створено інструкцію користувача. При запуску застосування користувач бачить стартовий екран, де є завантажені базовий корпус слів української мови та психоемоційний вектор для впевненості (рисунок 3.12).

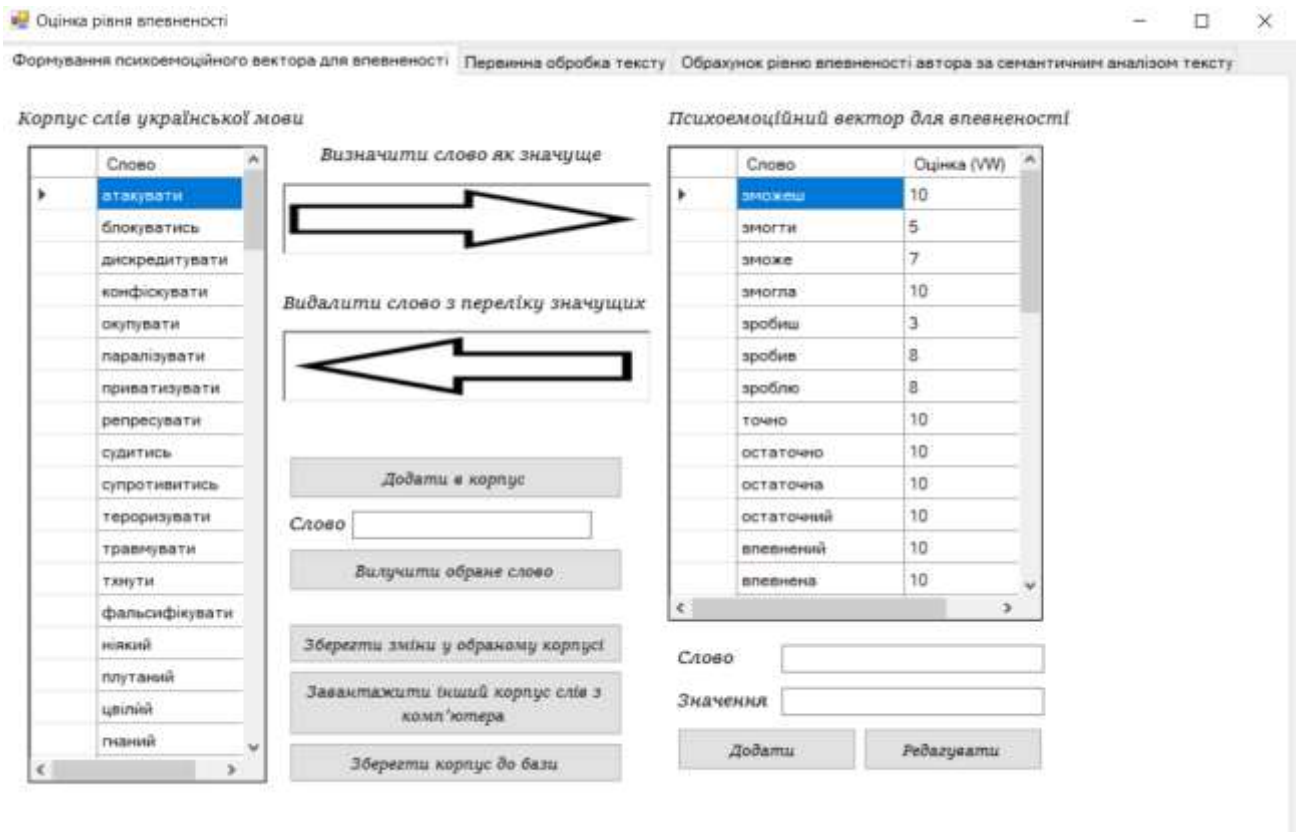


Рисунок 3.11 – Стартовий екран застосування

На першій вкладці є можливість додати нове слово до корпусу слів української мови, для цього потрібно вписати у поле «Слово» слово, яке потрібно додати до корпусу на натиснути кнопку «Додати в корпус». Введене слово буде додано до корпусу.

Аналогічно до додавання нового слова, є можливість видаляти слова з корпусу. Для цього потрібно виділити слово, яке користувачу потрібно видалити та натиснути кнопку «Вилучити обране слово» (рисунок 3.12). Обране слово «тест було вилучено».

Оцінка рівня впевненості

Формування психоемоційного вектора для впевненості | Первинна обробка тексту | Обрахунок рівню впевненості автора за семантичним

**Корпус слів української мови**

Слово
неблагополуччя
невігластво
невміння
непорозуміння
неприйняття
нерозуміння
ниття
обвалення
обурення
озброєння
отруєння
падіння
перевищення
передозування
переповнення
п'яцтво
<b>тестове</b>

**Психоемоційний вектор для впевненості**

Слово	Оцінка (VW)
зможеш	10
зможти	5
зможе	7
змогла	10
зробиш	3
зробив	8
зроблю	8
точно	10
остаточно	10
остаточна	10
остаточний	10
впевнений	10
впевнена	10

Слово

Значення

**Визначити слово як значуще**

**Видалити слово з переліку значущих**

**Додати в корпус**

Слово

Рисунок 3.12 – Додавання слова в корпус

**Корпус слів української мови**

Слово
небажання
неблагополуччя
невігластво
невміння
непорозуміння
неприйняття
нерозуміння
ниття
обвалення
обурення
озброєння
отруєння
падіння
перевищення
передозування
переповнення
<b>п'яцтво</b>

**Визначити слово як значуще**

**Видалити слово з переліку значущих**

Слово

Рисунок 3.13 – Вилучення слова з корпусу

Також є можливість завантажити інший корпус слів для аналізу, натиснувши на кнопку «Завантажити інший корпус слів з комп'ютера». У діалоговому вікні потрібно обрати корпус, який користувачу потрібно завантажити та натиснути у діалоговому вікні «Відкрити» (рисунк 3.14).

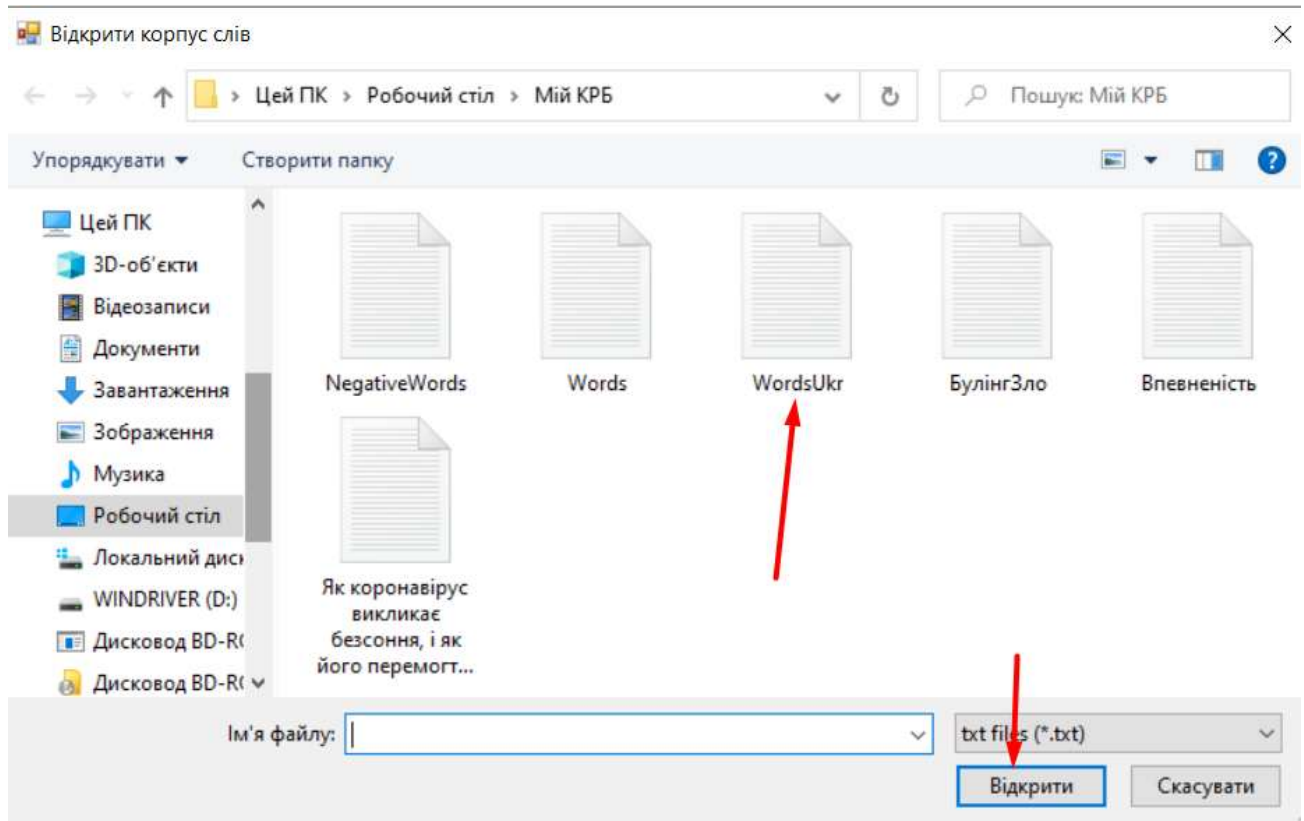


Рисунок 3.14 – Завантаження нового

Є можливість роботи з корпусом (окрім додавання нового слова, можна змінювати існуючі). Для цього потрібно курсор миші поставити на слово яке потрібно редагувати та натиснути на кнопку «Зберегти зміни». Також можна зберегти завантажений з комп'ютера корпус до бази даних, натиснувши кнопку «Зберегти корпус до бази». Після збереження корпусу користувач побачить повідомлення про те що корпус збережено (рисунк 3.15).

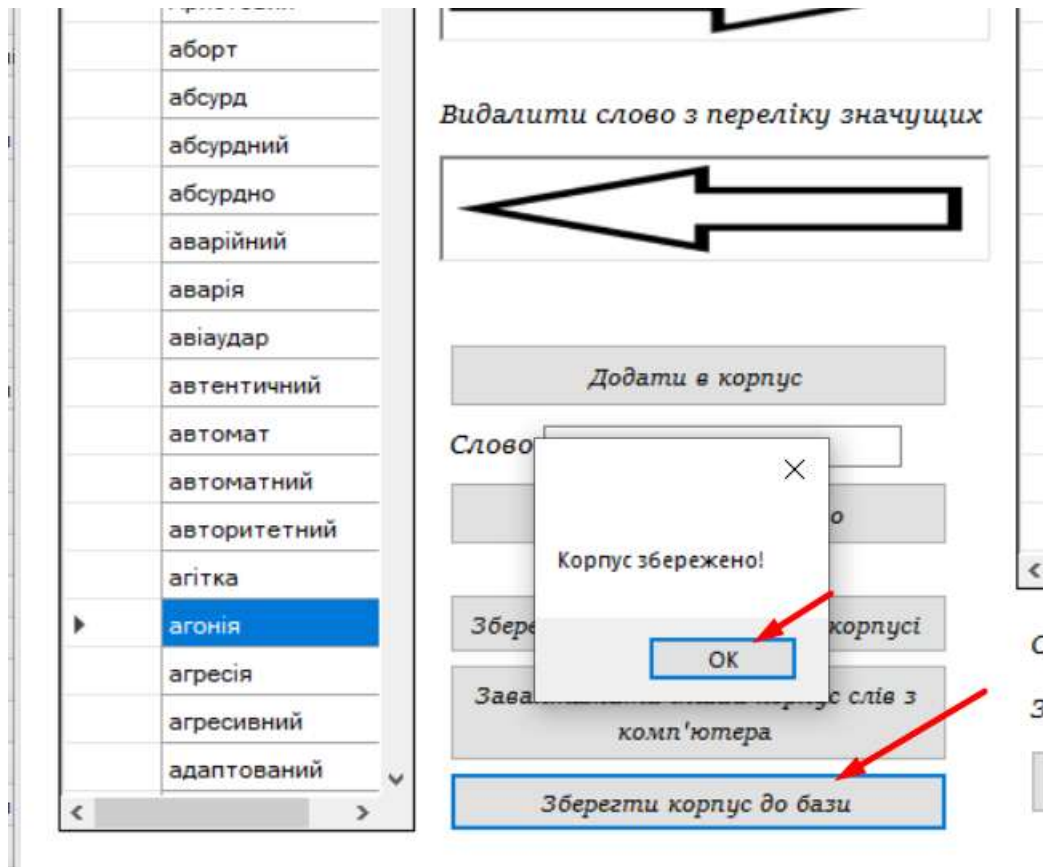


Рисунок 3.15 – Збереження корпусу до бази

Аналогічним чином представлена робота з психоемоційним вектором для впевненості. Для редагування вже існуючих у векторі слів потрібно мишею натиснути на слово, яке потрібно редагувати. Відредаговано може бути як саме слово, так і його числове значення оцінки (рисунок 3.16).

Також можна додати слово до психоемоційного вектора впевненості, вписавши у поля «Слово» та «Значення» відповідні характеристики. Після уведення слова та значення потрібно натиснути на кнопку «Додати».

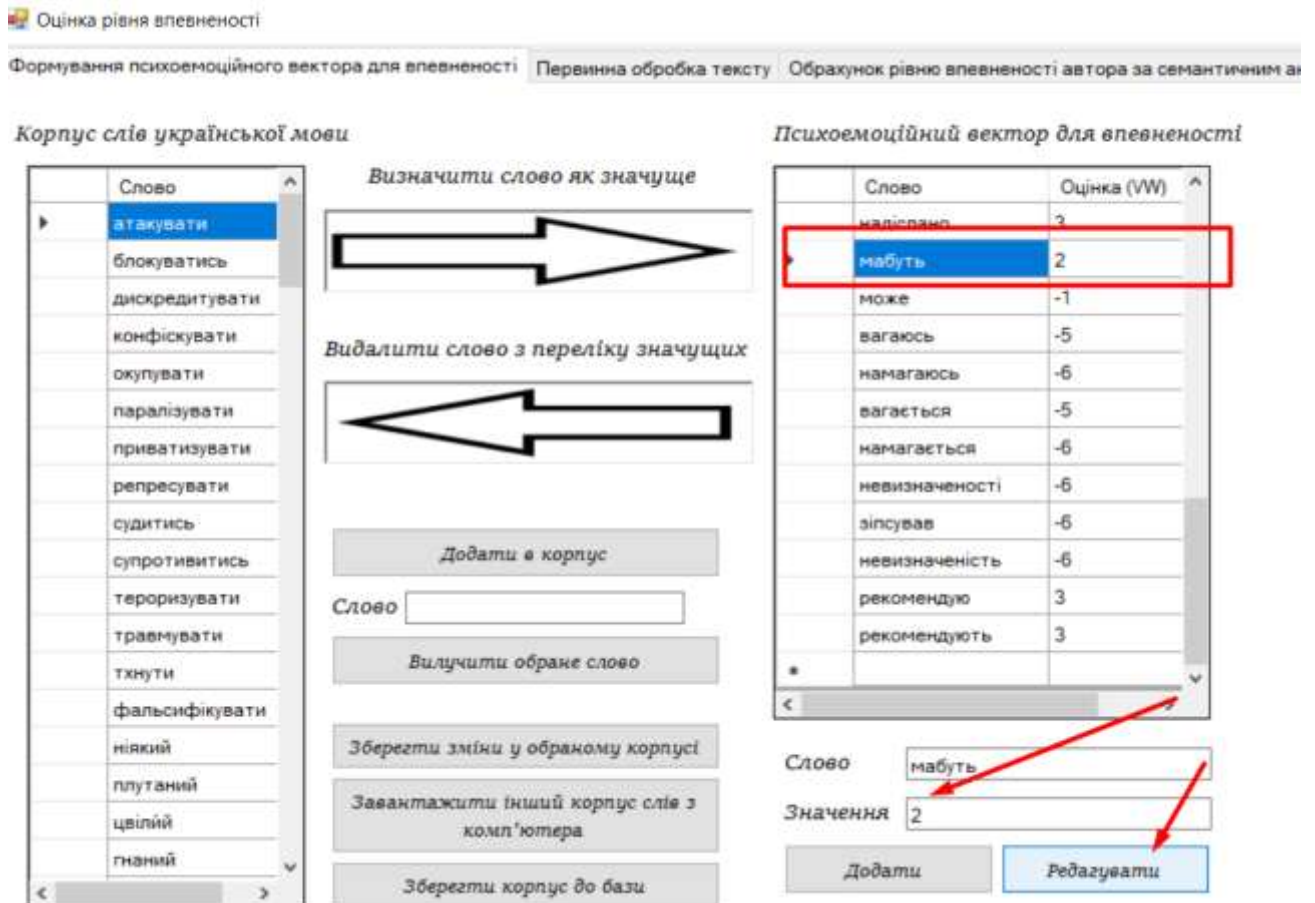


Рисунок 3.16 – Редагування оцінок психоемоційного вектора для впевненості

Друга вкладка «Первинна обробка тексту» представляє собою вкладку для обробки тексту. Текст для обробки можна або вписати вручну, або зчитати з файлу, або скопіювати у поле «Текст для аналізу». Для завантаження з файлу потрібно натиснути на кнопку «Завантажити», після чого відкриється діалогове вікно для завантаження тексту з провідника (рисунок 3.17). Для підтвердження відкриття обраного тексту потрібно натиснути кнопку «Відкрити».

Одразу після натискання кнопки «Відкрити», на екрані буде відображено текст файлу (рисунок 3.18).

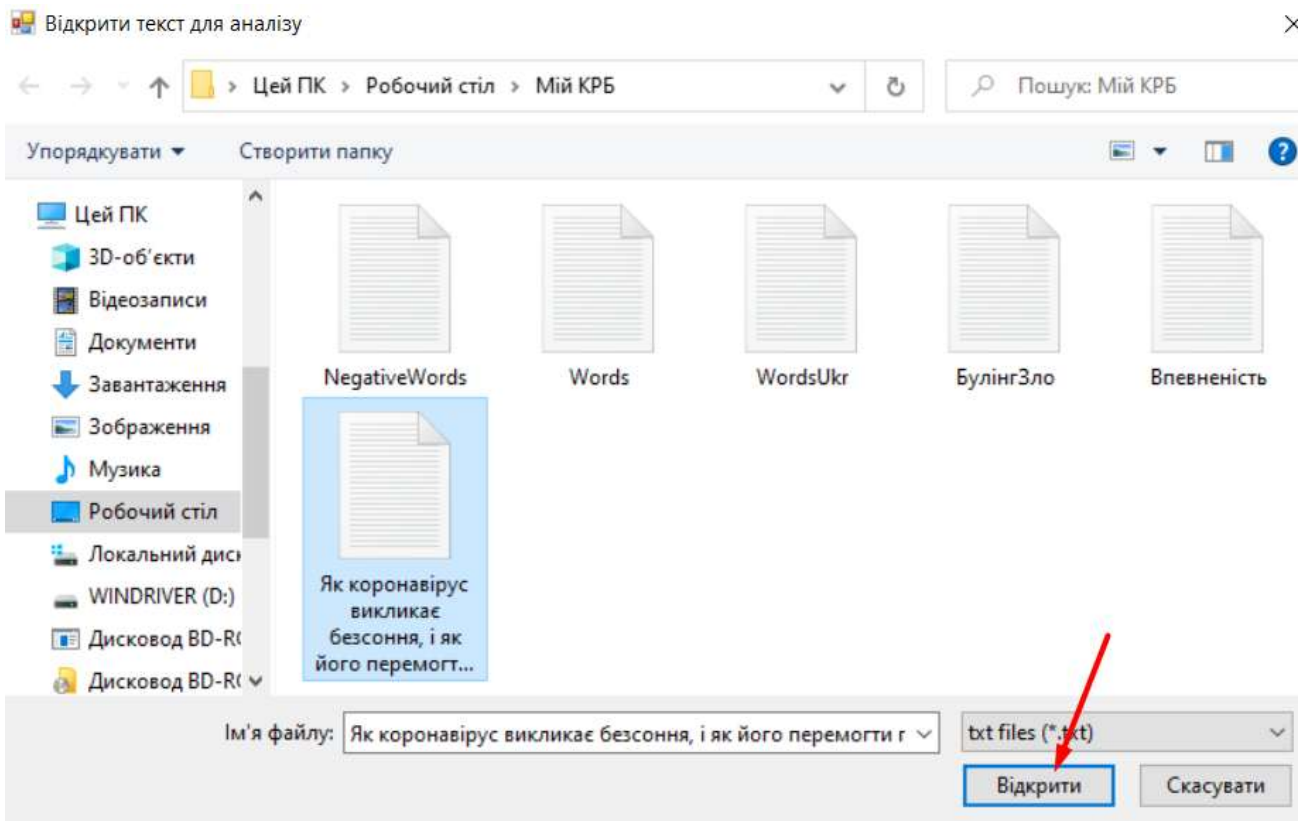


Рисунок 3.17 – Завантаження тексту для аналізу

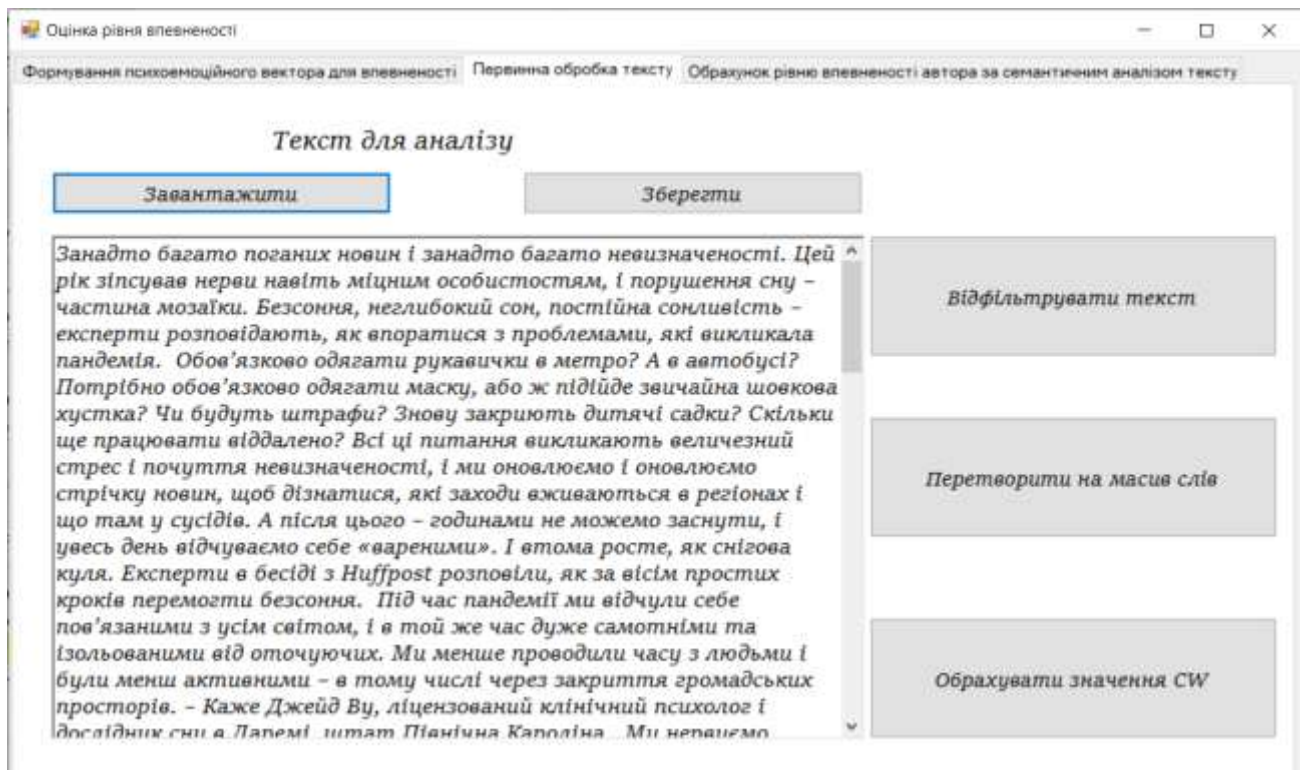


Рисунок 3.18 – Виведення тексту з файлу

Для фільтрації тексту (прибирання спец-символів, переведення слів до нижнього регістру) потрібно натиснути кнопку «Відфільтрувати текст». Після чого текст буде відфільтрований, та виведений у це саме поле «Текст для аналізу» (рисунок 3.19).

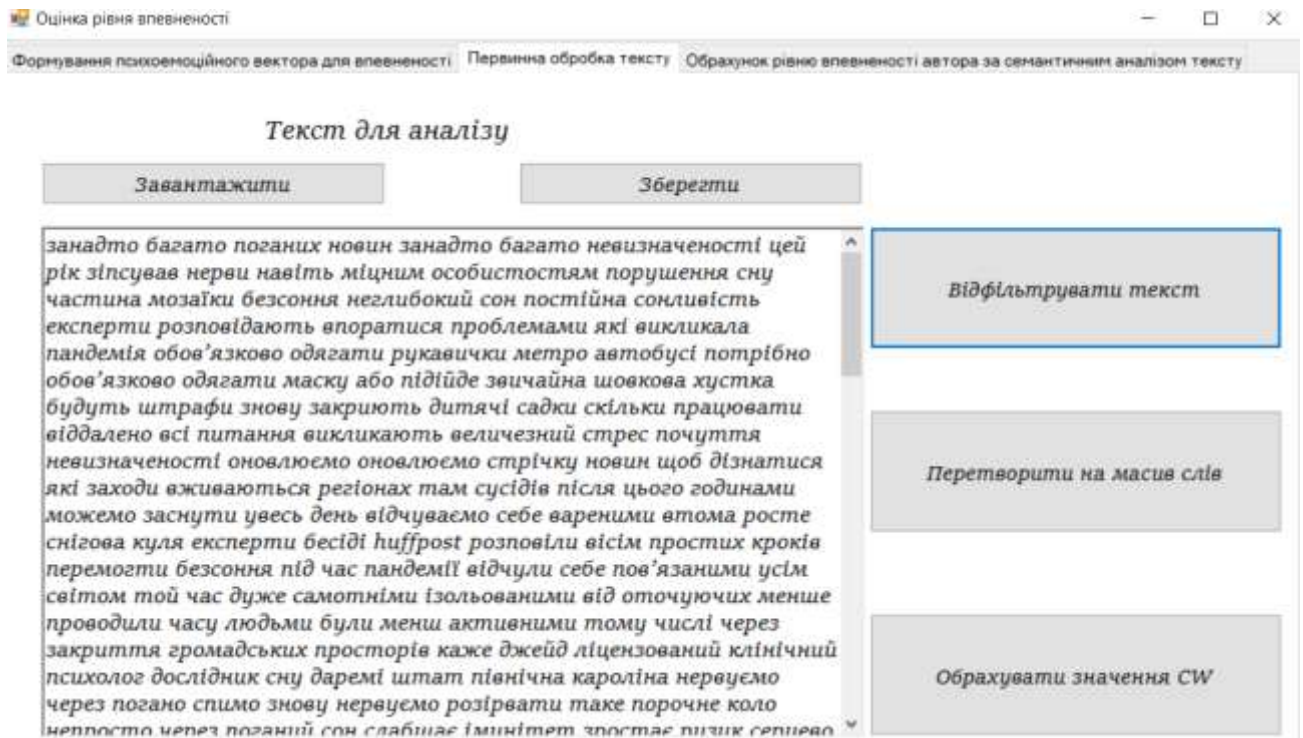


Рисунок 3.19 – Відфільтрований текст

Для представлення тексту у вигляді масиву символів потрібно натиснути кнопку «Перетворити на масив слів». Результат зображено на рисунку 3.20.

Для обрахування частоти кожного із слів тексту потрібно натиснути на кнопку «Обрахувати значення CW». Після чого всі частоти з'являться разом зі словами тексту (рисунок 3.21).

На третій вкладці «Обрахунок рівню впевненості автора за семантичним аналізом тексту» реалізовано виведення висновків по рівню впевненості автора за семантичним аналізом. Тут користувач може подивитись слова із їх частотою зустрічання, для перегляду яких потрібно натиснути кнопку «Показати слова з CW» (рисунок 3.22).

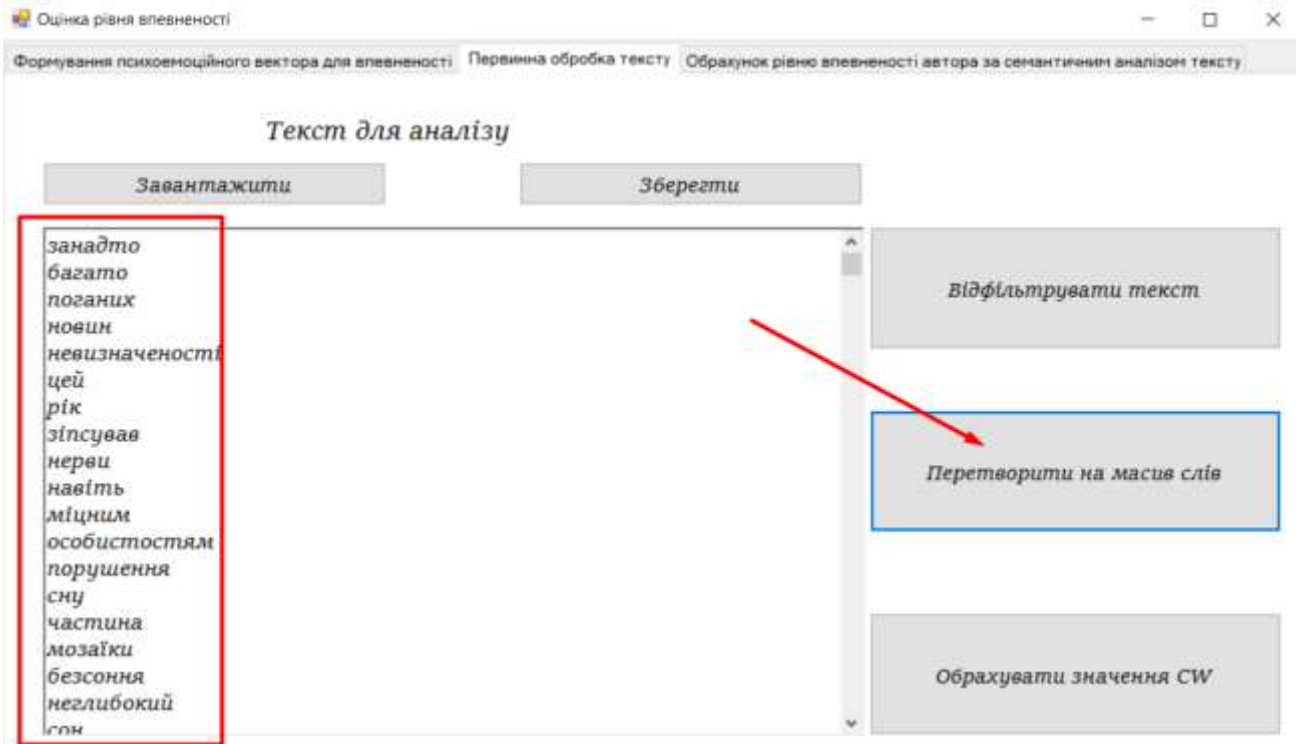


Рисунок 3.20 – Перетворення на масив символів

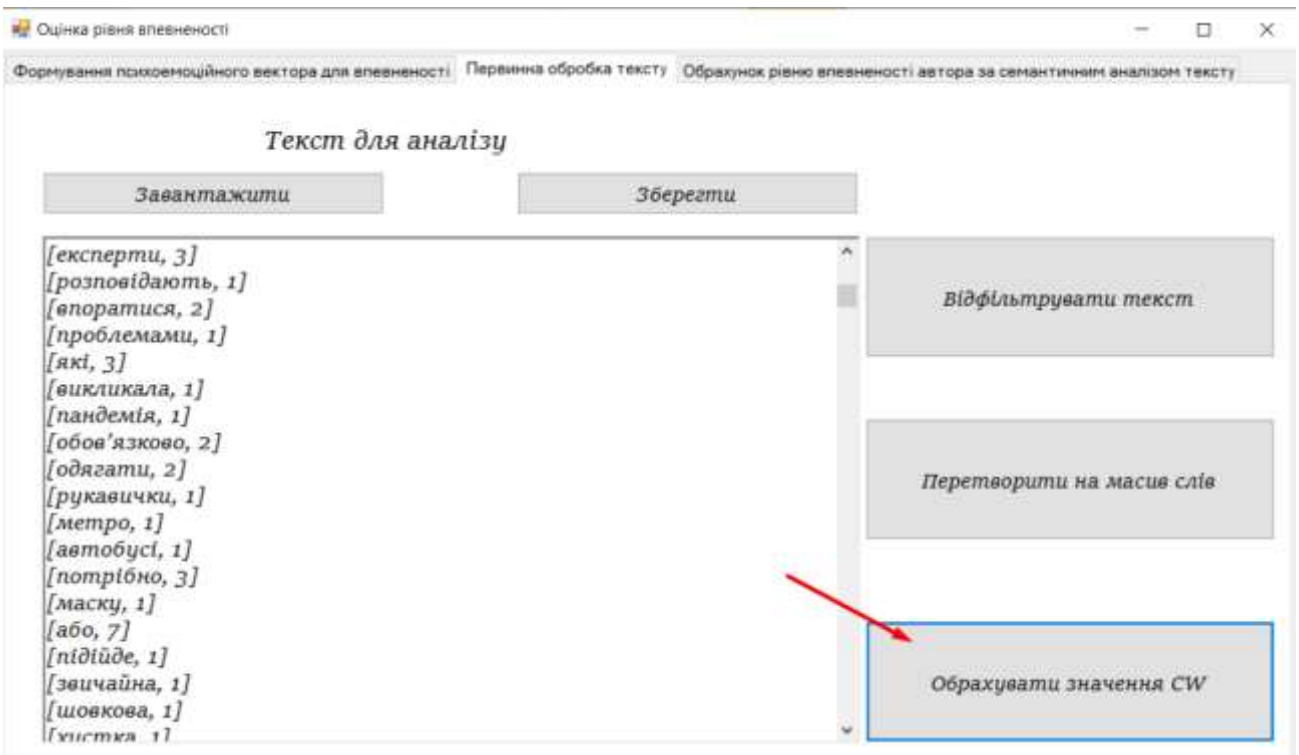


Рисунок 3.21 – Частота зустрічання слів тексту

Оцінка рівня впевненості

Формування психоемоційного вектора для впевненості | Первинна обробка тексту | **Обрахунок рівню впевненості автора за семантичним аналізом тексту**

### Обрахунок рівню впевненості автора за семантичним аналізом тексту

Слова	CW (кількість зустрічань)	VW (оціночне значення впевненості)
занадто	2	
багато	3	
поганих	2	
новин	3	
невизначеності	2	
цей	1	
рік	1	
зіпсував	1	

Показати слова з CW

Обрахунок оцінки рівню впевненості автора

Видалити незначущі слова

Обрахунок оцінки рівню емоційної напруженості

Значення оцінки рівню впевненості автора: Значення оцінки рівню емоційної напруженості

Градуси	Значення
-1...-0,6	Надвисокий рівень непевненості
-0,59...-0,3	Високий рівень непевненості
-0,3...-0,1	Середній рівень непевненості
-0,09...0	Низький рівень непевненості
0,01...0,1	Низький рівень впевненості
0,11...0,3	Середній рівень впевненості
0,31...0,59	Високий рівень непевненості
0,6...1	Надвисокий рівень непевненості

Градуси	Значення
0,00 - 0,1	Низький рівень емоційної напруженості
0,11 - 0,3	Середній рівень емоційної напруженості
0,31 - 0,59	Високий рівень емоційної напруженості
0,6 - 1	Надвисокий рівень емоційної напруженості

Рисунок 3.22 – Відображення слів з CW

Для обрахунку оцінки рівню впевненості потрібно натиснути на кнопку «Обрахунок оцінки рівню впевненості автора», після чого у третій колонці таблиці будуть пораховані усі значення, для яких слова в тексті належать вектору впевненості. Результат проілюстровано на рисунку 3.23.

Оцінка рівня впевненості

Формування психоемоційного вектора для впевненості | Первинна обробка тексту | **Обрахунок рівню впевненості автора за семантичним аналізом тексту**

### Обрахунок рівню впевненості автора за семантичним аналізом тексту

Слова	CW (кількість зустрічань)	VW (оціночне значення впевненості)
занадто	2	
багато	3	
поганих	2	
новин	3	
невизначеності	2	-0.0291970802919708
цей	1	
рік	1	
зіпсував	1	-0.0145985401459854

Показати слова з CW

**Обрахунок оцінки рівню впевненості автора**

Видалити незначущі слова

Обрахунок оцінки рівню емоційної напруженості

Значення оцінки рівню впевненості автора: Значення оцінки рівню емоційної напруженості

Рисунок 3.23 – Відображення слів з CW

Також доступним функціоналом для користувача є видалення усіх незначущих слів, де немає оцінки. Для цього потрібно натиснути на кнопку «Видалити незначущі слова». Результат проілюстровано на рисунку 3.24.

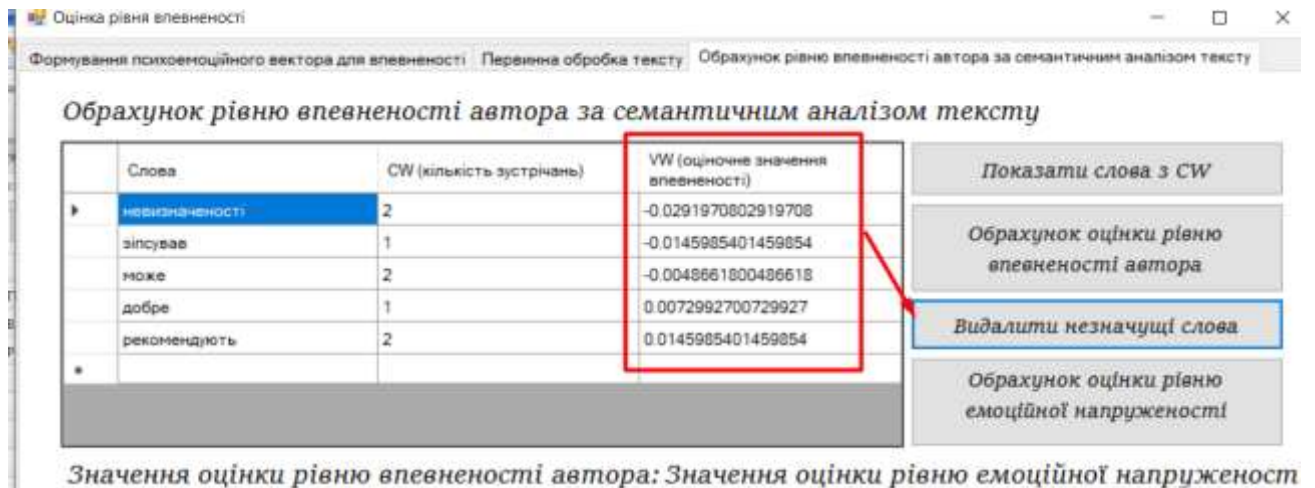


Рисунок 3.24 – Відображення тільки значущих слів

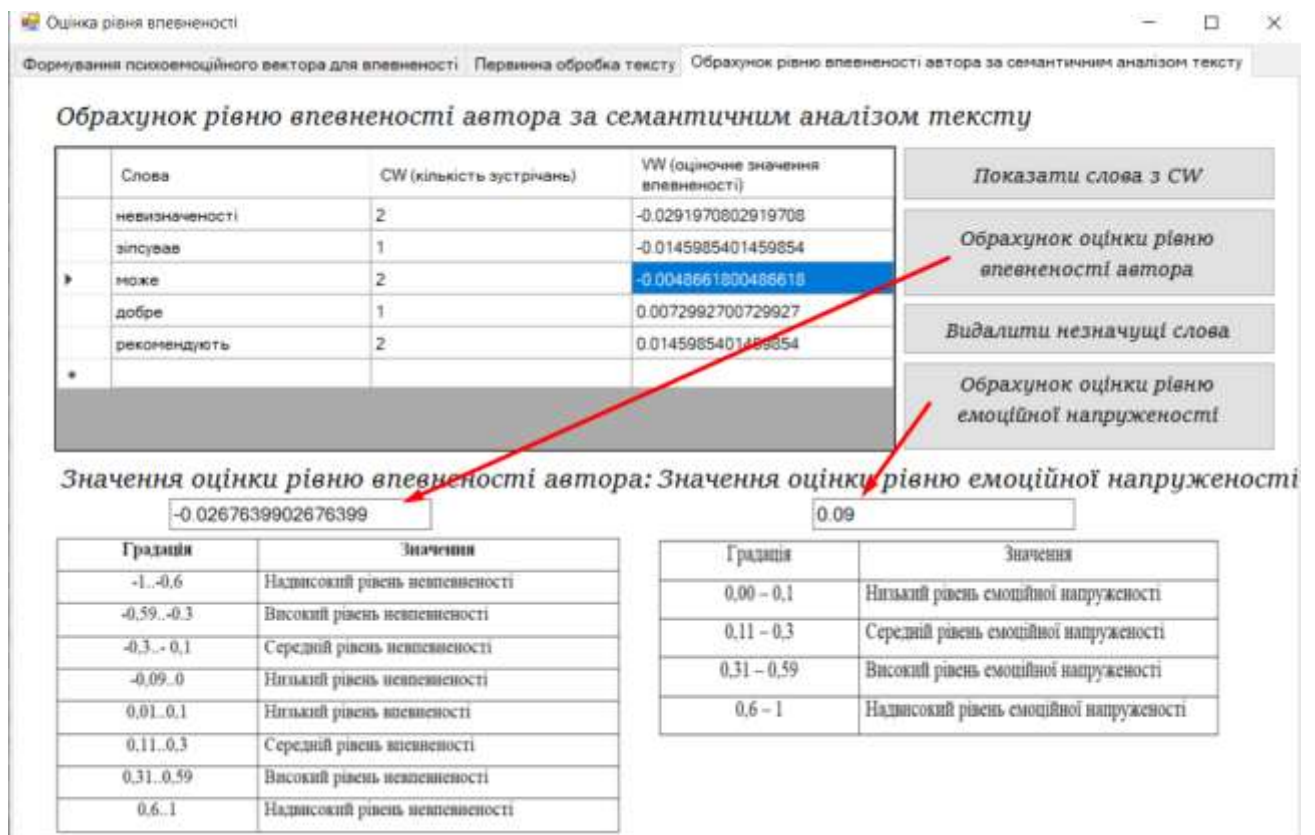


Рисунок 3.25 – Результат обчислень

Також по натисненню кнопки «Обрахунок оцінки рівню впевненості автора» користувачу буде обраховано результат, який можна інтерпретувати за таблицею нижче. По натисненні на кнопку «Обрахунок оцінки рівню емоційної напруженості» користувачу буде виведено оцінку рівню напруженості тексту, яку можна звірити з таблицею розшифровки (рисунок 3.25).

Отже, розроблена інструкція користувача значно спростить використання розробленим застосуванням реалізації методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту.

## Висновки

Розробка методів та засобів для автоматизованого визначення психоемоційних характеристик людини як метаданих до створеного нею текстового контенту дозволяє сприяти вирішенню цілої низки безпекових та критичних питань, як на рівні локальних задач, так і на державному рівні. Розглянуте як приклад автоматизоване визначення рівню впевненості автора за семантичним аналізом текстового контенту є одним із випадків витягу психоемоційних складових автора за наявним цифровим текстом, що дозволяє робити висновок щодо актуальності та достовірності наявних у тексті відомостей.

Для автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту пропонується використовувати психоемоційний вектор слів, який складається із слів та оцінок їх семантичної важливості у психоемоційному розрізі фактора впевненості. Для формування цього вектора пропонується використати наявні у відомих розмічених корпусах текстів української мови і слів відомості, значення оцінок семантичної важливості у психоемоційному розрізі фактора впевненості яких коригується вручну.

У ході виконання КРБ були вирішені такі задачі:

1. Проведено аналіз предметної області
2. Проаналізовано існуючі програмні рішення схожі з тематикою КРБ.
3. Проаналізовано сучасні засоби для створення програмних рішень та обрано комплект засобів для створення власної програмної реалізації.

4. Запропоновано, реалізовано та протестовано метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту, що виконує наступні функції:

– робота з корпусом слів української мови (перегляд, додавання видалення);

– експорт та імпорт слів між корпусом слів української мови та психоемоційного вектору для категорії впевненості;

- робота з елементами психоемоційного вектору для категорії впевненості (перегляд, привласнення оцінок впливу, коригування, додавання слів);

- завантаження чи введення й збереження дослідного тексту для аналізу;
- первинна обробка дослідного тексту (стоп-символи, регістр);
- перетворення дослідного тексту у масив слів;
- обрахунок значень кількостей слів для кожного унікального слова;
- побудова вектору слів тексту й зіставлення кожному унікальному слову як значень кількостей слів так і відповідних значень оцінок впливу;

- видалення незначущих слів тексту, для яких відсутні значення оцінок впливу;

- обрахунок інтегрованих значень добутку кількостей слів впевненості автора на відповідних значень оцінок впливу для кожного значущого слова тексту;

- обрахунок оцінки рівню впевненості автора;
- обрахунок оцінки рівню емоційної напруженості тексту;
- відображення висновку для користувача стосовно рівню впевненості та рівню емоційної напруженості тексту.

5. Для зручності використання створено інструкцію користувача до розробленої програмної реалізації.

Розроблена відповідно до методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту інформаційна система виконує усі прогнозовані функції.

## Перелік посилань

1. Your Article Library. Essay on Man: As a Social Animal. URL: <https://www.yourarticlelibrary.com/man/essay-on-man-as-a-social-animal-1623-words/6260>
2. PNAS. Social relationships and physiological determinants of longevity across the human life span. URL: <https://www.pnas.org/doi/10.1073/pnas.1511085112>
3. Libraries. What Is a Social Problem? URL: <https://open.lib.umn.edu/socialproblems/chapter/1-1-what-is-a-social-problem/>
4. Your dictionary. Common Examples of Social Issues. URL: <https://examples.yourdictionary.com/common-examples-of-social-issues.html>
5. В. І. Франчук. Теорія безпеки соціальних систем. Львів : ЛьвДУВС, 2016
6. Wikipedia. Піраміда потреб Абрагама Маслоу. URL: [https://uk.wikipedia.org/wiki/Піраміда\\_потреб\\_Абрагама\\_Маслоу](https://uk.wikipedia.org/wiki/Піраміда_потреб_Абрагама_Маслоу)
7. Mens line Men and emotions. URL: <https://mensline.org.au/mens-mental-health/men-and-emotions/>
8. Wikipedia. Психічний стан. URL: [https://uk.wikipedia.org/wiki/Психічний\\_стан](https://uk.wikipedia.org/wiki/Психічний_стан)
9. Wikipedia. Confidence. URL: <https://en.wikipedia.org/wiki/Confidence>
10. KidsHealth. Confidence. URL: <https://kidshealth.org/en/teens/confidence.html>
11. Департамент охорони здоров'я Тернопільської облдержадміністрації. Нормальний психічний стан – основа здоров'я людини. URL: <https://uozter.gov.ua/ua/pages/277>
12. Персонал. Журнал інтелектуальної еліти. Тест А. Ассінгера (оцінювання агресивності у стосунках). URL: <http://www.personal.in.ua/article.php?ida=526>
13. Підручники для студентів онлайн. Методика «Диференціальна діагностика депресивних станів». URL:

[https://stud.com.ua/124956/psihologiya/metodika\\_diferentsialna\\_diagnostika\\_depresivnih\\_staniv](https://stud.com.ua/124956/psihologiya/metodika_diferentsialna_diagnostika_depresivnih_staniv)

14. Online Test Pad. Методика визначення стресостійкості та соціальної адаптації Холмса і Page. URL: <https://onlinetestpad.com/ua/testview/94623-metodika-viznachennya-stresost%D1%96jkost%D1%96-ta-soc%D1%96alnoi-adaptac%D1%96i-kholmsa-%D1%96-rage>

15. Uchika. Методика дослідження впевненості на усвідомлюваному рівні. URL: <https://uchika.in.ua/kocan-g-v-lojkin-m.html?page=43>

16. Сотников А. В. Емоційний стан мовця та його вплив на просодичні показники мовлення. Збірник матеріалів круглого столу «Сучасні тенденції фонетичних досліджень» (12 травня 2017 р.). 107-109 сс. 2017 р

17. Кузенко Г. Мовні засоби вираження емотивності. URL: [http://ekmair.ukma.edu.ua/bitstream/handle/123456789/9836/Kuzenko\\_Movni\\_zasoby\\_vyrazhennya.pdf?sequence=1&isAllowed=y](http://ekmair.ukma.edu.ua/bitstream/handle/123456789/9836/Kuzenko_Movni_zasoby_vyrazhennya.pdf?sequence=1&isAllowed=y)

18. Маслова С.В. Виявленні емоцій у мовленні. <https://cyber.ru/article/n/vyyavlenni-emotsiy-u-movlenni/viewer>

19. M. A. Mohsin, A. P. Beltiukov Analysis of emotions from a text using syntactic methods. URL: <https://www.researchgate.net/publication/333489703>

20. Wikipedia. Семантика. URL: <https://uk.wikipedia.org/wiki/Семантика>

21. Wikipedia. Ключове слово. URL: [https://uk.wikipedia.org/wiki/Ключове\\_слово](https://uk.wikipedia.org/wiki/Ключове_слово)

22. Wikipedia. Словосполучення. URL: <https://uk.wikipedia.org/wiki/Словосполучення>

23. Wikipedia. Термін. URL: <https://uk.wikipedia.org/wiki/Термін>

24. Cran. Textrank for summarizing text. URL: <https://cran.r-project.org/web/packages/textrank/vignettes/textrank.html>

25. Big Bro. Частотний аналіз тексту: особливості і приклади. URL: <https://bigbro.com.ua/chastotnij-analiz-tekstu-osoblivosti-i-prikladi/>

26. Monkey learn. Understanding TF-IDF: A Simple Introduction. URL: <https://monkeylearn.com/blog/what-is-tf-idf/>

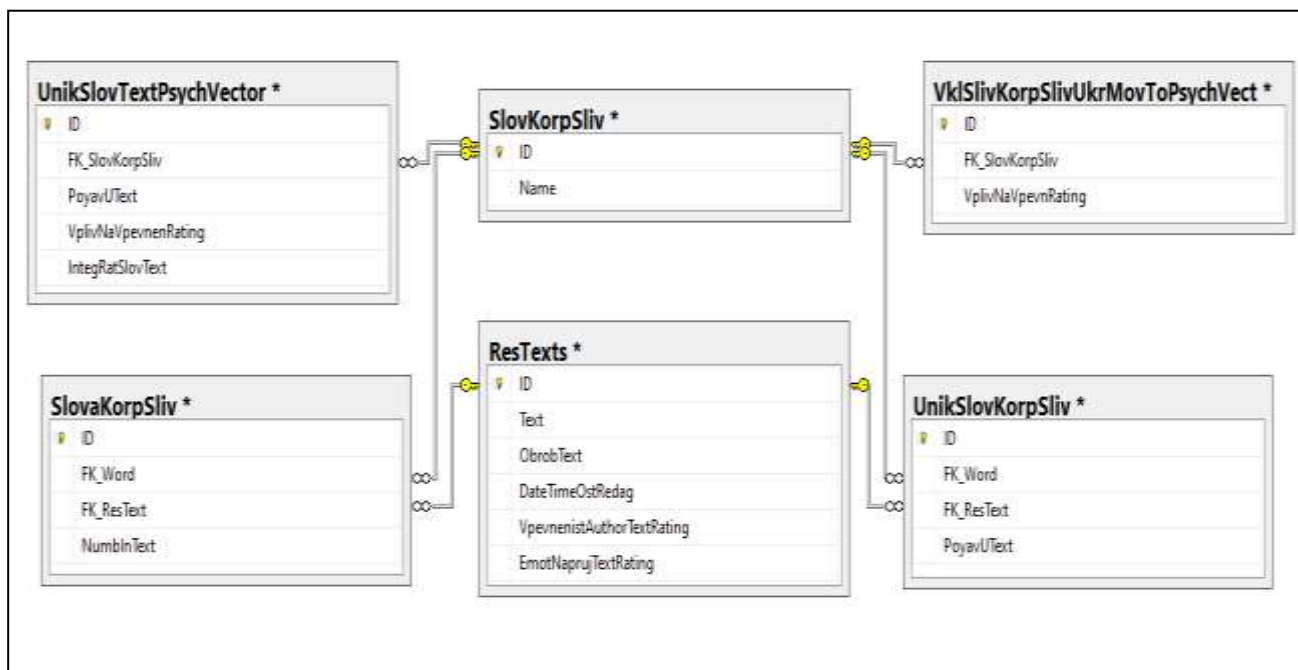
27. Д.В. Ланде. Оцінки і візуальність рівня дискримінантної сили слів. URL: <https://docplayer.com/41071205-Ocenki-i-vizualizaciya-urovnya-diskriminantnoy-sily-slov-d-v-lande-institut-problem-registracii-informacii-nan-ukrainy.html>
28. Генеральний регіонально анотований корпус української мови. URL: <http://uacorporus.org/Kyiv/ua>
29. MOVA.info. URL: <http://www.mova.info/corpus.aspx>
30. UA-GEC: перший анотований GEC-корпус української мови. URL: <https://github.com/grammarly/ua-gec>
31. Проєкт lang-uk. URL: <https://lang.org.ua/uk/>.
32. Словник Українського веб-корпусу Лейпцизького університету URL: [https://wortschatz.uni-leipzig.de/en/download/Ukrainian#ukr\\_mixed\\_2014](https://wortschatz.uni-leipzig.de/en/download/Ukrainian#ukr_mixed_2014)
33. Словопедія. URL: <http://slovopedia.org.ua/>
34. Горох. URL: <https://goroh.pp.ua/>
35. Великий електронний словник української мови. URL: <https://r2u.org.ua/vesum/>
36. IBM. Tone Analyzer. URL: <https://tone-analyzer-demo.ng.bluemix.net/>
37. Twin. Automate communications with Twin bots. URL: <https://twin24.ai/>
38. OneAI. Embed language ai in yur product. URL: <https://www.oneai.com/>
39. AWS. Amazon Competitive Features. URL: <https://aws.amazon.com/ru/comprehend/features/>
40. Smart Logger. Program for the operation of a multi-channel automatic digital recorder. URL: <https://platforms.su/platform/2470>
41. WEZOM. Desktop-додатки. URL: <https://wezom.com.ua/blog/desktop-prilozhenie>
42. Romi center. Мобільний додаток. URL: <https://romi.center/ru/learning/glossary/mobile-app/>
43. Decipher Zone. Top 10 Benefits of Having a Web Application. URL: <https://www.decipherzone.com/blog-detail/benefits-web-application-development>

44. Explorate Global. 12 Benefits of Dot Net Framework for Developing Business Applications. URL: <https://www.explorateglobal.com/blog/benefits-of-dot-net-framework>
45. Guru99. What is Java? Definition, Meaning & Features of Java Platforms URL: <https://www.guru99.com/java-platform.html>
46. Grand Circus. 10 Things to Know About the .NET Framework. URL: <https://www.grandcircus.co/blog/things-to-know-about-dot-net/>
47. Litslink. C# vs Python: Choosing the Right Language For Your Project URL: <https://litslink.com/blog/csharp-vs-python-choosing-right-language-for-your-project>
48. Compare Camp. Microsoft Visual Studio Review. URL: <https://comparecamp.com/microsoft-visual-studio-review-pricing-pros-cons-features/>
49. Tech Target. Definition Microsoft SQL Server. URL: <https://www.techtarget.com/searchdatamanagement/definition/SQL-Server>
50. United Solutions Group. 4 advantages of using Microsoft SQL Server: building for the future. URL: <https://blog.u-s-i.com/blog/4-advantages-of-using-microsoft-sql-server-building-for-the-future>

# ДОДАТКИ

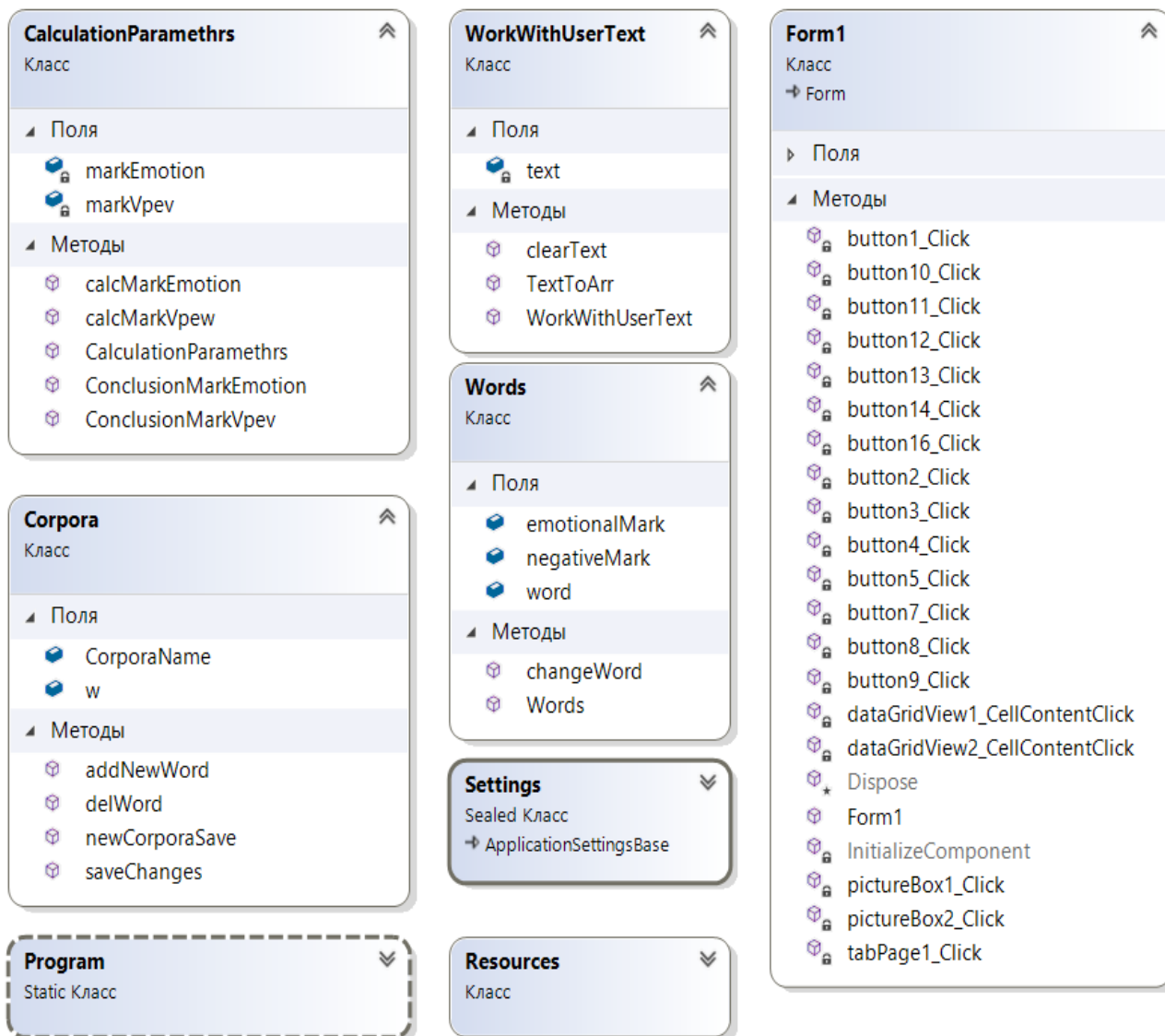
## Додаток А

## Структура бази даних інформаційної системи автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту



## Додаток Б

### Розгорнута структура класів інформаційної системи автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту



## Додаток В

### Презентаційний матеріал

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

## МЕТОД АВТОМАТИЗОВАНОГО ВИЗНАЧЕННЯ РІВНЮ ВПЕВНЕНOSTІ АВТОРА ЗА СЕМАНТИЧНИМ АНАЛІЗОМ ТЕКСТУ



**Виконав:**

*студент 4 курсу, група КН-18-1*  
КЛЕКАЧ СТЕПАН СЕРГІЙОВИЧ



**Керівник:**

*викладач кафедри КН*  
СОБКО ОЛЕНА ВІТАЛІЇВНА

## Актуальність

Розробка методів та засобів для автоматизованого визначення психоемоційних характеристик людини як метаданих до створеного нею текстового контенту дозволяє сприяти вирішенню цілої низки безпекових та критичних питань, як на рівні локальних задач, так і на державному рівні.

Ця задача є актуальна, адже з розвитком мережі Інтернет та різного роду соціальних мереж люди все частіше висловлюють свої думки в текстовій формі. Людям простіше висловлювати своє ставлення до якихось подій не в реальному житті, а у віртуальному, особливо коли є можливість приховати особистість, зробити це анонімно. З метою забезпечення безпеки на державному рівні, такого роду повідомлення можуть перехоплюватися з метою запобігання різного роду злочинів, як проти держави так і проти людяності.

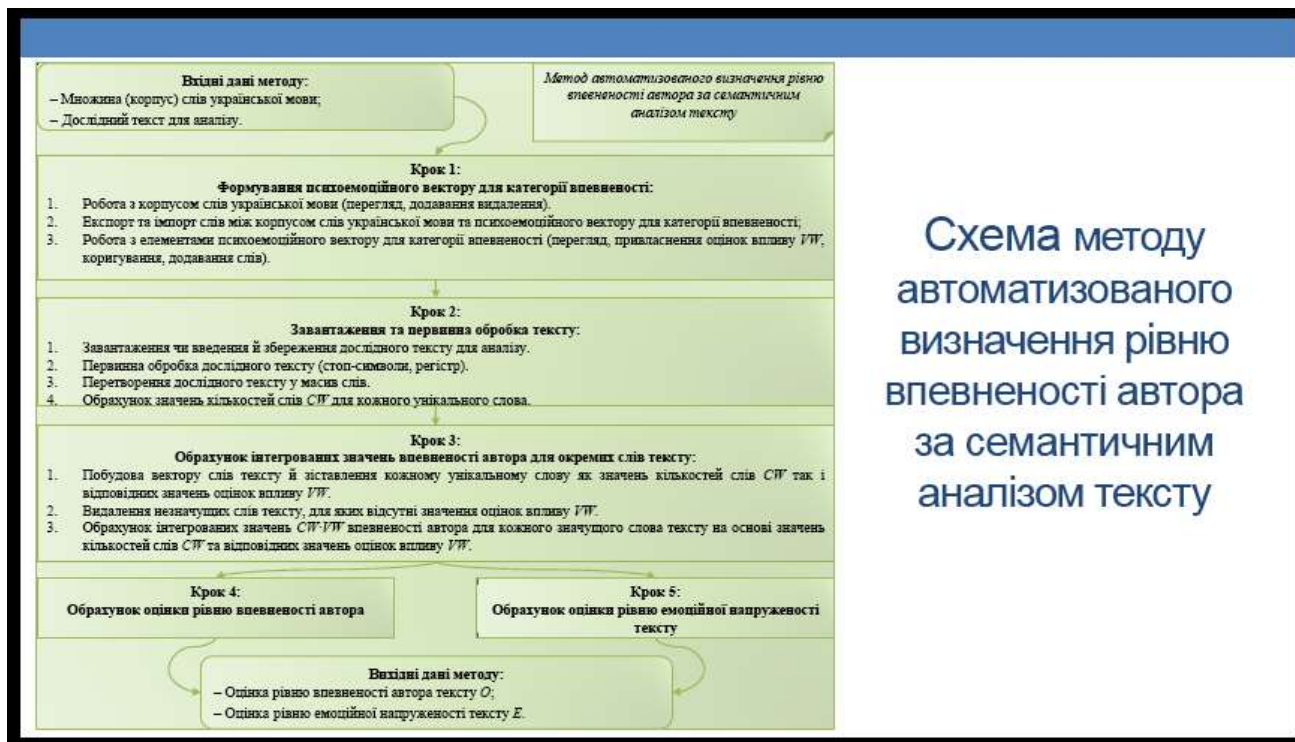
Розглянуте як приклад автоматизоване визначення рівню впевненості автора за семантичним аналізом текстового контенту є одним із випадків витягу психоемоційних складових автора за наявним цифровим текстом, що дозволяє робити висновок щодо актуальності та достовірності наявних у тексті відомостей.

- **Мета кваліфікаційної роботи бакалавра** – створення й програмна реалізація методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту.
- **Об'єкт дослідження** – процес семантичного аналізу цифрового тексту.
- **Предмет дослідження** – інформаційні технології, моделі, методи та засоби для автоматизованого визначення рівню впевненості автора у результаті семантичного аналізу цифрового тексту.

## Мета і задачі роботи

Метою кваліфікаційної роботи бакалавра є розробка методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту. Для досягнення мети потрібно вирішити такі задачі:

1. Провести аналіз предметної області, в процесі якого виконати огляд теоретичних підходів до розв'язку задач автоматизованого визначення рівню впевненості за семантичним аналізом тексту.
2. Проаналізувати існуючі програмні рішення для вирішення задачі КРБ.
3. Проаналізувати сучасні засоби для створення програмних рішень та обрати комплект засобів для створення власної програмної реалізації.
4. Розробити метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту, виконати на його основі проєктування інформаційної системи.
5. Створити та протестувати відповідну програмну реалізацію, яка буде реалізовувати запропонований метод. Для зручності використання створити інструкцію користувача до розроблюваного програмного продукту.



## Математична модель оцінки рівню впевненості автора тексту

Оцінка рівню впевненості автора тексту  $O$  визначається наступним чином:

$$O = \frac{\sum_{i=1}^n (SW_i \cdot W_i)}{m},$$

де  $n$  – кількість унікальних значущих слів дослідного тексту,  $m$  – загальна кількість слів у дослідному тексті для аналізу,  $SW_i$  – значення кількості зустрічань слова  $i$  дослідному тексті,  $W_i$  – значення привласненої оцінки впливу слова  $i$  на рівень впевненості згідно психоемоційного вектору

для категорії впевненості.

Градація	Значення
-1..-0,6	Надвисокий рівень невпевненості
-0,59..-0,3	Високий рівень невпевненості
-0,3..-0,1	Середній рівень невпевненості
-0,09..0	Низький рівень невпевненості
0,01..0,1	Низький рівень впевненості
0,11..0,3	Середній рівень впевненості
0,31..0,59	Високий рівень впевненості
0,6..1	Надвисокий рівень впевненості

## Математична модель оцінки рівню емоційної напруженості тексту

Оцінка рівню емоційної напруженості тексту  $E$  визначається наступним чином :

$$E = \frac{\sum_{i=1}^n |CW_i \cdot VW_i|}{m},$$

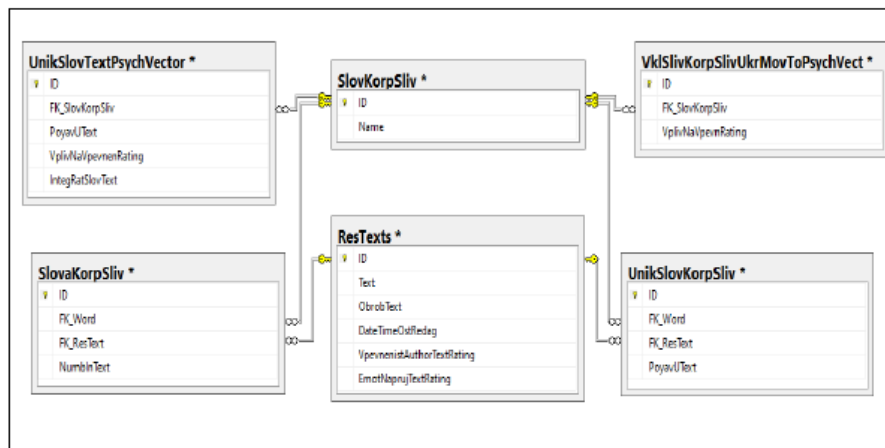
де  $n$  – кількість унікальних значущих слів дослідного тексту,  $m$  – загальна кількість слів у дослідному тексті для аналізу,  $CW_i$  – значення кількості зустрічань слова  $i$  дослідному тексті,  $VW_i$  – значення привласненої оцінки впливу слова  $i$  на рівень впевненості згідно психоемоційного вектору для категорії впевненості.

Градація	Значення
0,00 – 0,1	Низький рівень емоційної напруженості
0,11 – 0,3	Середній рівень емоційної напруженості
0,31 – 0,59	Високий рівень емоційної напруженості
0,6 – 1	Надвисокий рівень емоційної напруженості

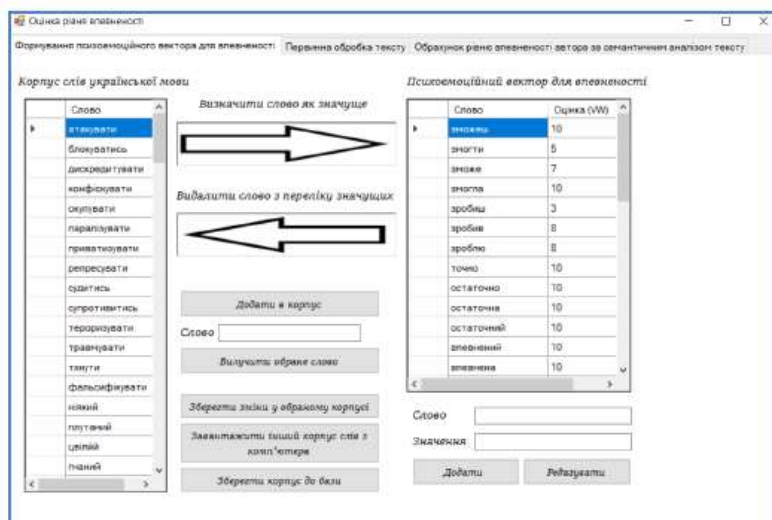
## Інформаційна структура системи автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту



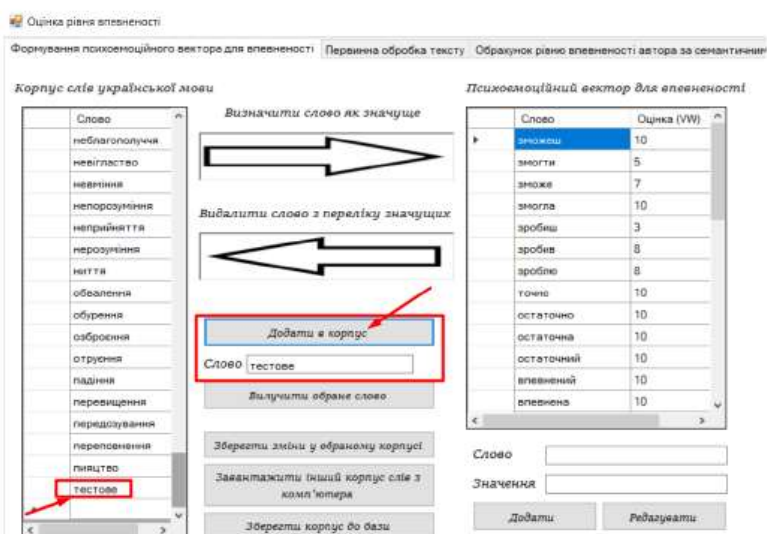
## Даталогічна модель бази даних системи автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту



## Стартовий екран застосування



## Додавання слова в корпус



## Виведення тексту з файлу

Оцінка рівня впевненості

Формування психоенціпічного вектора для впевненості | Первинна обробка тексту | **Обрахунок рівня впевненості автора за семантичним аналізом тексту**

Текст для аналізу

Завантажити | Зберегти

Занадто багато поганих новин і занадто багато невизначеності. Цей рік зіпсував нерви навіть міцним особистостям, і порушення сну – частина мозаїки. Безсоння, неглибокий сон, постійна сонливість – експерти розповідають, як впоратися з проблемами, які викликала пандемія. Обов'язково одягати рукавички в метро? А в автобусі? Потрібно обов'язково одягати маску, або ж підійде звичайна шовкова хустка? Чи будуть штрафи? Знову закриють дитячі садки? Скільки ще працювати віддалено? Всі ці питання викликають величезний стрес і почуття невизначеності, і ми оновлюємо і оновлюємо стрічку новин, щоб дізнатися, які заходи вживаються в регіонах і що там у сусідів. А після цього – годинами не можемо заснути, і увесь день відчуваємо себе «вареними». І втома росте, як снігова куля. Експерти в бесіді з Huffpost розповіли, як за вісім простих кроків перемогти безсоння. Під час пандемії ми відчули себе пов'язаними з усім світом, і в той же час дуже самотніми та ізольованими від оточуючих. Ми менше проводили часу з людьми і були менш активними – в тому числі через закриття громадських просторів. – Каже Джейд Ву, ліцензований клінічний психолог і дослідник сну в Ларемі штат Північна Кароліна. Ми не можемо

Відфільтрувати текст

Перетворити на масив слів

Обрахувати значення СВ

## Частота зустрічання слів тексту

Оцінка рівня впевненості

Формування психоенціпічного вектора для впевненості | Первинна обробка тексту | **Обрахунок рівня впевненості автора за семантичним аналізом тексту**

Текст для аналізу

Завантажити | Зберегти

[експерти, 3]  
[розповідають, 1]  
[впоратися, 2]  
[проблемами, 1]  
[які, 3]  
[викликала, 1]  
[пандемія, 1]  
[обов'язково, 2]  
[одягати, 2]  
[рукавички, 1]  
[метро, 1]  
[автобусі, 1]  
[потрібно, 3]  
[маску, 1]  
[або, 7]  
[підійде, 1]  
[звичайна, 1]  
[шовкова, 1]  
[хустка, 1]

Відфільтрувати текст

Перетворити на масив слів

Обрахувати значення СВ

## Результат обрахунків

Оцінка рівня впевненості

Формування психоемоційного вектора для впевненості | Паралельна обробка тексту | Обрахунок рівню впевненості автора за семантичним аналізом тексту

**Обрахунок рівню впевненості автора за семантичним аналізом тексту**

Слова	СВ (кількість зустрічаних)	VW (одичне значення впевненості)
невизначеності	2	0.0291970802919708
згубив	1	-0.0145985401459854
може	2	<b>0.0648681806486818</b>
добре	1	0.0072992700729927
рекомендують	2	0.0145985401459854

Показати слова з СВ

Обрахунок оцінки рівню впевненості автора

Видалити незначущі слова

Обрахунок оцінки рівню емоційної напруженості

Значення оцінки рівню впевненості автора: Значення оцінки рівню емоційної напруженості

-0.0267639902676399      0.09

Границі	Значення
-1 -0.6	Надвисокий рівень непевненості
-0.59 -0.3	Високий рівень непевненості
-0.3 -0.1	Середній рівень непевненості
-0.09 0	Низький рівень непевненості
0.01 0.1	Низький рівень впевненості
0.11 0.3	Середній рівень впевненості
0.31 0.59	Високий рівень впевненості
0.6 1	Надвисокий рівень впевненості

Границі	Значення
0.00 - 0.1	Низький рівень емоційної напруженості
0.11 - 0.3	Середній рівень емоційної напруженості
0.31 - 0.59	Високий рівень емоційної напруженості
0.6 - 1	Надвисокий рівень емоційної напруженості

## Висновки

Розробка методів та засобів для автоматизованого визначення психоемоційних характеристик людини як метаданих до створеного нею текстового контенту дозволяє сприяти вирішенню цілої низки безпекових та критичних питань, як на рівні локальних задач, так і на державному рівні. Розглянуте як приклад автоматизоване визначення рівню впевненості автора за семантичним аналізом текстового контенту є одним із випадків витягу психоемоційних складових автора за наявним цифровим текстом, що дозволяє робити висновок щодо актуальності та достовірності наявних у тексті відомостей.

У ході виконання КРБ були вирішені такі задачі:

1. Проведено аналіз предметної області
2. Проаналізовано існуючі програмні рішення схожі з тематикою КРБ.
3. Проаналізовано сучасні засоби для створення програмних рішень та обрано комплект засобів для створення власної програмної реалізації.
4. Запропоновано, програмно реалізовано та протестовано метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту.
5. Створено інструкцію користувача до розробленої програмної реалізації.

Ім'я користувача:  
Кафедра КН

Дата перевірки:  
12.06.2022 10:21:55 EEST

Дата звіту:  
12.06.2022 10:22:35 EEST

ID перевірки:  
1011549012

Тип перевірки:  
Doc vs Internet + Library

ID користувача:  
100005671

Назва документа: Клекач\_ЗАПИСКА\_short

Кількість сторінок: 67 Кількість слів: 10297 Кількість символів: 78039 Розмір файлу: 4.80 MB ID файлу: 1011421045

Виявлено модифікації тексту (можуть впливати на відсоток схожості)

## 7.06% Схожість

Найбільша схожість: 3.23% з джерелом з Бібліотеки (ID файлу: 1011420928)

2.37% Джерела з Інтернету 69 ..... Сторінка 69

6.43% Джерела з Бібліотеки 135 ..... Сторінка 69

## 0% Цитат

Вилучення цитат вимкнене

Вилучення списку бібліографічних посилань вимкнене

## 0% Вилучень

Немає вилучених джерел

## Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Замінені символи 2

Підозріле форматування 21 сторінка

## Anti-Plagiarism v-15.257

**Максимальное совпадение с одним документом 11.0%**

Словари проверки: en\_US, ru\_RU, ua\_UA. **Ошибок в документах: 10%**

ID: 105050 Название: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту Добавлено в БД: 2022-06-12 Авторы: С.С. Клекач Руководители: О.В. Собко Консультанты: Оponentы:	Документ		Суммарное совпадение по Базе Данных	
	Символы	Лексемы	Символы	Лексемы
	61132	893	9387 (15%)	137 (15%)

### Источник плагиата

ID	Описание	Наличие плагиата в документе	
		Символы	Лексемы
104983	Название: ЗВІТ з професійної практики база практики ТОВ “ДЕВЛУП.ПРО” м. Хмельницький Добавлено в БД: 2022-06-10 Авторы: С.С. Клекач Руководители: Т.К. Скрипник Консультанты: Оponentы:	6691 (11.0%)	109 (12.0%)

**РІШЕННЯ ЕКСПЕРНОЇ КОМІСІЇ КАФЕДРИ КОМП'ЮТЕРНИХ НАУК  
ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ**

Підтверджуємо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: Метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту

Автор: студент групи КН-18-1 Клекач Степан Сергійович

Спеціальність: 122 – Комп'ютерні науки

Освітня програма: освітньо-професійна

Науковий керівник: викладач кафедри КН Собко Олена Віталіївна

Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом. Робота приймається до захисту.	<i>відповідає</i>
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи.	
3	Виявлені запозичення не є плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. В зв'язку з цим мета роботи та поставлені завдання не були досягнені. Робота може бути допущена до захисту (наступного року) після того як буде відкоригована та допрацьована і успішно пройде повторну перевірку на академічний плагіат.	
4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	

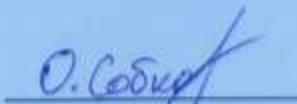
*Підтвердження: запозичення, виявлені в роботі С.С. Клекача, є законними і не є плагіатом, оскільки:*

*1) за програмою Anti-Plagiarism виявлені 15% запозичень вказують на документ автора роботи та містять його ж Звіт з практики.*

*2) За програмою UNICHECK виявлені 7,06%, які є фрагментарними, не більше 3.03% на джерело – містять поширені конструкції, загальновідомі терміни та визначення.*

*3) запозичення розміщені в розділах аналізу існуючих аналогів та прототипів, які не описують безпосередньо авторське дослідження і не стосуються результатів роботи;*

Керівник роботи



Олена СОБКО

Гарант ОП



Олександр МАЗУРЕЦЬ

Завідувач кафедри КН



Олександр БАРМАК



## ВІДГУК НАУКОВОГО КЕРІВНИКА на кваліфікаційну роботу бакалавра

студента *гр. КН-18-1 Клекача Степана Сергійовича*

за темою Метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту

### 1. Актуальність теми

Виявлення психоемоційного стану автора тексту є важливою задачею, адже це не тільки вирішує задачу аналізу особистості, а й може допомогти в запобіганні порушення безпеки інших особистостей. Такі методи можуть застосовуватись локально, або у більш глобальних задачах забезпечення державної безпеки. У даній роботі розглядається визначення рівню впевненості автора тексту, як приклад вирішення задачі з визначення психоемоційного стану людини. Відповідно, автоматизація процесу визначення рівню впевненості автора за семантичним аналізом тексту є актуальною задачею комп'ютерних наук.

### 2. Відповідність роботи предметній області спеціальності 122 Комп'ютерні науки

Поставлена у кваліфікаційній роботі бакалавра мета стосується розробки методів і технологій отримання, зберігання, обробки, передачі та використання інформації, інтелектуального аналізу даних і прийняття рішень, а саме розробки методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту. При цьому при вирішенні поставлених задач використовуються математичні моделі, методи та алгоритми розв'язання теоретичних і прикладних задач, що виникають при розробці інформаційних технологій. Тому результати виконання кваліфікаційної роботи бакалавра відповідають стандарту бакалавра спеціальності 122 – Комп'ютерні науки.

### 3. Професійні та особистісні якості бакалавра

При роботі над кваліфікаційною роботою бакалавра Клекач Степан Сергійович проявив себе підготовленим, кваліфікованим фахівцем та дисциплінованим студентом, вчасно виконуючи поставлені етапи дослідження. Як в процесі написання пояснювальної записки, так і при розробці прикладного програмного забезпечення проявив достатні для одержання успішного результату компетентності.

#### **4. Ступінь самостійності під час виконання кваліфікаційної роботи**

Одержані в роботі результати є наслідком особистої діяльності студента, який самостійно виконував всі поставлені задачі.

#### **5. Ступінь оволодіння методами дослідження**

В роботі при розробці та прикладній реалізації методу автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту виявлено достатній ступінь оволодіння студентом необхідними інструментами та обладнанням, методами, методиками та технологіями предметної області комп'ютерних наук.

#### **6. Повнота та якість розкриття теми роботи**

Тема роботи в повній мірі обґрунтована й розкрита, проведено аналіз актуальності та відомих досліджень в межах обраної теми, поставлені завдання, які у роботі виконані, та розроблено програмне забезпечення для автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту.

#### **7. Логічність, послідовність, аргументованість, літературна грамотність викладу матеріалу**

Структура роботи та послідовність викладення логічні та відповідні поставленій меті. Викладення матеріалу грамотне та виявляє високий ступінь відповідності стилю.

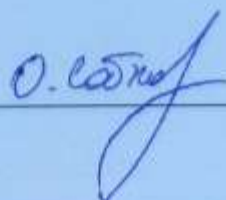
#### **8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин**

Запропонований метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту може мати практичне використання. Розробка методів та засобів для автоматизованого визначення психоемоційних характеристик людини як метаданих до створеного нею текстового контенту дозволяє сприяти вирішенню цілої низки безпекових та критичних питань, як на рівні локальних задач, так і на державному рівні. Розглянуте як приклад автоматизоване визначення рівню впевненості автора за семантичним аналізом текстового контенту є одним із випадків витягу психоемоційних складових автора за наявним цифровим текстом, що дозволяє робити висновок щодо актуальності та достовірності наявних у тексті відомостей.

#### **9. Висновок про можливість допуску кваліфікаційної роботи бакалавра до захисту, на яку оцінку заслуговує робота**

Враховуючи високий рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «добре».

Керівник



викладач каф. КН Олена СОБКО



## РЕЦЕНЗІЯ

### на кваліфікаційну роботу бакалавра

студента гр. КН-18-1 Клекач Степан Сергійович

за темою: Метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту

#### 1. Актуальність обраної теми

Розробка методів та засобів для автоматизованого визначення психоемоційних характеристик людини як метаданих до створеного нею текстового контенту дозволяє сприяти вирішенню цілої низки безпекових та критичних питань, як на рівні локальних мереж, так і на державному рівні. Розглянуте як приклад автоматизоване визначення рівню впевненості автора за семантичним аналізом текстового контенту є одним із способів вирішення психоемоційних складових автора за наявним цифровим текстом, що є актуальною задачею комп'ютерних наук.

#### 2. Повнота розкриття мети та завдань дослідження

Кваліфікаційна робота бакалавра студента Клекача С.С. виконана в повному обсязі, було проаналізовано предметну область, створено метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту та реалізовано програмний застосунок на базу вищезазначеного методу.

#### 3. Зміст кожного розділу роботи

Перший розділ присвячений проведенню аналізу предметної області та визначенню основних параметрів для розв'язку поставленої задачі. Другий розділ присвячений проєктуванню функціональної структури інформаційної системи. Третій розділ присвячений програмній реалізації спроектованої функціональної структури інформаційної системи. Також сформовано основний висновок роботи.

#### 4. Оцінка розробленої інформаційної системи, її практична цінність

Створений метод автоматизованого визначення рівню впевненості автора за семантичним аналізом тексту та програмний застосунок на його основі дозволяють користувачеві отримати швидкі відповіді на запитання. Реалізований метод та програмний продукт на його основі можна застосувати в наукових та прикладних цілях.

#### 5. Якість оформлення кваліфікаційної роботи бакалавра

Робота виконана на належному науково-методичному рівні та відповідає встановленим вимогам щодо оформлення такого роду праць.

#### 6. Недоліки кваліфікаційної роботи бакалавра

Кваліфікаційна робота бакалавра має кілька незначних відхилень від норм оформлення пояснювальних записок, втім це не впливає на одержані результати.

8. Загальний висновок (допускається чи не допускається до захисту), та оцінка якої оцінки заслуговує кваліфікаційна робота.

Враховуючи рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «добре».

Рецензент Бедрагон А.Р., д.ф.-м.н.