

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний аерокосмічний університет ім. М.Є. Жуковського  
«Харківський авіаційний інститут»

Факультет програмної інженерії та бізнесу

Кафедра інженерії програмного забезпечення

## Пояснювальна записка до дипломної роботи

магістра  
(освітній ступінь)

на тему «Дослідження сервісів оцінки тональності тексту»

XAI.603.667П1.121.156345.200

Виконав: студент б курсу групи № 667П1  
Спеціальність 121 – Інженерія програмного  
забезпечення

(код та найменування)

Освітня програма Хмарні обчислення  
та Інтернет речей

(найменування)

Петрак О. В.

(прізвище й ініціали студента)

Керівник: Шевченко І. В.

(прізвище й ініціали)

Рецензент: Ткачов В. М.

(прізвище й ініціали)

Харків – 2020

**Міністерство світи і науки України**  
**Національний аерокосмічний університет ім. М. Є. Жуковського**  
**«Харківський авіаційний інститут»**

Факультет програмної інженерії та бізнесу  
(повне найменування)

Кафедра інженерії програмного забезпечення  
(повне найменування)

Рівень вищої освіти другий (магістерський)

Спеціальність 121 – інженерія програмного забезпечення  
(код та найменування)

Освітня програма хмарні обчислення та Інтернет речей  
(найменування)

**ЗАТВЕРДЖУЮ**  
**Завідувач кафедри**  
**Туркін І. Б.**

(підпис) \_\_\_\_\_ (ініціали та прізвище)  
“ \_\_\_\_\_ ” \_\_\_\_\_ 2020 року

**З А В Д А Н Н Я**  
**НА ДИПЛОМНУ РОБОТУ СТУДЕНТУ**

Петраку Олексію Васильовичу  
(прізвище, ім'я, по батькові)

1. Тема дипломної роботи Дослідження сервісів оцінки тональності тексту  
керівник дипломної роботи Шевченко Ілона Володимирівна к.т.н., доцент,  
( прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом Університету № \_\_\_\_\_ від “ \_\_\_\_\_ ” \_\_\_\_\_ 2020 року

2. Термін подання студентом роботи 02.12.2020

3. Вихідні дані до роботи Дослідження сервісів оцінки тональності тексту

4. Зміст пояснювальної записки (перелік питань, які потрібно розробити)

1. Аналіз сучасного стану проблеми аналізу тональності тексту. Мета і задачі дослідження.

2. Планування експерименту для дослідження ефективності дослідження тональності тексту.

3. Аналіз результатів дослідження ефективності сервісів оцінки тональності тексту та надання практичних рекомендацій.

5. Перелік графічного матеріалу Мета і задачі дипломної роботи – 1 слайд,  
об'єкт, предмет та методи дослідження – 1 слайд, актуальність роботи – 3  
слайди, аналіз сучасного стану проблеми оцінки тональності тексту – 3 слайди,  
планування експерименту для оцінки ефективності сервісів – 4 слайди, аналіз  
результатів дослідження – 3 слайди, висновки, практична значущість та наукова  
новизна – 1 слайд.

## 6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1÷3	к.т.н. доц.каф.603 Шевченко І.В.		

Нормоконтроль \_\_\_\_\_ Постернакова В.А. «\_\_» \_\_\_\_\_ 20\_\_ р.  
(підпис) (ініціали та прізвище)

7. Дата видачі завдання « 1 » вересня 2020 р.

## КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломної роботи	Строк виконання етапів роботи	Примітка
1	Отримання і затвердження теми диплому	01.09.2020	
2	Обґрунтування актуальності дослідження ефективності сервісів для оцінки тонального забарвлення текстів	12.09.20	
3	Обґрунтування мети, об'єкту, предмету та методів дослідження	15.09.20	
4	Аналіз існуючих алгоритмів, рішень і методів визначення тонального забарвлення текстів	20.09.20	
5	Аналіз існуючих алгоритмів, рішень і методів визначення тонального забарвлення текстів	01.10.20	
6	Проектування експерименту з визначення ефективності сервісів визначення тонального забарвлення текстів	10.10.20	
7	Проведення експерименту з визначення ефективності сервісів визначення тонального забарвлення текстів;	21.10.20	
8	Порівняння отриманих результатів експерименту	05.11.20	
9	Оформлювання пояснювальної записки до дипломного проекту	22.11.20	
10	Передзахист дипломного проекту	27.11.20	
11	Захист дипломного проекту	07.11.20	

Студент \_\_\_\_\_

(підпис)

Петрак О.В.

(прізвище та ініціали)

Керівник проекту \_\_\_\_\_

(підпис)

Шевченко І.В.

(прізвище та ініціали)

## РЕФЕРАТ

Пояснювальна записка на дипломну роботу: 88 с., 30 рис., 8 табл., 2 Додатки, 37 джерел.

Мета дослідження – визначення ефективності загальнодоступних сервісів, що дозволяють оцінювати задоволеність клієнтів без втручання людини на основі аналізу тональності текстових відгуків за спожиті товари чи послуги.

Об'єкт дослідження – процеси аналізу тональності тексту.

Предмет дослідження – сервіси аналізу тональності тексту.

Поставлена мета буде досягнута за рахунок дослідження сервісів з оцінки тональності тексту. Для цього необхідно вирішити такі завдання:

- проаналізувати стан проблеми оцінки тональності тексту;
- розробити план експериментального дослідження ефективності роботи існуючих сервісів оцінки тональності тексту;
- провести експериментальне дослідження та надати практичні рекомендації щодо вибору конкретного сервісу аналізу тональності тексту для певної ситуації.

У дипломній роботі досліджено ефективність роботи сучасних комерційних сервісів з оцінки тональності тексту (sentiment analysis). Ці сервіси базуються на досягненнях у сфері обробки природньої мови (NLP – natural language processing). Проведено аналіз сучасних технологій, які використовуються зазначеними сервісами, а також зазначено відомі проблеми, все ще не вирішені на сьогоднішній день.

Проведено планування тесту для дослідження сервісів, у рамках якого розроблено вимоги та обмеження для тестового набору відних даних та визначено формат вихідних результатів тестування на підставі якого буде аналізуватися ефективність вищезазначених сервісів.

Згідно зі сформованим планом проведено тестування сервісів оцінки тональності тексту. За результатами тестування проведено аналіз та зроблено висновки про ефективність досліджених сервісів на наборі реальних тестових даних.

**ОЦІНКА ТОНАЛЬНОСТІ ТЕКСТУ, ОБРОБКА ПРИРОДНЬОЇ МОВИ, АНАЛІЗ ТЕКСТУ, ЕФЕКТИВНІСТЬ СЕРВІСІВ, NLP**

## РЕФЕРАТ

Пояснительная записка на дипломную работу: 88 с., 30 рис., 8 табл., 2  
Додатки, 37 источника.

Цель исследования – определение эффективности общедоступных сервисов, позволяющих оценивать удовлетворенность клиентов без вмешательства человека на основе анализа тональности текстовых отзывов за потребленные товары или услуги.

Объект исследования – процессы анализа тональности текста.

Предмет исследования – сервисы анализа тональности текста.

Поставленная цель будет достигнута за счет исследования сервисов по оценке тональности текста. Для этого необходимо решить следующие задачи:

- проанализировать состояние проблемы оценки тональности текста;
- разработать план экспериментального исследования эффективности работы существующих сервисов оценки тональности текста;
- провести экспериментальное исследование и дать рекомендации по выбору конкретного сервиса анализа тональности текста для определенной ситуации.

В дипломной работе исследована эффективность работы современных коммерческих сервисов по оценке тональности текста (sentiment analysis). Эти сервисы базируются на достижениях в сфере обработки естественного языка (NLP - natural language processing). Проведен анализ современных технологий, которые используются указанными сервисами, а также обозначены известные проблемы, все еще не решенные на сегодняшний день.

Произведено планирование эксперимента для исследования сервисов, в рамках которого разработаны требования и ограничения для тестового набора тестовых данных и определили формат выходных результатов тестирования на основании которого будет анализироваться эффективность вышеуказанных сервисов.

Согласно со сложившемся планом проведено тестирование сервисов оценки тональности текст, по результатам тестирования проведен анализ и сделаны выводы об эффективности исследованных сервисов на наборе реальных тестовых данных.

**ОЦЕНКА ТОНАЛЬНОСТИ ТЕКСТА, ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА, АНАЛИЗ ТЕКСТА, ЭФФЕКТИВНОСТЬ СЕРВИСОВ, NLP**

## ABSTRACT

Explanatory note for the diploma work: 88 p., 30 pic., 8 table., 2 Applications, 37 sources.

The purpose of the study is to determine the effectiveness of publicly available services that allow you to assess customer satisfaction without human intervention based on the analysis of the tone of text reviews for the goods or services consumed.

The object of study - the processes of analysis of the tone of the text.

The subject of research - text tone analysis services.

This goal will be achieved through the study of services to assess the tone of the text. To do this, you must solve the following tasks:

- analyze the state of the problem of assessing the tone of the text;
- develop a plan for an experimental study of the effectiveness of existing services for assessing the tone of the text;
- to conduct an experimental study and provide practical recommendations for choosing a specific service for analyzing the tone of the text for a particular situation.

In the thesis, the effectiveness of the work of modern commercial services for assessing the sentiment of a text (sentiment analysis) was investigated. These services build on advances in natural language processing (NLP). An analysis of modern technologies that are used by these services is carried out, and also known problems are identified that have not yet been resolved to date.

An experiment was planned for the study of services, within the framework of which requirements and restrictions for a test set of test data were developed and the format of the output test results was determined on the basis of which the effectiveness of the above services will be analyzed.

According to the established plan, the services for assessing the sentiment of the text were tested, based on the test results, an analysis was carried out and conclusions were drawn about the effectiveness of the services studied on a set of real test data.

**SENTIMENT ANALYSIS, NATURAL LANGUAGE PROCESSING, TEXT ANALYSIS, EFFICIENCY OF SERVICES, NLP**

## ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ ВИМІРЮВАНЬ ФІЗИЧНИХ ВЕЛИЧИН, СКОРОЧЕНЬ І ТЕРМІНІВ .....	9
ВСТУП .....	10
1 АНАЛІЗ СУЧАСНОГО СТАНУ ПРОБЛЕМИ АНАЛІЗУ ТОНАЛЬНОСТІ ТЕКСТУ. МЕТА І ЗАДАЧІ ДОСЛІДЖЕННЯ.....	12
1.1 Проблематика задач, які вирішуються сервісами обробки природньої мови .....	12
1.1.1 Передумови появи інтелектуальних систем обробки природньої мови.....	12
1.1.2 Особливості роботи систем обробки природньої мови .....	14
1.1.3 Задачі, які виконують системи обробки природньої мови .....	19
1.2 Проблематика задач, які вирішуються сервісами оцінки тональності тексту.....	22
1.2.1 Різновиди задач, що вирішуються сервісами оцінки тональності тексту.....	22
1.2.2 Визначення проблематики сучасних сервісів оцінки тональності тестів.....	28
1.3 Висновки до розділу 1 .....	32
2 ПЛАНУВАННЯ ЕКСПЕРИМЕНТУ ДЛЯ ДОСЛІДЖЕННЯ ЕФФЕКТИВНОСТІ СЕРВІСІВ ОЦІНКИ ТОНАЛЬНОСТІ ТЕКСТУ .....	33
2.1 Загальний опис проблеми оцінки тональності тексту.....	33
2.2 Розгляд обраних сервісів .....	33
2.2.1 Microsoft Text Analytics API.....	33
2.2.2 Google Cloud Natural Language API.....	36
2.2.3 Sentiment Analysis від Tweenword .....	38
2.2.4 Sentiment Analysis від Text Processing.....	39
2.3 Вибір вхідних факторів і відгуків для визначення ефективності роботи сервісів оцінки тональності текстової інформації .....	40
2.3.1 Розгляд даних з якими використовуються досліджувані алгоритми.....	40
2.4 Визначення факторів і відгуків.....	52
2.5 Планування експерименту .....	52
2.6 Програмні засоби для проведення експерименту.....	53
2.7 Висновок до розділу 2 .....	55
3 АНАЛІЗ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ ЕФФЕКТИВНОСТІ СЕРВІСІВ ОЦІНКИ ТОНАЛЬНОСТІ ТЕКСТУ ТА НАДАННЯ ПРАКТИЧНИХ РЕКОМЕНДАЦІЙ .....	56
3.1 Формування цілей і постановка задачі для розробки програмного забезпечення .....	56
3.2 Огляд технологій і визначення формату для розробки програмного забезпечення .....	56
3.3 Побудова діаграми варіантів використання.....	58

3.3 Вимоги до ПО .....	59
3.3.1 Функціональні вимоги .....	59
3.3.2 Нефункціональні вимоги .....	60
3.4 Специфікація варіантів використання .....	60
3.5 Огляд досліджуваних сервісів .....	62
3.6 Початкові дані .....	63
3.7 Тестування сервісів з аналізу тональності тексту .....	65
3.8 Формат даних сервісів для оцінки тональності тексту .....	65
3.8.1 Формат даних сервісу Text Analytics API від Microsoft .....	66
3.8.2 Формат даних сервісу Sentiment Analysis від Tweenword .....	67
3.8.3 Формат даних сервісу Sentiment Analysis від Text Processing .....	68
3.8.4 Формат даних сервісу Natural Language API від Google .....	69
3.9 Результати тестування сервісів оцінки тональності тексту .....	69
3.10 Аналіз результатів дослідження сервісів оцінки тональності тексту .....	71
3.10.1 Аналіз результатів сервісу Microsoft Text Analytics API .....	71
3.10.2 Аналіз результатів сервісу Tweenword Sentiment Analysis .....	73
3.10.3 Аналіз результатів сервісу Text Processing Sentiment Analysis .....	75
3.10.4 Аналіз результатів сервісу Natural Language API від Google .....	75
3.10.5 Загальний порівняльний аналіз результатів сервісів оцінки тональності тексту .....	76
3.11 Рекомендації щодо використання результатів дослідження .....	77
3.12 Висновки до розділу 3 .....	78
ВИСНОВКИ .....	80
ПЕРЕЛІК ПОСИЛАНЬ .....	81



## **ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ ВИМІРЮВАНЬ ФІЗИЧНИХ ВЕЛИЧИН, СКОРОЧЕНЬ І ТЕРМІНІВ**

API – Application Programming Interface – прикладний програмний інтерфейс;

HTTP – Hyper Text Transfer Protocol – протокол передачі гіпертекстових документів;

NLP – Natural Language Processing – обробка природньої мови;

REST – Representational State Transfer – передача репрезентативного стану;

URI – Uniform Resource Identifier – уніфікований ідентифікатор ресурсів;

URL – Uniform Resource Locator – уніфікований локатор ресурсів;

ПЗ – програмне забезпечення.

## ВСТУП

**Актуальність дослідження.** Зворотній зв'язок з клієнтами грає визначальну роль у розвитку та життєдіяльності будь-якого вигляду бізнесу. Уважний контроль за рівнем задоволеності споживачів дозволяє досягти стійкого довготривалого розвитку. У певних галузях може бути один клієнт, з яким підтримується прямий зв'язок. Але частіше за все у середнього та великого бізнеса кількість клієнтів досягає тисяч користувачів. Аби дослідити думку користувача у такому випадку, доводиться використовувати спеціальний інструментарій.

Власники сайтів та інших різновидів електронних сторінок мають можливість дізнатися відношення своїх покупців до власних послуг або товарів за допомогою відгуків. Відгуки представляють з себе доволі простий і корисний інструмент, однак при великій кількості користувачів він може стати доволі важким у контролі, вимагати значних витрат часу.

З розвитком штучного інтелекту з'явилась галузь яка займається питаннями пов'язаними з обробкою природної мови. Це напрям, який займається вивченням питань пов'язаних з синтезом природніх мов, а також їх комп'ютерним аналізом.

Її різновидом став аналіз тональності тексту. Це клас методів контент-аналізу в комп'ютерній лінгвістиці, призначений для автоматизованого виявлення в текстах емоційно забарвленої лексики і емоційної оцінки авторів (думок) по відношенню до об'єктів, мова про які йде в тексті [1].

Цей інструмент дозволяє оцінювати задоволеність клієнтів без втручання людини. Таким чином можна використовувати людський ресурс найманих працівників у більш пріоритетних напрямках, які вимагають більш складних операцій.

Системи пов'язані зі штучним інтелектом знаходяться у постійному розвитку, і системи аналізу тональності тексту – не виняток. На сьогоднішній день існують готові програмні рішення, які існують у широкому доступі для усіх бажаючих, та які можна використати для власних цілей. Вони вже почали використовуватися у реальних умовах для вирішення бізнес задач, однак неточності і помилки в їх роботі все ще трапляються. Також різні сервіси пропонують різний час обробки інформації і різні тарифні плани за надані послуги. Відрізняється також і точність оцінки текстових даних.

Аби визначити ефективні сервіси, серед існуючих у загальному доступі на сьогодні, важливо провести порівняльний аналіз на реальних відібраних даних. Визначити рівень правильності оцінки даних. Це дозволить побачити ефективність існуючих систем та оцінити коректність їх роботи у реальних умовах.

**Мета дослідження** – визначення ефективності загальнодоступних сервісів, що дозволяють оцінювати задоволеність клієнтів без втручання людини на основі аналізу тональності текстових відгуків за спожиті товари чи послуги.

Поставлена мета буде досягнута за рахунок дослідження сервісів з оцінки тональності тексту. Для цього необхідно вирішити такі **завдання**:

- проаналізувати стан проблеми оцінки тональності тексту;
- розробити план експериментального дослідження ефективності роботи існуючих сервісів оцінки тональності тексту;
- провести експериментальне дослідження та надати практичні рекомендації щодо вибору конкретного сервісу аналізу тональності тексту для певної ситуації.

**Об’єкт дослідження** – процеси аналізу тональності тексту.

**Предмет дослідження** – сервіси аналізу тональності тексту.

**Наукова новизна** – отримало подальший розвиток експериментальне дослідження найновіших доступних сервісів оцінки тональності тексту. Дослідження порівнює результати роботи сервісів на наборі реальних даних і дозволяє оцінити точність сучасних бізнес-інструментів оцінки тональності тексту.

**Практичне значення отриманих результатів** – результати дослідження можуть бути корисними при виборі сервісу оцінки тональності тексту для оптимізації бізнес-процесів.

# 1 АНАЛІЗ СУЧАСНОГО СТАНУ ПРОБЛЕМИ АНАЛІЗУ ТОНАЛЬНОСТІ ТЕКСТУ. МЕТА І ЗАДАЧІ ДОСЛІДЖЕННЯ

## 1.1 Проблематика задач, які вирішуються сервісами обробки природньої мови

### 1.1.1 Передумови появи інтелектуальних систем обробки природньої мови

Все, що висловлюється людьми (усно або письмово), містить величезну кількість інформації. Тема, яку обирають для розмови, тон розмови, підбір слів, все додає певний тип інформації, яку можна інтерпретувати та витягнути з неї цінність. Теоретично можливо зрозуміти і навіть передбачити поведінку людини, використовуючи цю інформацію [2].

Але є проблема: одна людина може генерувати сотні чи тисячі слів у декларації, кожне речення з відповідною складністю. Якщо виникає бажання масштабувати та аналізувати кілька сотень, тисяч, мільйонів людей чи декларацій у певній географії, тоді ситуація некерована.

Дані, отримані з розмов, декларацій або навіть твітів, є прикладами неструктурованих даних. Неструктуровані дані не вписуються в традиційну структуру рядків і стовпців реляційних баз даних і представляють переважну більшість даних, наявних у реальному світі. Це складні дані якими важко маніпулювати. Тим не менше, завдяки досягненню таких дисциплін, як машинне навчання, щодо цієї теми відбувається велика революція. У сучасному світі мова вже не про спроби інтерпретувати текст чи промову на основі його ключових слів (старомодний механічний спосіб), а про розуміння значення цих слів (когнітивний спосіб). Таким чином можна виявити такі фігури мови, як іронія, або навіть провести аналіз настроїв.

Обробка природньої мови або NLP - це область штучного інтелекту, яка надає машинам можливість читати, розуміти та виводити значення з людських мов.

Це дисципліна, яка зосереджена на взаємодії між наукою про дані та людською мовою, і охоплює багато галузей. Сьогодні НЛП процвітає завдяки значно розширеному доступу до даних та збільшенню обчислювальної потужності обчислювальних систем, що дозволяє досягти розробникам новітніх систем значних результатів у таких сферах, як охорона здоров'я, медіа, фінанси та людські ресурси, серед інших.

Обробка природньої мови сягає корінням у 1950-ті роки. Вже в 1950 році Алан Тьюрінг опублікував статтю під назвою "Обчислювальна техніка та інтелект", в якій запропоновано те, що зараз називається тестом Тьюрінга, як

критерій інтелекту, завдання, яке включає автоматизовану інтерпретацію та генерацію природної мови, але на той воно ще не було сформульоване як проблема, окрема від штучного інтелекту [3].

Період з 1950 - початок 1990 років ознаменований розвитком символічного НЛП.

Передумова символічного НЛП добре підсумована в експерименті з китайськими кімнатами Джона Сірла: з огляду на збірник правил (наприклад, китайський розмовник, із запитаннями та відповідними відповідями), комп'ютер імітує розуміння природної мови (або інші завдання НЛП), застосовуючи ці правила щодо даних, з якими він стикається.

Джорджтаунський експеримент у 1954 р. Передбачав повністю автоматичний переклад понад шістдесяті російських речень англійською мовою. Автори стверджували, що протягом трьох-п'яти років машинний переклад буде вирішеною проблемою. Однак реальний прогрес був набагато повільнішим, і після звіту ALPAC у 1966 р., в якому було встановлено, що десятирічні дослідження не виправдали очікувань, фінансування машинного перекладу різко скоротилося. Подальші дослідження машинного перекладу проводились лише наприкінці 1980-х років, коли були розроблені перші статистичні системи машинного перекладу

Протягом 1970-х років багато програмістів почали писати "концептуальні онтології", які структурували реальну інформацію в зрозумілі для комп'ютера дані. Прикладами є MARGIE (Schank, 1975), SAM (Cullingford, 1978), PAM (Wilensky, 1978), TaleSpin (Meehan, 1976), QUALM (Lehnert, 1977), Politics (Carbonell, 1979), and Plot Unit (Lehnert 1981).

Саме 1980-ті і початок 1990-х років відзначають розквіт символічних методів у НЛП. Основні напрямки цього часу включали дослідження синтаксичного аналізу на основі правил (наприклад, розвиток HPSG як обчислювальної операціоналізації генеративної граматики), морфологію (наприклад, дворівневу морфологію), семантику (наприклад, алгоритм Леска), посилення (наприклад, у рамках Теорії центрування) та інших областей розуміння природної мови (наприклад, у Теорії риторичної структури). Важливою подією (яка врешті призвела до статистичного повороту в 90-х роках) було зростаюче значення кількісної оцінки в цей період.

До 1980-х років більшість систем обробки природних мов базувались на складних наборах рукописних правил. Однак, починаючи з кінця 1980-х, відбулася революція в обробці природної мови із запровадженням алгоритмів машинного навчання для обробки мови. Це було зумовлено як постійним збільшенням обчислювальної потужності, так і поступовим зменшенням домінування теорій лінгвістики Хомського (наприклад, трансформаційна граMATика), теоретичні основи яких знеохочували той тип корпусної лінгвістики, який лежить в основі підходу до машинного навчання до мовної обробки.

Багато помітних ранніх успіхів щодо статистичних методів у НЛП відбулись у галузі машинного перекладу, особливо завдяки роботі в IBM Research. Ці системи змогли скористатися перевагами існуючих багатомовних текстових корпусів, вироблених Парламентом Канади та Європейським Союзом в результаті законів, що передбачають переклад усіх урядових процедур на всі офіційні мови відповідних систем управління. Однак більшість інших систем залежали від корпусів, спеціально розроблених для завдань, що реалізуються цими системами, що було (і часто продовжує бути) основним обмеженням успіху цих систем. Як результат, велика кількість досліджень зайнялася методами ефективнішого навчання на обмежених обсягах даних.

Із зростанням Інтернету з середини 1990-х років стає доступним зростаючий обсяг необроблених (без коментарів) мовних даних. Таким чином, дослідження все більше зосереджуються на алгоритмах навчання без нагляду та напівконтролю. Такі алгоритми можуть вчитися на даних, які не були анотовані вручну бажаними відповідями або використовуючи комбінацію анотованих та неанотованих даних. Як правило, це завдання набагато складніше, ніж навчання під контролем і, як правило, дає менш точні результати для певного обсягу вхідних даних. Однак існує величезна кількість анотованих даних (включаючи, серед іншого, весь вміст Всесвітньої павутини), які часто можуть компенсувати гірші результати, якщо використовуваний алгоритм має досить низьку часову складність, щоб бути практичним.

У 2010-х роках представницьке навчання та методи машинного навчання у стилі нейронних мереж набули широкого поширення в обробці природних мов, частково завдяки безлічі результатів, які показують, що такі методи можуть досягти найсучасніших результатів у багатьох завданнях з природної мови, наприклад, у моделюванні мови, синтаксичному аналізі, та багатьох інших.

### **1.1.2 Особливості роботи систем обробки природної мови**

Розглянемо базові процеси, які відбуваються у системах природної обробки мови при аналізі текстової інформації.

Токенізація (іноді - сегментація) за реченнями - це процес поділу писемної мови на речення-компоненти. Ідея виглядає досить простою. У багатьох мовах ми можемо виокремлювати речення кожен раз, коли знаходимо певний знак пунктуації – точку [4].

Але навіть в англійській мові ця задача нетривіальна, так як точка використовується і в скороченнях. Таблиця скорочень може значною мірою допомогти під час обробки тексту, щоб уникнути невірної розстановки кордонів речень. У більшості випадків для цього використовуються бібліотеки.

Розглянемо приклад токенизації на невеликому тексті:

Backgammon is one of the oldest known board games. Its history can be traced back nearly 5,000 years to archeological discoveries in the Middle East. It is a two

player game where each player has fifteen checkers which move between twenty-four points according to the roll of two dice.

Токенізація речень дає нам розподіл на 3 окремих речення:

Backgammon is one of the oldest known board games.

Its history can be traced back nearly 5,000 years to archeological discoveries in the Middle East.

It is a two player game where each player has fifteen checkers which move between twenty-four points according to the roll of two dice.

Токенізація (іноді - сегментація) за словами - це процес поділу речень на слова-компоненти. В англійській і багатьох інших мовах, що використовують ту чи іншу версію латинського алфавіту, пробіл - це непоганий роздільник слів.

Проте, можуть виникнути проблеми, якщо ми будемо використовувати тільки пробіл - в англійському складові іменники пишуться по-різному і іноді через пробіл. І тут знову нам допомагають бібліотеки.

Якщо розглянути текст з попереднього прикладу, то токенізація за словами забезпечить наступний результат:

```
['Backgammon', 'is', 'one', 'of', 'the', 'oldest', 'known', 'board', 'games', '.']
```

```
['Its', 'history', 'can', 'be', 'traced', 'back', 'nearly', '5,000', 'years', 'to', 'archeological', 'discoveries', 'in', 'the', 'Middle', 'East', '.']
```

```
['It', 'is', 'a', 'two', 'player', 'game', 'where', 'each', 'player', 'has', 'fifteen', 'checkers', 'which', 'move', 'between', 'twenty-four', 'points', 'according', 'to', 'the', 'roll', 'of', 'two', 'dice', '.']
```

Зазвичай тексти містять різні граматичні форми одного і того ж слова, також можуть зустрічатися спільнокореневі слова. Лематизації і стемінг мають на меті привести все словоформи, що зустрічаються, до однієї, нормальної словникової формі.

На прикладі одного слова це виглядає наступним чином:

dog, dogs, dog's, dogs' => dog

На прикладі цілого речення результат такий:

the boy's dogs are different sizes => the boy dog be differ size

Лематизації і стемінг - це окремі випадки нормалізації і вони відрізняються.

Стемінг - це грубий евристичний процес, який відрізає «зайве» від кореня слів, часто це призводить до втрати словотворчих суфіксів.

Лематизація - це більш тонкий процес, який використовує словник і морфологічний аналіз, щоб в результаті привести слово до його канонічної формі - лемми.

Відмінність в тому, що Стеммер (конкретна реалізація алгоритму стемінг) діє без знання контексту і, відповідно, не розуміє різницю між словами, які мають різний зміст в залежності від частини мови. Однак у Стеммера є і свої переваги: їх простіше впровадити і вони працюють швидше. Плюс, більш низька «акуратність» може не мати значення в деяких випадках.

Розглянемо кілька прикладів:

- слово good - це лема для слова better. Стеммер не побачить цей зв'язок, так як тут потрібно звіритися зі словником;

- слово play - це базова форма слова playing. Тут впораються і стемінг, і лематизація;
- слово meeting може бути як нормальною формою іменника, так і формою дієслова to meet, в залежності від контексту. На відміну від стемінгу, лематизація спробує вибрати правильну лемму, спираючись на контекст.

Стоп-слова - це слова, які викидаються з тексту до чи після обробки тексту. При застосуванні машинного навчання до текстів, такі слова можуть додати багато шуму, тому необхідно позбавлятися від нерелевантних слів.

Під терміном стоп-слова зазвичай розуміють артиклі, вигуки, сполучники і т.д., які не несуть смислового навантаження. При цьому треба розуміти, що не існує універсального списку стоп-слів, все залежить від конкретного випадку.

Розглянемо наступне речення:

Backgammon is one of the oldest known board games.

Після фільтрації на стоп-слова воно буде мати наступний вигляд:

['Backgammon', 'one', 'oldest', 'known', 'board', 'games', '.']

Також для додаткової фільтрації можна використовувати регулярні вирази. Наприклад, можна прибрати всі символи, які не є словами. У багатьох випадках пунктуація не потрібна і її легко прибрати за допомогою регулярних виразів.

Розглянемо приклад речення зі знаками пунктуації.

The development of snowboarding was inspired by skateboarding, sledding, surfing and skiing.

З допомогою регулярних виразів після обробки воно може мати наступний вигляд:

'The development of snowboarding was inspired by skateboarding sledding surfing and skiing '

Алгоритми машинного навчання не можуть безпосередньо працювати з сирим текстом, тому необхідно конвертувати текст в набори цифр (вектори). Це називається витяганням ознак.

Мішок слів - це популярна і проста техніка вилучення ознак, яка використовується при роботі з текстом. Вона описує входження кожного слова в текст.

Щоб використовувати модель, потрібно:

- визначити словник відомих слів (токенів);
- вибрати ступінь присутності відомих слів.

Будь-яка інформація про порядок або структуру слів ігнорується. Ось чому це називається мішком слів. Ця модель намагається зрозуміти, чи зустрічається знайоме слово в документі, але не знає, де саме воно зустрічається.

Інтуїція підказує, що схожі документи мають схожу вміст. Також, завдяки вмісту, можна дізнатися дещо про сенс документа.

Складність цієї моделі в тому, як визначити словник і як підрахувати входження слів.

Коли розмір словника збільшується, вектор документа теж зростає.

У деяких випадках, може бути неймовірно великий обсяг даних і тоді вектор може складатися з тисяч або мільйонів елементів. Більш того, кожен документ може містити лише малу частину слів зі словника.



Як наслідок, у векторному поданні буде багато нулів. Вектори з великою кількістю нулів називаються розрідженими векторами (sparse vectors), вони вимагають більше пам'яті і обчислювальних ресурсів.

Однак можна зменшити кількість відомих слів, при використанні цієї моделі, щоб знизити вимоги до обчислювальних ресурсів. Для цього можна використовувати ті ж техніки, що вже розглядалися до створення мішка слів:

- ігнорування регістра слів;
- ігнорування пунктуації;
- викидання стоп-слів;
- приведення слів до їх базовим формам (лематизації і стемінг);
- виправлення неправильно написаних слів.

Інший, більш складний спосіб створення словника - використовувати згруповані слова. Це змінить розмір словника і дасть мішку слів більше деталей про документ. Такий підхід називається «N-грами».

N-грами це послідовність будь-яких сутностей (слів, букв, чисел, цифр і т.д.). В контексті мовних корпусів, під N-грамою зазвичай розуміють послідовність слів. Юніграма це одне слово, біграма це послідовність двох слів, триграма - три слова і так далі. Цифра N позначає, скільки згрупованих слів входить в N-граму. У модель потрапляють не всі можливі N-грами, а тільки ті, що фігурують в корпусі.

Розглянемо наступне речення:

The office building is open today

Його біграми матимуть наступний вигляд:

- the office;
- office building;
- building is;
- is open;
- open today.

Біграми пропонують більш дієвий підхід, ніж мішок слів.

У частотного скорингу (оцінки) є проблема: слова з найбільшою частотністю мають, відповідно, найбільшу оцінку. У цих словах може бути не так багато інформаційного виграшу для моделі, як в менш частих словах. Один із способів виправити ситуацію - знижувати оцінку слова, яке часто зустрічається у всіх подібних документах. Це називається TF-IDF.

TF-IDF (скорочення від term frequency - inverse document frequency) - це статистична міра для оцінки важливості слова в документі, який є частиною колекції або корпусу.

Скоринг по TF-IDF зростає пропорційно частоті появи слова в документі, але це компенсується кількістю документів, що містять це слово.

Формула скорингу для слова X в документі Y зображена на рисунку 1.1:

$$w_{x,y} = tf_{x,y} \times \log \left( \frac{N}{df_x} \right)$$

**TF-IDF**

Term  $x$  within document  $y$

$tf_{x,y}$  = frequency of  $x$  in  $y$   
 $df_x$  = number of documents containing  $x$   
 $N$  = total number of documents

Рисунок 1.1 – Формула скорингу для методу TF-IDF

TF (term frequency - частота слова) - відношення числа входжень слова до загальної кількості слів документа.

$$TF(\text{term}) = \frac{\text{Number of times term appears in a document}}{\text{Total number of items in the document}}$$

IDF (inverse document frequency - зворотна частота документа) - інверсія частоти, з якою деяке слово зустрічається в документах колекції.

$$IDF(\text{term}) = \log \left( \frac{\text{Total number of documents}}{\text{Number of documents with term in it}} \right)$$

У підсумку, обчислити TF-IDF для слова term можна так:

$$TFIDF(\text{term}) = TF(\text{term}) * IDF(\text{term})$$

Для прикладу розглянемо наступний текст:

I like this movie, it's funny.

I hate this movie.

This was awesome! I like it.

Nice one. I love it.

Після застосування методу TF-IDF до данного тексту у результаті отримаємо наступний вектор (рис. 1.2).

	awesome	funny	hate	it	like	love	movie	nice	one	this	was
0	0.000000	0.571848	0.000000	0.365003	0.450852	0.000000	0.450852	0.000000	0.000000	0.365003	0.000000
1	0.000000	0.000000	0.702035	0.000000	0.000000	0.000000	0.553492	0.000000	0.000000	0.448100	0.000000
2	0.539445	0.000000	0.000000	0.344321	0.425305	0.000000	0.000000	0.000000	0.000000	0.344321	0.539445
3	0.000000	0.000000	0.000000	0.345783	0.000000	0.541736	0.000000	0.541736	0.541736	0.000000	0.000000

Рисунок 1.2 – Вектор для тексту сформований методом TF-IDF

### 1.1.3 Задачі, які виконують системи обробки природної мови

НЛП представляє автоматичну обробку такої людської мови, як мова чи текст, і хоча сама концепція захоплює, справжня цінність цієї технології походить від випадків використання [5].

НЛП може допомогти у виконанні багатьох завдань, а сфери застосування просто збільшуються щодня. Згадаймо кілька прикладів:

НЛП дозволяє розпізнавати та прогнозувати захворювання на основі електронних медичних карт та власного мовлення пацієнта. Ця здатність досліджується в умовах здоров'я, які переходять від серцево-судинних захворювань до депресії і навіть шизофренії. Наприклад, Amazon Comprehend Medical - це послуга, яка використовує НЛП для вилучення захворювань, ліків та результатів лікування із записів пацієнтів, звітів про клінічні випробування та інших електронних медичних карт.

Розуміння настроїв споживачів має важливе значення для розробки бізнес-стратегії. Програмне забезпечення, засноване на NLP, може бути використано для аналізу вмісту в соціальних мережах, оглядів продуктів та контенту клієнтів для розробки статистичних даних.

Аналіз конкурентів зазвичай проводиться під час заснування бізнесу або виходу на новий ринковий сектор. Аналіз може надати глибше розуміння ринку, ким будуть конкуренти та хто потенційні клієнти [6].

Механізми, що працюють на основі НЛП, можуть значно спростити та автоматизувати процес сканування конкурентного середовища. Доступні інструменти для моніторингу конкуренції, як правило, шляхом сканування в Інтернеті статей про цей сектор та використання інформації для подачі модуля NLP, який виявляє семантичні взаємозв'язки між компаніями.

Винахідник IBM розробив когнітивного помічника, який працює як персоналізована пошукова система, вивчаючи все про себе, а потім нагадуючи ім'я, пісню або все, що ви не можете згадати, коли вам це потрібно.

Такі компанії, як Yahoo та Google, фільтрують та класифікують електронні листи за допомогою NLP, аналізуючи текст у електронних листах, що проходять через їх сервери, та зупиняючи спам ще до того, як вони потрапляють у вашу поштову скриньку.

Щоб допомогти ідентифікувати фальшиві новини, група NLP у Массачусетському технологічному інституті розробила нову систему, яка визначає, чи є джерело точним чи політично упередженим, виявляючи, чи можна довіряти джерелу новин чи ні.

У США проживає понад 110 мільйонів користувачів голосових асистентів, а голосові асистенти, такі як Alexa, Siri, Cortana та Google Assistant, отримують мільйони запитів щомісяця. Десятки мільйонів користувачів також придбали голосові системи, такі як Amazon Echo та Google Home, а голосові помічники підприємств є одними з найбільш швидкозростаючих категорій продуктів.

У секторі голосових асистентів NLP використовується для перекладу розмови в текст, семантичного узгодження з баз знань та повернення відповідей після перекладу тексту в мову. Використовуючи розумні колонки як маркетинговий канал, компанії можуть активізувати маркетингову діяльність та залучити аудиторію.

На сцену виходять нові бренди з можливостями, які можна додати до пристроїв Alexa через магазин Skills. Alexa Skills можна додати до існуючого програмного забезпечення за допомогою набору заздалегідь визначених голосових команд. Також можна створювати нові голосові команди на основі вимог конкретних бізнес-кейсів. Після створення Alexa Skills можна ділитися з магазином Alexa Skills Store і передавати її користувачам по всьому світу. Для електронної комерції це означає, що люди, які мають навички Alexa на своїх пристроях, можуть здійснювати голосові покупки продуктів чи послуг.

Маючи розуміння того, що відбувається, і про що говорять люди, може бути дуже цінним для фінансових трейдерів. НЛП використовується для відстеження новин, звітів, коментарів щодо можливих злиттів між компаніями, і все це потім можна включити в торговий алгоритм для отримання величезних прибутків. Запам'ятайте: купуйте чутки, продавайте новини.

НЛП також використовується як на етапах пошуку, так і на етапі відбору персоналу, виявляючи навички потенційних працівників, а також визначаючи перспективи до того, як вони стануть активними на ринку праці.

Працюючи на основі технології IBM Watson NLP, LegalMation розробив платформу для автоматизації рутинних судових процесів та допомагає юридичним командам економити час, знижувати витрати та переносити стратегічний фокус.

НЛП особливо бурхливо розвивається у галузі охорони здоров'я. Ця технологія покращує надання медичної допомоги, діагностику захворювань та знижує витрати, в той час як організації охорони здоров'я проходять все більш широке впровадження електронних медичних карт. Той факт, що клінічну документацію можна вдосконалити, означає, що пацієнтів можна краще зрозуміти та отримати вигоду завдяки кращому медичному обслуговуванню. Метою має бути оптимізація їх досвіду, і кілька організацій вже працюють над цим.

Кількість публікацій, що містять речення «обробка природною мовою» у PubMed у період 1978–2018 рр. Станом на 2018 рік PubMed включав понад 29 мільйонів посилань на біомедичну літературу (рис. 1.3).

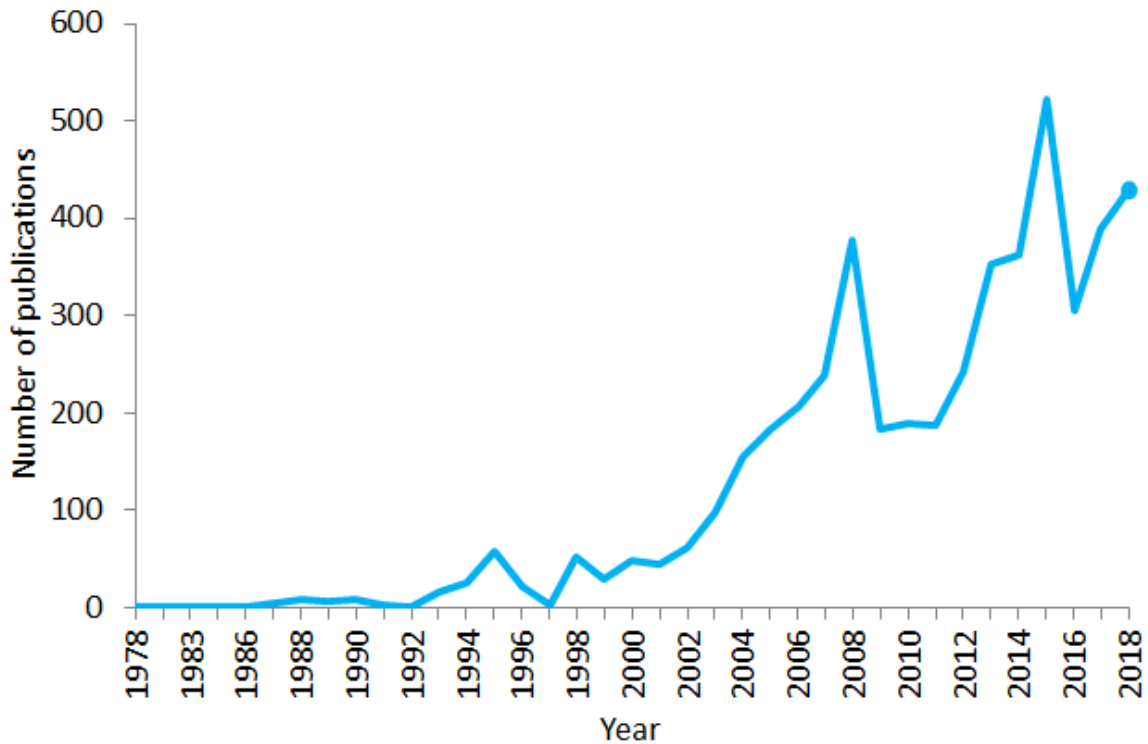


Рисунок 1.3 – Кількість публікацій про NLP у різні роки

Такі компанії, як Winterlight Labs, роблять значні вдосконалення у лікуванні хвороби Альцгеймера, відстежуючи когнітивні порушення за допомогою мови, а також можуть підтримувати клінічні випробування та дослідження широкого кола розладів центральної нервової системи. Слідуючи подібному підходу, Стенфордський університет розробив Woebot, терапевта-чат-бота з метою допомоги людям з тривогою та іншими розладами.

Але навколо цього питання виникають серйозні суперечки. Пару років тому Microsoft продемонструвала, що, проаналізувавши великі зразки запитів пошукових систем, вони можуть ідентифікувати користувачів Інтернету, які страждають на рак підшлункової залози ще до того, як вони отримали діагноз захворювання. Як реагують користувачі на такий діагноз? А що було б, якби вас перевірили як хибнопозитивний? (мається на увазі, що у вас може бути діагностовано захворювання, навіть якщо у вас його немає). Це нагадує випадок Google Flu Trends, який у 2009 році був оголошений як спрогнозувати грип, але пізніше зник через низьку точність та неможливість досягти прогнозованих показників.

NLP може бути ключем до ефективної клінічної підтримки в майбутньому, але в короткостроковій перспективі все ще існує багато проблем.

## **1.2 Проблематика задач, які вирішуються сервісами оцінки тональності тексту**

Аналіз настрою - це тип досліджень тексту, ще відомий як майнінг. Він застосовує поєднання статистичних даних, обробки природної мови (NLP) та машинного навчання для виявлення та вилучення суб'єктивної інформації із текстових файлів, наприклад, почуттів, думок, суджень або оцінок рецензента щодо певної теми, події чи компанії. Цей тип аналізу також відомий як аналіз думок (з акцентом на видобуток) або афективний рейтинг. Деякі фахівці також використовують терміни класифікації та вилучення настроїв. Незалежно від назви, мета аналізу настроїв однакова: дізнатися думку користувача чи аудиторії щодо цільового об'єкта, проаналізувавши величезну кількість тексту з різних джерел.

Існує можливість аналізувати текст на різних рівнях деталізації, і рівень деталізації залежить від обраних цілей. Наприклад, можливо визначити середній емоційний тон групи оглядів, щоб знати, якому відсотку клієнтів сподобався нова колекція одягу. Якщо потрібно знати, що відвідувачам подобається чи не подобається у певному одязі та чому, або чи порівнюють вони його з подібними предметами інших брендів, потрібно проаналізувати кожне рецензійне речення з акцентом на конкретні аспекти та використання чи конкретні ключові слова.

Залежно від масштабу можна використовувати два типи аналізу: грубозернистий та дрібнозернистий. Грубозернистий аналіз дозволяє визначити настрої на рівні документа або речення. А за допомогою дрібнозернистого аналізу ви можете отримати почуття в кожній частині речення.

### **1.2.1 Різновиди задач, що вирішуються сервісами оцінки тональності тексту**

#### **1.2.1.1 Задача розвитку бренду**

Розуміння настроїв споживачів має важливе значення для розробки бізнес-стратегії. Програмне забезпечення, засноване на NLP, може бути використано для аналізу вмісту в соціальних мережах, оглядів продуктів та вмісту клієнтів для розробки статистичних даних.

Належний інструмент для аналізу настроїв пропонує чітку візуалізацію даних, щоб існувала можливість зробити загальні висновки щодо голосу клієнта. Для прикладу можна розглянути аналіз настроїв щодо епізодів сервіалу «Гри престолів» (рис. 1.4), і прогрес настрою очевидний: поєднання позитивних чи негативних емоцій на основі того, що сталося [7].

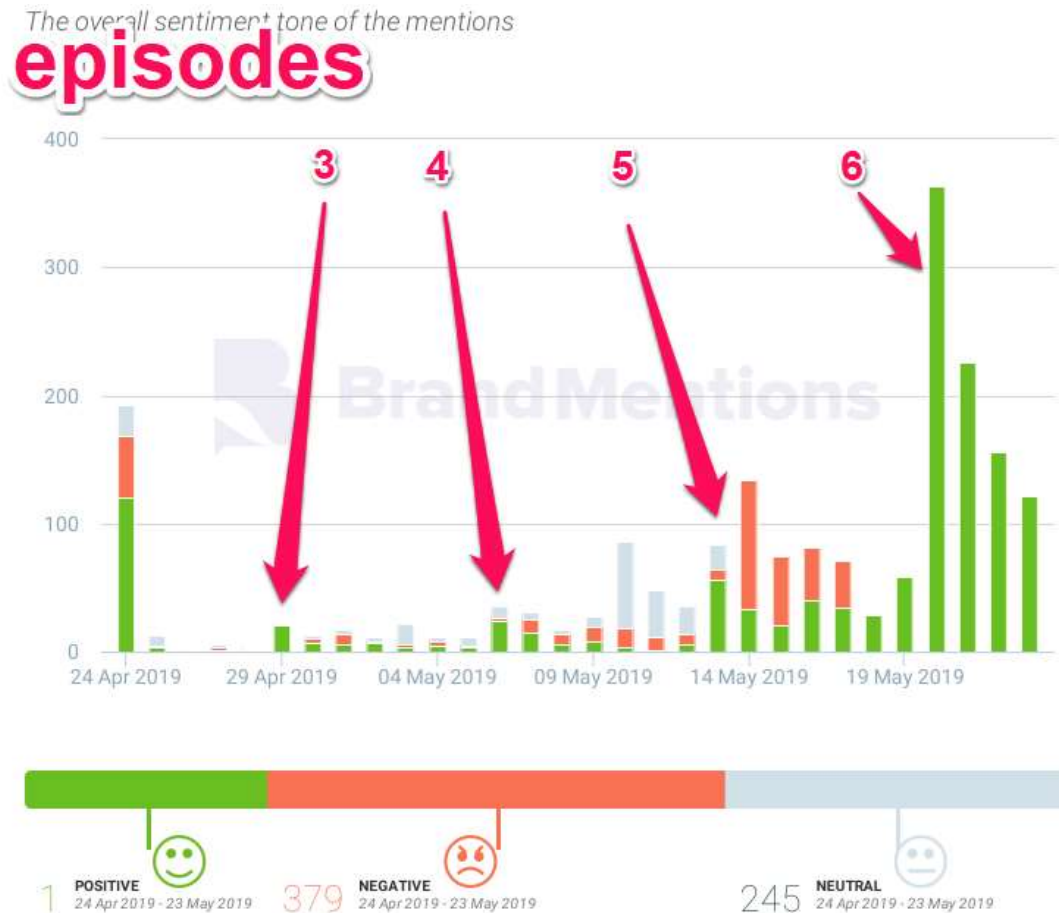


Рисунок 1.4 – Гістограма аналізу настроїв

Аналіз настрою використовується для розгляду контекстів позитивних та негативних відгуків, спрямованих на бренд. Алгоритми працюють шляхом побудови моделей аналізу настроїв з коментарів. Класифікатори витягуються за допомогою найбільш часто вживаних слів та перегляду відомих негативних та позитивних фраз. Потім кожній частині інформації присвоюється значення, як правило, число, яке вказує на позитивне, негативне чи нейтральне почуття. Маючи на руках такі дані, маркетологи можуть приймати більш обґрунтовані рішення при розробці стратегій та прогнозуванні попиту на товари та послуги.

Соціальні платформи заповнені людьми, що заявляють про себе. Соціальні мережі спонукали людей бути відкритими, чесними, прямими, а іноді навіть критичними до всього, навіть до товарів брендів. Це один із відповідних критеріїв для вимірювання присутності бренду в соціальних мережах та допомога в оцінці зусиль витрачених на просування бренду.

Ось як можна застосувати системи аналізу настроїв у моніторингу соціальних мереж:

- аналізувати твіти та публікації на facebook, щоб побачити аналіз настроїв у часових межах для конкретної аудиторії;
- запустити аналіз настроїв на всі ваші згадки та впорядкувати терміново;





застосована до ряду передбачувальних завдань. Від рівня макрокампанії аж до мікроформулювання цільової сторінки, аналіз настрою дозволяє точно налаштувати повідомлення на найбільший вплив.

Наприклад, сектор фінансових послуг вже вивчає це. Поєднуючи здатність машинного навчання аналізувати настрої корпоративних звітів та обробляти історичні дані в режимі реального часу, фінансові установи дедалі більше можуть покладатися на ШІ для прийняття швидких рішень, які керують ринком. Незалежно від того, чи прагнете ви побудувати довгострокові тенденції чи зрозумієте, як зробити, щоб один фрагмент вмісту мав миттєвий вплив, аналіз настроїв може надати додатковий вимір зусиллям (рис. 1.5) [8].

Дослідження ринку розглядається як систематичний підхід, що включає збір даних та аналіз даних щодо будь-яких відповідних питань, пов'язаних з маркетингом. У маркетингу дослідження використовуються з різними цілями, включаючи: отримання уявлення про ставлення та переконання споживачів, вимірювання задоволеності споживачів, перевірку ефективності реклами тощо. Деякі великі компанії мають власні відділи дослідження ринку, тоді як менші компанії зазвичай передають цю функцію спеціалісти з досліджень (Kotler, 2010).

Дослідження проводяться двома основними способами: якісним та кількісним. При якісному підході дослідник висуває претензії на знання насамперед на конструктивістські перспективи (тобто множинні значення індивідуального досвіду, значення, побудовані соціально та історично, з метою розвитку теорії чи моделі) або перспективи адвокації / участі (тобто політичні, орієнтовані на проблеми, спільні дії чи орієнтовані на зміни) або обидва (Кресвелл, 2007, с.18).

Якісне дослідження передбачає з'ясування того, що люди думають і як вони почуваються - або в будь-якому разі, що вони кажуть, що думають і як кажуть, що відчують. Цей вид інформації є суб'єктивним. Він включає почуття та враження, а не цифри. З іншого боку, кількісні дослідження зосереджені на вимірюванні об'єктивного факту. Ключовим для проведення кількісного дослідження є визначення змінних, що представляють інтерес, і значною мірою відчуття відокремленості дослідником у збиранні даних. Кількісне дослідження аналізує дані за допомогою статистичних даних і спирається на великі вибірки для узагальнення тверджень [9].

Аналіз настрою є корисним інструментом для вирішення проблем триангуляції в Інтернет-середовищі. У поєднанні з якісними дослідженнями аналіз настроїв може бути використаний як інструмент, який сприяє підвищенню строгості та структури до іншого, гнучкого та суб'єктивного процесу збору та аналізу даних.

Як альтернатива, якщо інтегрувати з кількісними дослідженнями, аналіз настроїв сприяє глибокому глибокому розумінню небажаних думок та емоцій, тим самим полегшуючи більш значуще розуміння будь-якого явища. Застосовуючи технології машинного навчання, аналіз настрою дає можливість надати систематичний підхід до проектування змішаних методів. Ми

стверджуємо, що за допомогою аналізу настроїв маркетологу пропонується багата можливість отримувати значущі та проникливі відгуки про почуття, думки, думки та настрої клієнтів у режимі реального часу.

Аналіз настрою забезпечує швидшу, простішу та дешевшу альтернативу традиційним якісним методам дослідження ринку, таким як спостереження, інтерв'ю та навіть етнографія, а також надає інформацію в режимі реального часу. У той же час він пропонує переваги традиційних кількісних методів, включаючи вимірюваність та об'єктивність. Дані також збираються повністю ненав'язливо порівняно з методами, що використовуються як в якісних, так і в кількісних дослідженнях.

### **1.2.1.3 Задача аналізу конкурентів**

Аналіз конкурентів зазвичай проводиться під час заснування бізнесу або виходу на новий ринковий сектор. Аналіз може надати глибше розуміння ринку, ким будуть конкуренти та хто потенційні клієнти.

Механізми, що працюють на основі НЛП, можуть значно спростити та автоматизувати процес сканування конкурентного середовища. Доступні інструменти для моніторингу конкуренції, як правило, шляхом сканування в Інтернеті статей про цей сектор та використання інформації для подачі модуля NLP, який виявляє семантичні взаємозв'язки між компаніями.

Існує можливість відстежувати та досліджувати, як суспільство оцінює конкурентів, так само, як і аналізувати їх ставлення до іншого бізнесу. Що споживачі найбільше цінують у інших гравцях галузі, що у конкурентів бракує чи робиться ними неправильно, які канали використовують клієнти для взаємодії з іншими компаніями – використовувати ці знання можна для вдосконалення комунікаційних та маркетингових стратегій, загального сервісу та надання послуг та продуктів, які клієнти оцінять [10].

Аналіз конкуренції, який включає аналіз настроїв, також може допомогти зрозуміти слабкі та сильні сторони і, можливо, знайти способи виділитися.

### **1.2.1.4 Задача аналізу продукту**

Кожен підприємець бажає побачити шанувальників, які стоять у чергах, поки магазини відкриються, щоб вони могли забігти всередину, захопити цей новий продукт і стати одним з перших гордих власників у світі. Як вивести бажаний товар на ринок? Єдиний підхід - запитати людей, що вони хочуть. Успішні компанії створюють мінімально життєздатний продукт (MVP), збирають зворотній зв'язок на ранніх термінах, постійно вдосконалюючи продукт навіть після його випуску. Дані про відгуки надходять з опитувань,

соціальних медіа та форумів та взаємодії зі службою підтримки. Виникають питання, як визначити, які групи споживачів запитувати, проаналізувати цей океан даних та класифікувати огляди.

В подібних ситуаціях аналіз настроїв стає в нагоді. Це дозволяє дізнатись про переваги та недоліки товару. Наприклад, студент Університету штату Оклахома проаналізував відгуки Amazon про дві моделі телефонів Samsung (Galaxy S6 і Galaxy S7) та два пристрої від Apple (iPhone 6 та iPhone 7), щоб з'ясувати, чому клієнти віддають перевагу одній торговій марці перед іншою. Він з'ясував, що користувачі, пріоритетами яких є надійний акумулятор і якісний екран, вибирають телефони Samsung. А клієнти, яким більше цікавий дизайн та камери, купують iPhone.

Фільтруючи коментарі за темами та настроями, також можна дізнатися, які функції необхідні, а які потрібно усунути. Озброївшись результатами аналізу настроїв, команда розробників продуктів точно знатиме, як доставити товар, який споживачі купуватимуть і насолоджуватимуться.

На думку покупців щодо продуктів впливає безліч різних факторів. Вони також змінюються на льоту, як покращення функцій, підвищення цін та маркетингові кампанії. Всі ці фактори викликають у користувачів різноманітні емоції.

Перш ніж використовувати модель аналізу настроїв, необхідно знайти відгуки про товари, які необхідно проаналізувати.

Відгуки про товари є скрізь в Інтернеті. Можна натрапити на назву іншого бренду в Capterra, G2Crowd, Sifter, Yelp, Amazon та Google Play, щоб назвати лише декілька, тому про збір даних вручну, не може бути й мови, якщо щохвилини з'являються нові відгуки [11].

На сьогоднішній день для цієї операції не потрібно здійснювати копіювання та вставку. Веб-скрапінг допомагає автоматизувати та впорядкувати весь цей процес. Веб-скрапінг - це набір інструментів, що використовуються для збору інформації з Інтернету. Ці інструменти імітують, як люди користуються Інтернетом, щоб збирати конкретні дані з різних веб-сайтів. По суті, вони автоматично знаходять те, що вам довелося б скопіювати та вставити вручну з будь-якого веб-сайту.

Менеджери продуктів повинні постійно стежити за тим, як користувачі реагують на товари та зміни, які вони зазнають. Аналіз настрою корисний і в цій галузі.

### **1.2.1.5 Задача аналізу заохоченості працівників**

Деякі організації виходять за рамки використання аналізу настроїв для дослідження ринку або оцінки досвіду споживачів, застосовуючи його внутрішньо для процесів, пов'язаних з персоналом. Ці компанії вимірюють задоволеність співробітників, виявляють фактори, які знеохочують членів

команди і врешті-решт знижують ефективність роботи компанії. Фахівці автоматизують аналіз опитувань працівників за допомогою програмного забезпечення SA, що дозволяє їм швидше вирішувати проблеми та проблеми. Менеджери з управління персоналом можуть виявляти та відстежувати загальний тон відповідей, групувати результати за підрозділами та ключовими словами, а також перевіряти, чи змінилися настрої співробітників з часом чи ні.

Аналіз заохоченості працівників - це використання наявних даних для розкриття повної потужності співробітників компанії. Це забезпечує розуміння шляхом аналізу всіх даних ваших співробітників і перетворення результатів на щось відчутне і, якщо це зроблено добре, на практику.

Аналіз заохоченості працівників запровадив нову дисципліну у відділі кадрів, що дозволяє директорам з управління персоналом приймати обґрунтовані рішення, базуючись на фактах, а не лише на теорії. Кожна велика корпорація може отримати величезну вигоду через це.

Багато чого, до чого прагне аналіз заохоченості працівників, - це незначні прибутки, - настільки широко популяризовані сером Дейвом Брейлсфордом. Отримання 1% покращення тут, ще 1% покращення в іншій сфері. Зрештою все це складається у відчутні поліпшення.

Аналіз настрою виводить моніторинг настрою співробітників на новий рівень за допомогою можливостей моніторингу в режимі реального часу. Наприклад, члени команди можуть щомісяця заповнювати анкети з одним запитом, щоб оцінити умови їх роботи. Вони також можуть аналізувати свої публікації в соціальних мережах, щоб виявити можливий зв'язок між їхнім душевним станом та робочим життям.

## **1.2.2 Визначення проблематики сучасних сервісів оцінки тональності тестів**

### **1.2.2.1 Проблема виявлення сарказму**

У саркастичному тексті люди висловлюють свої негативні настрої позитивними словами. Цей факт дозволяє сарказму легко обманювати моделі аналізу настроїв, якщо вони спеціально не розроблені з урахуванням його можливості [12].

Сарказм найчастіше трапляється в створеному користувачем вмісті, такому як коментарі Facebook, твіти тощо. Виявлення сарказму в аналізі настроїв дуже важко здійснити, не маючи хорошого розуміння контексту ситуації, конкретної теми та навколишнього середовища.

Це може бути важко зрозуміти не лише для машини, а й для людини. Постійна варіація слів, що використовуються в саркастичних реченнях, ускладнює успішну підготовку моделей аналізу настроїв. Щоб зробити

доступним сарказм, двоє людей повинні ділитися спільними темами, інтересами та історичною інформацією.

Спочатку розглянемо сарказм з точки зору лінгвістики, де сарказм широко вивчається. В одному з найбільш цитованих досліджень у цій галузі автор Елізабет Кемп пропонує наступні чотири типи сарказму:

- твердження: Сарказм, схоже, є пропозицією, яка не стосується сентиментів, але має в собі неявні настрої;
- вбудований: Сарказм має вкорінені почуття невідповідності у формі самих слів і фраз;
- префікс "Подобається": Подібна фраза забезпечує неявне заперечення аргументу;
- невербальні: немовні акти (мова тіла, жести), що сприяють сарказму.

Чотири типи сарказму Елізабет Кемп: пропозиційний ("Це виглядає як ідеальний план!"), Вбудований ("я люблю, коли мене ігнорують"), "префікс" ("Як ті хлопці вірять слову, яке вони говорять") і невербальні "(знизує плечима) Насправді дуже корисно!" (рис. 1.6).

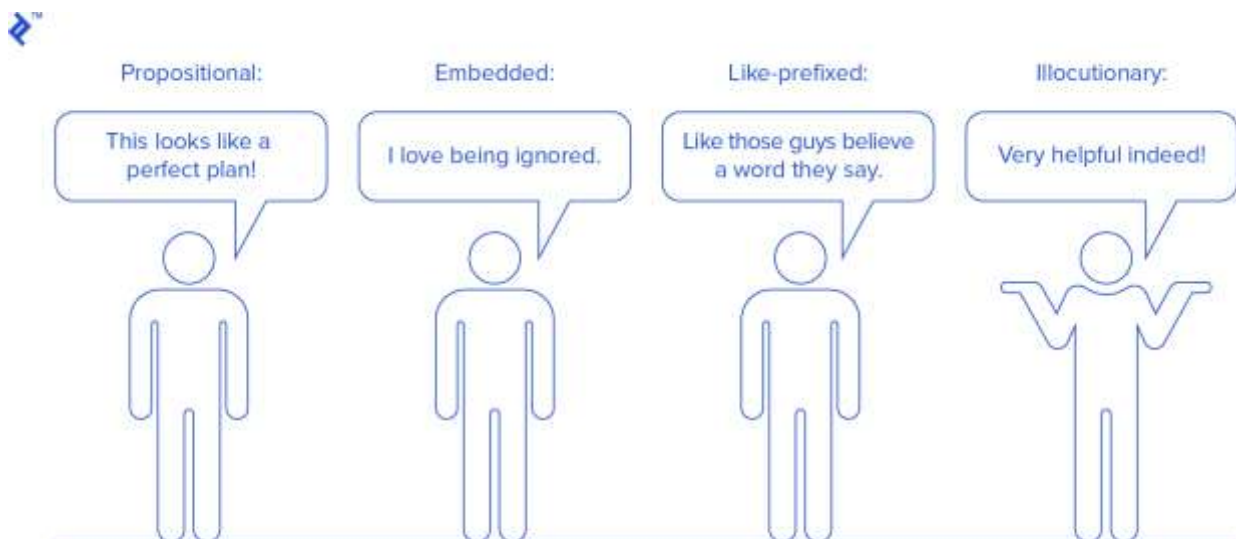


Рисунок 1.6 – Типи сарказму

Дослідження Кемпа було опубліковане в 2012 році. У 2017 році дослідники зі Стенфордського університету анонсували власне досить цікаве дослідження «Мати 2 години, щоб написати статтю - це весело!»: Виявлення сарказму в числових частинах тексту, де вони говорили про інший тип сарказму, який називається числовим сарказм. Чисельний сарказм дуже часто зустрічається в соціальних мережах. Ідея цього пов'язана зі зміною числових значень, які потім впливають на полярність тексту. Наприклад:

"Цей телефон має надзвичайну резервну батарею за 38 годин".  
(Несаркастичний)

"Цей телефон має надзвичайну резервну батарею на 2 години."  
(Саркастично)

"На вулиці +25, і мені так жарко". (Несаркастичний)

"Надворі -25, і мені так жарко". (Саркастично)

"Ми їхали так повільно --- лише 20 км / год." (Несаркастичний)

"Ми їхали так повільно --- лише 160 км / год." (Саркастично)

Як бачимо, ці речення відрізняються лише кількістю вживаних - отже, числовим сарказмом.

Існують різні підходи до автоматичного виявлення сарказму, зокрема:

- на основі правил;
- статистичний;
- алгоритми машинного навчання;
- глибоке навчання.

Підходи, засновані на глибокому навчанні, набирають популярності. Кумар, Сомані та Бхаттачарія зробили висновок у 2017 році, що конкретна модель глибокого навчання (архітектура CNN-LSTM-FF) перевершує попередні підходи, досягаючи найвищого рівня точності для чисельного виявлення сарказму.

Але глибокі нейронні мережі (DNN) були не тільки найкращими для чисельного сарказму - вони також перевершували інші підходи детектора сарказму загалом. Гош і Віл у своїх роботах за 2016 рік використовують комбінацію згорткової нейронної мережі, мережі довгострокової короткочасної пам'яті (LSTM) та DNN. Вони порівнюють свій підхід із рекурсивними машинами векторної підтримки (SVM) і роблять висновок, що їхня архітектура глибокого навчання є покращенням порівняно з такими підходами.

### 1.2.2.2 Проблема виявлення заперечень

У лінгвістиці заперечення - це спосіб змінити полярність слів, фраз і навіть речень. Дослідники використовують різні лінгвістичні правила, щоб визначити, чи відбувається заперечення, але також важливо визначити коло слів, на які впливають заперечні слова.

Немає фіксованого розміру для обсягу слів, що зазнають впливу. Наприклад, у реченні "Шоу не було цікавим" сфера дії - це лише наступне слово після слова заперечення. Але для таких речень, як «Я не називаю цей фільм комедійним фільмом», ефект заперечення слова «не» діє до кінця речення. Початкове значення слів змінюється, якщо позитивне чи негативне слово потрапляє в сферу заперечення - у цьому випадку буде повернута протилежна полярність.

Найпростіший підхід до вирішення заперечень у реченні, який використовується в більшості найсучасніших методів аналізу настроїв, полягає у позначенні як заперечених усіх слів із сигналу заперечення до наступної

пунктуаційної лексеми. Ефективність моделі заперечення може бути змінена через специфічну конструкцію мови в різних контекстах.

Існує кілька форм вираження негативної думки в реченнях:

- заперечення може бути морфологічним, якщо воно позначається префіксом (“dis-”, “non-”) або суфіксом (“-less”);
- заперечення може бути неявним, оскільки в «з цим вчинком це буде його перший і останній фільм» - воно несе негативні настрої, але негативні слова не використовуються;
- заперечення може бути явним, як у "це не добре".

Наявність зразків з різними типами описаних заперечень підвищить якість набору даних для навчання та тестування моделей класифікації настроїв у межах заперечення. Згідно з останніми дослідженнями рекурентних нейронних мереж (RNN), різні архітектури моделей LSTM перевершують усі інші підходи при виявленні типів заперечень у реченнях.

У статті Вплив заперечення в аналізі настроїв модель аналізу настроїв оцінила 500 оглядів, зібраних з Amazon та Trustedreviews.com. Автори показують порівняння моделей з виявленням заперечень та без нього. Їх оцінка демонструє, як врахування заперечення може значно підвищити точність моделі.

### 1.2.2.3 Проблема неоднозначності слів

Невизначеність слів - це ще одна пастка, з якою зтикаються дослідники, працюючи над проблемою аналізу настроїв. Проблема двозначності слів полягає в неможливості визначити полярність заздалегідь, оскільки полярність деяких слів сильно залежить від контексту речення.

Підходи до аналізу настроїв на основі лексикону популярні серед існуючих методів. Лексикон думок містить слова думки з їх значенням полярності. В Інтернеті є кілька лексиконів громадської думки: SentiWordNet, General Inquirer та SenticNet, серед інших. Оскільки полярність слова різниться в різних сферах, неможливо розробити універсальний лексикон думок, який має полярність для кожного слова. Наприклад:

"Історія непередбачувана".

"Кермо непередбачуване".

Ці два приклади показують, як контекст впливає на думки слів думки. У першому прикладі слово полярність "непередбачуваний" прогнозується як позитивне. У другому - полярність цього ж слова є негативною.

#### 1.2.2.4 Проблема багатополярності

Іноді дане речення чи документ - або будь-яка одиниця тексту, яку необхідно проаналізувати - матиме багатополярність. У цих випадках наявність лише загального результату аналізу може ввести в оману, дуже схоже на те, як середній показник може іноді приховувати цінну інформацію про всі цифри, які в нього потрапили.

Якщо уявити, як автори говорять про різних людей, товари чи компанії (або їх аспекти) у статті чи огляді. Загальноприйнято, що в межах фрагмента тексту деякі предмети зазнають критики, а деякі - похвали.

Тут у загальній полярності настрою буде бракувати ключової інформації. Ось чому необхідно вилучити всі сутності або аспекти у реченні з присвоєними мітками настроїв і обчислити загальну полярність лише за потреби.

Розглянемо приклад, який складається з декількох полярностей: "Якість звуку мого нового ноутбука настільки класна, але кольори дисплея не надто хороші".

Деякі моделі аналізу настроїв дадуть цьому реченню негативну або нейтральну полярність. Для вирішення таких ситуацій модель аналізу настрою повинна призначати полярність кожному аспекту речення.

### 1.3 Висновки до розділу 1

У даному розділі проводиться аналіз сучасного стану проблеми оцінки тональності текстів, відбувається постановка задач для дипломної роботи з дослідження сервісів оцінки текстової інформації.

Для дослідження сучасного стану проблеми розглядаються напрями обробки природної мови та її підвиду оцінки тональності тексту. Проводиться огляд ряду задач, які вирішуються застосуванням сучасних сервісів, побудованих на алгоритмах зазначених вище технологій. На сьогоднішній день, системи активно використовуються для бізнес задач та забезпечують значне підвищення ефективності розвитку бізнесу.

Проводиться огляд основних критичних моментів, які негативно впливають на рівень точності сучасних систем оцінки тональності тексту. У цих напрямках активно проводяться дослідницькі роботи, однак, у повному обсязі дані проблеми досі лишаються не вирішеними.

У зв'язку з цим представлена робота спрямована на дослідження роботи сервісів оцінки тональності тексту, які представлені на сьогоднішній день широкому загалу для вирішення бізнес-цілей. Це дозволить побачити ефективність представлених систем та оцінити коректність їх роботи у реальних умовах.



## **2 ПЛАНУВАННЯ ЕКСПЕРИМЕНТУ ДЛЯ ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ СЕРВІСІВ ОЦІНКИ ТОНАЛЬНОСТІ ТЕКСТУ**

### **2.1 Загальний опис проблеми оцінки тональності тексту**

Оцінка тональності текстової інформації будується на алгоритмах, які у свою чергу базуються на моделі обробки природної мови Natural Language Processing (NLP). Вона представляє собою загальний напрям інформатики, штучного інтелекту та математичної лінгвістики. Вивчає проблеми комп'ютерного аналізу та синтезу природної мови. Стосовно штучного інтелекту аналіз означає розуміння мови, а синтез — генерацію розумного тексту. Розв'язок цих проблем буде означати створення зручнішої форми взаємодії комп'ютера та людини.

### **2.2 Розгляд обраних сервісів**

Було обрано 4 сервіси, які на сьогоднішній день мають практичне застосування у вирішенні бізнес задач, мають широкий спектр застосування та доступні для використання широкому загалу. Частина з них спрямована лише на вирішення питання оцінки тональності тексту, решта має ширший спектр застосування і можливості. Ключовий компонент NLP моделі заключається у концепції моделювання мови.

#### **2.2.1 Microsoft Text Analytics API**

Користувачі Microsoft Text Analytics API можуть витягувати ключові фрази, сутності (наприклад, людей, компанії чи місцеположення), настрої, а також визначати, на якій із 120 підтримуваних мов написаний їх текст. API аналізу настроїв повертає результати, використовуючи оцінку настрою від 0 (негативний) до 1 (позитивний). На сьогоднішній день програмне забезпечення може виявляти настрої в текстах англійською, іспанською, німецькою та французькою мовами. Розробники вказують, що аналіз слід робити на цілому документі, і рекомендують використовувати документи, що складаються з одного або двох речень, для досягнення більш високої точності.

Ось як Microsoft Text Analytics API аналізує огляд фільму The Nun. Він виявив англійську мову із 100-відсотковою впевненістю, і настрої вимірюються у відсотках. Результати аналізу також повертаються у форматі JSON (рис 2.1).

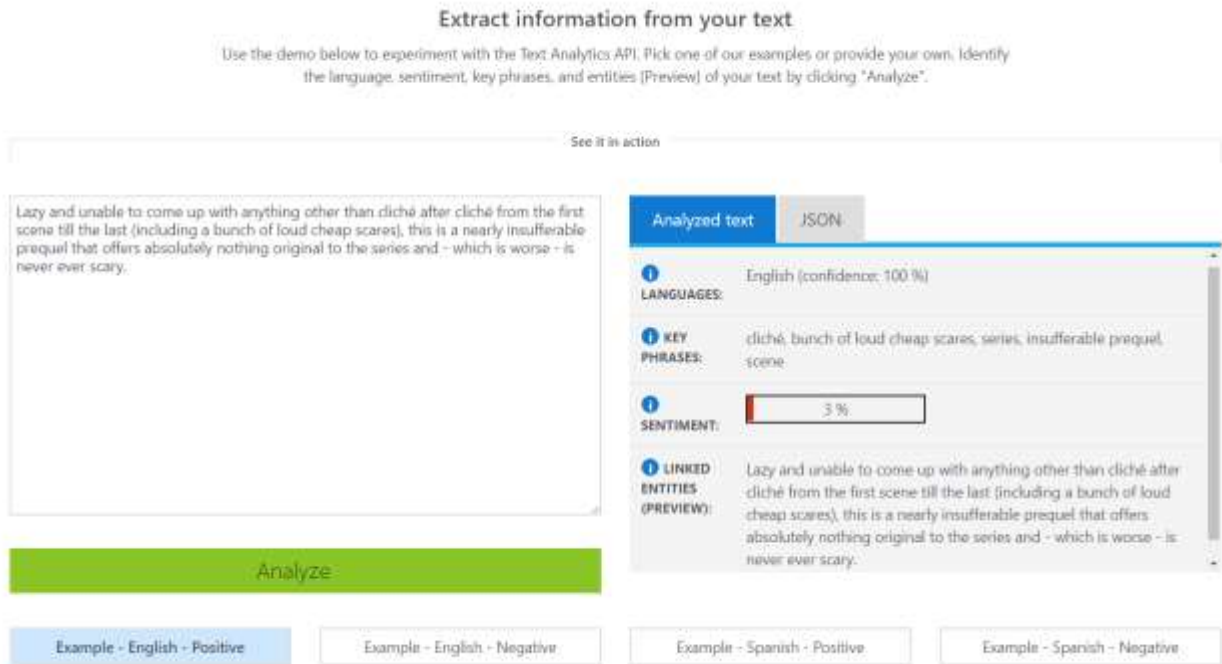


Рисунок 2.1 – Аналіз тексту від Microsoft Text Analytics API

Функція Text Analytics API аналізу настроїв забезпечує два способи виявлення позитивних і негативних настроїв. Якщо надіслати запит на аналіз настроїв, API поверне мітки настроїв (наприклад, "негативні", "нейтральні" та "позитивні") та оцінки достовірності на рівні пропозиції та рівня документа. Також можна надсилати запити на видобуток думки, використовуючи кінцеву точку Sentiment Analysis, яка надає детальну інформацію про думки, пов'язані з аспектами (такими як атрибути продуктів чи послуг) у тексті [13].

Аналіз настрою у версії 3.x застосовує мітки настроїв до тексту, які повертаються на рівні речення та документа, з оцінкою достовірності для кожного.

Мітки можуть бути позитивними, негативними та нейтральними. На рівні документа мітку змішаних настроїв також можна повернути. Сентимент документа визначається нижче (рис. 2.2):

Sentence sentiment	Returned document label
At least one <b>positive</b> sentence is in the document. The rest of the sentences are <b>neutral</b> .	<b>positive</b>
At least one <b>negative</b> sentence is in the document. The rest of the sentences are <b>neutral</b> .	<b>negative</b>
At least one <b>negative</b> sentence and at least one <b>positive</b> sentence are in the document.	<b>mixed</b>
All sentences in the document are <b>neutral</b> .	<b>neutral</b>

Рисунок 2.2 – Система оцінки тональності документа від Microsoft Text Analytics API

Оцінки впевненості коливаються від 1 до 0. Оцінки ближче до 1 вказують на вищу впевненість у класифікації етикетки, тоді як нижчі оцінки вказують на меншу впевненість. Для кожного документа або кожного речення прогнозовані бали, пов'язані з мітками (позитивні, негативні та нейтральні), сумарно дорівнюють 1.

Іншою особливістю сервісу Microsoft Text Analytics API є видобуток думок. Видобуток думок - це особливість аналізу настроїв, починаючи з версії 3.1-preview.1. Також відомий як Аспектний аналіз настрою в обробці природних мов (NLP), ця функція надає більш детальну інформацію про думки, пов'язані з аспектами (такими як атрибути продуктів чи послуг) у тексті.

Наприклад, якщо клієнт залишає відгуки про готель, такий як: "Кімната була чудова, але персонал був непривітний", "Майнінг думки" знайде аспекти в тексті, а також пов'язані з ними думки та настрої. Аналіз настрою може повідомляти лише про негативні настрої.

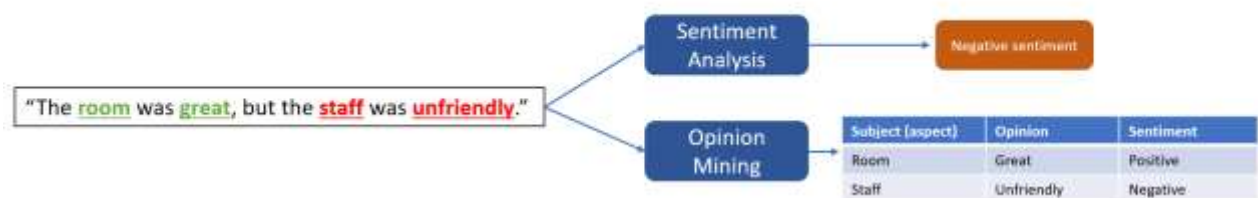


Рисунок 2.3 – Аспективний аналіз сервісу Microsoft Text Analytics API

Щоб отримати видобуток думок у результатах, потрібно включити прапорець `opinionMining = true` у запит на аналіз настроїв. Результати видобутку думок будуть включені до відповіді на аналіз настроїв.

Аналіз настрою дає більш якісний результат, за умови надання йому меншого обсягу тексту для роботи. Цей підхід протилежний виділенню ключових фраз, який краще працює на більших блоках тексту. Щоб отримати найкращі результати від обох операцій, варто розглянути можливість реструктуризації вхідних даних відповідно.

Для аналізу повинні бути документи JSON у такому форматі: `ID`, `text`, `language` (рис 2.4). Аналіз настрою підтримує широкий спектр мов, багато яких з можливістю попереднього перегляду. Розмір документа повинен бути менше 5120 символів на документ.

```

{
  "documents": [
    {
      "language": "en",
      "id": "1",
      "text": "The restaurant had great food and our waiter was friendly."
    }
  ]
}

```

Рисунок 2.4 – Формат запиту для сервісу Microsoft Text Analytics API

### 2.2.2 Google Cloud Natural Language API

Google Cloud Natural Language API витягує настрої з електронних листів, текстових документів, статей новин, соціальних мереж та публікацій у блогах. Його використання включає витяг статистичних даних із аудіофайлів, сканованих документів та документів іншими мовами у поєднанні з іншими хмарними службами.

Інструмент призначає оцінку настрою та величину для кожного речення, що дозволяє легко зрозуміти, що споживачеві найбільше сподобалось чи не сподобалось, а також відрізнити речення про настрої від речень, що не мають настрою.

Google надає свої послуги машинного навчання та AI на двох рівнях: Google Cloud Machine Learning для технічно підкованих фахівців у галузі даних та платформа Cloud AutoML, яка замінила застарілий API Prediction.

Google Cloud AutoML - це хмарна платформа ML, призначена для недосвідчених користувачів. Клієнти можуть завантажувати свої набори даних, навчати власні моделі та розміщувати їх на веб-сайті. AutoML повністю інтегрований з усіма сервісами Google і зберігає дані у хмарі. Навчені моделі можна розгорнути через інтерфейс REST API.

З AutoML доступно кілька продуктів, до яких ви можете отримати доступ через графічний інтерфейс. Якщо коротко їх описати, це послуги з обробки зображень та відео, механізм обробки та перекладу природних мов та навчальні моделі щодо структурованих даних. Оскільки доступ до кожного продукту можна отримати через API, ми розглянемо їх окремо в розділі API.

Розробники пропонують користувачам можливість спробувати послугу відразу і подивитися, що вона може зробити. Ось приклад огляду клієнтами навушників від Amazon (рис. 2.5).

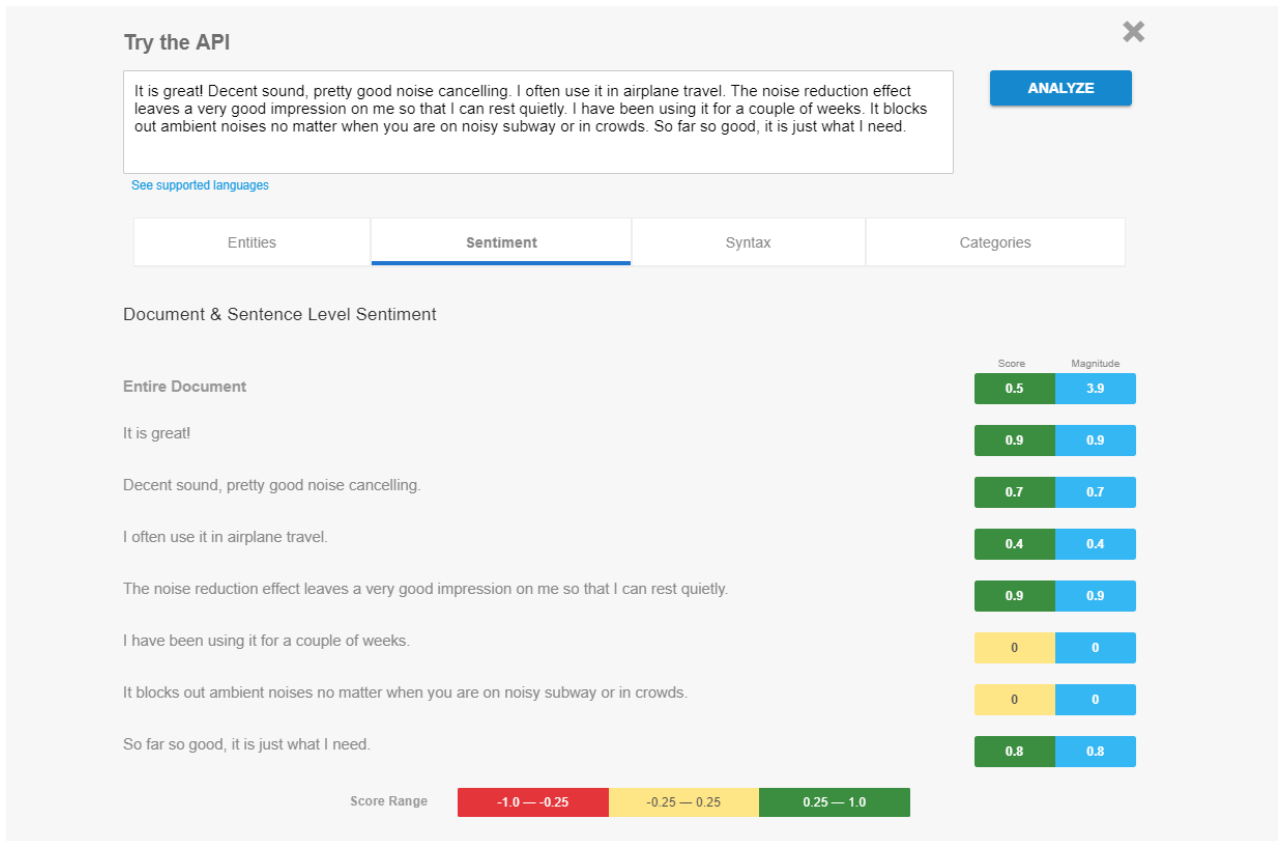


Рисунок 2.5 – Приклад оцінки від сервісу

Однією з найбільших проблем в обробці природних мов (NLP) є дефіцит навчальних даних. Оскільки NLP - це багатогалузеве поле з безліччю різних завдань, більшість наборів даних, що стосуються конкретних завдань, містять лише кілька тисяч або кілька сотень тисяч прикладів навчання, позначених людиною. Однак сучасні моделі НЛП на основі глибокого навчання бачать переваги від значно більших обсягів даних, покращуючись під час навчання на мільйонах або мільярдах анотованих прикладів навчання. Щоб допомогти усунути цей розрив у даних, дослідники розробили різноманітні методики для навчання моделей представлення мови загального призначення, використовуючи величезну кількість ненаголошеного тексту в Інтернеті (відомий як попередня підготовка). Заздалегідь підготовлена модель може бути точно налаштована на завдання NLP з невеликими даними, такі як відповіді на запитання та аналіз настроїв, що призводить до значних поліпшень точності порівняно з навчанням на цих наборах даних з нуля.

Потужні попередньо навчені моделі API природничої мови дозволяють розробникам легко застосовувати розуміння природної мови (NLU) до своїх додатків із такими функціями, як аналіз настроїв, аналіз сутності, аналіз настроїв сутності, класифікація вмісту та аналіз синтаксису.

### 2.2.3 Sentiment Analysis від Tweenword

Sentiment Analysis від Tweenword API аналізу настроїв - слугує, щоб з'ясувати тон пропозиції чи абзацу. Цей API може бути корисним, якщо існує велика кількість відповідей або відгуків користувачів і необхідно швидко знайти негативні коментарі, щоб побачити, що не подобається клієнтам, і навпаки [14].

Окрім того, що сервіс надає відповідь типу («негативна», «нейтральна» чи «позитивна»), існує можливість самостійно визначити, що саме ви вважаєте позитивним чи негативним.

Оцінка вказує, наскільки негативним чи позитивним є загальний аналізований текст. Все, що нижче показника  $-0,05$ , сервіс позначає як негатив, а все, що перевищує  $0,05$  - як позитив. Все, що між ними включно, позначається як нейтральне [15].

Таким чином, маючи у результаті фактичні бали, які сервіс використовує для визначення типу, користувач можете відхилити інтерпретацію, яку надає сервіс, та інтерпретувати фактичні бали самостійно, створити власні негативні та позитивні мінімальні бали.

Співвідношення (Ratio) - це загальний бал негативних слів порівняно із загальним сумарним балом позитивних слів, який коливається від  $-1$  до  $1$ . Наведемо для пояснення приклад.

Припустимо, що необхідно проаналізувати слова A, B, C та D.

Сценарій 1) Якщо слова A і B є негативними словами, що мають загальний загальний бал  $-5,0$ , а слова C і D - позитивні слова, що мають загальний загальний бал  $5,0$ , тоді співвідношення буде  $0$  (або  $1: 1$ ) так як обидві оцінки збалансовані.

Сценарій 2) Якщо загальні загальні оцінки негативних слів складають  $-5,0$ , тоді як позитивні слова мають загальну загальну оцінку  $2,5$ , тоді співвідношення буде  $-0,667$  (або  $2: 1$ ).

Сценарій 3) Якщо загальні негативні слова складають  $-5,0$ , а позитивні -  $0,0$ , тоді співвідношення буде  $-1$  (або  $1: 0$ ).

Зрештою, оцінка та співвідношення дозволяють вийти за межі того, що вважається негативним або того, що вважається позитивним. У своєму коді користувач може врахувати вплив оцінки або вплив коефіцієнта, або навіть поєднати їх усі, щоб встановити власні позитивні та негативні пороги відповідно задач.

Наприклад, замість того, щоб використовувати нашу інтерпретацію, що менше  $-0,05$  є негативним, а більше  $0,05$  - позитивним, є можливість написати свій власний алгоритм, який говорить, що нижче  $-0,9$  є негативним, а вище  $-0,8$  - позитивним.

## 2.2.4 Sentiment Analysis від Text Processing

Щоб правильно зрозуміти статтю, іноді потрібно посилатися на слово або речення, яке зустрічається на кілька тисяч слів назад. Це приклад залежності великої дальності - поширене явище, яке зустрічається в послідовних даних, - яке слід розуміти для вирішення багатьох реальних завдань. Хоча люди роблять це природно, моделювання тривалої залежності від нейронних мереж залишається проблемою. РНН на основі решітки та градієнтна техніка відсікання покращують здатність моделювати довгострокову залежність, але все ще недостатньо для повного вирішення цього питання.

Одним із способів вирішити цю проблему є використання Трансформаторів, що дозволяє здійснювати прямі зв'язки між одиницями даних, пропонуючи обіцянку кращого захоплення довгострокової залежності. Однак у мовному моделюванні Трансформатори зараз реалізуються з контекстом фіксованої довжини, тобто довга послідовність тексту врізається в сегменти фіксованої довжини у кілька сотень символів, і кожен сегмент обробляється окремо.

### Повторність на рівні сегмента

Під час навчання представлення, обчислені для попереднього сегмента, фіксуються та кешуються для повторного використання у розширеному контексті, коли модель обробляє наступний новий сегмент. Це додаткове з'єднання збільшує найбільшу можливу довжину залежності в  $N$  разів, де  $N$  - глибина мережі, оскільки контекстна інформація тепер може протікати через межі сегмента. Більше того, цей механізм рецидивування також вирішує проблему фрагментації контексту, надаючи необхідний контекст для лексем перед новим сегментом.

### Відносні позиційні кодування

Наївне застосування повторення на рівні сегменту не працює, тому що позиційні кодування не є когерентними, коли ми використовуємо попередні сегменти. Наприклад, розглянемо старий сегмент з контекстними позиціями [0, 1, 2, 3]. Коли новий сегмент обробляється, ми маємо позиції [0, 1, 2, 3, 0, 1, 2, 3] для двох комбінованих сегментів, де семантика кожного ідентифікатора позиції не є когерентною через поза послідовність. З цією метою ми пропонуємо нову схему відносної позиції кодування, щоб зробити можливим механізм рецидивування. Більше того, на відміну від інших відносних схем позиційного кодування, наша формулювання використовує фіксовані вбудовування з навчальними перетвореннями замість навчальних вкладень і, таким чином, є більш узагальненою для довших послідовностей у тестовий час.

## 2.3 Вибір вхідних факторів і відгуків для визначення ефективності роботи сервісів оцінки тональності текстової інформації

### 2.3.1 Розгляд даних з якими використовуються досліджувані алгоритми

Серед багатоцільових методів для обробки природньої мови існує декілька видів визначення ефективності, залежно від цільової галузі застосування. Якщо вести мову про застосування даних методів у цілях ранжування посилань у мережі Інтернет та підбору найбільш релевантних результатів, що відповідають запиту за змістом, то тут також спостерігається різноманіття методів для аналізу.

Враховуючи той факт, що розробки у напрямі застосування систем штучного інтелекту у пошукових сервісах почалися приблизно з 2013 року, коли були проведені перші дослідження можливостей системи Word2Vec для семантичного аналізу [16], та швидкий розвиток технологій цього сегмента, у якому використовуються різні підходи до аналізу інформації, та продовжується пошук найліпшої архітектури для моделей, не існує єдиної загальної метрики оцінки ефективності технологій. Так, зокрема, відомі лабораторії з розробки інтелектуальних систем застосовують власні методології для оцінки ефективності алгоритмів.

Однак можна виокремити наступні напрями, за якими в останні роки проводилися найбільші дослідження у даній галузі, та у яких відбувалися порівняння різних методів:

**Part-of-speech (POS) tagging:** У корпусній лінгвістиці, розмічування частин мови (англ. part-of-speech tagging, POS tagging, POST), також називається граматичним позначенням або розбором на частини мови, це процес позначення слова в тексті (корпусі) належним до певної частини мови, заснований як на його визначенні, так і на його контексті — тобто, на його зв'язку з суміжними і спорідненими словами у фразі, реченні, або абзаці.[17] Спрощена форма корпусної лінгвістики зазвичай викладається дітям шкільного віку, у вигляді визначення слів як іменник, дієслово, прикметник, прислівник, тощо (рис. 2.6).



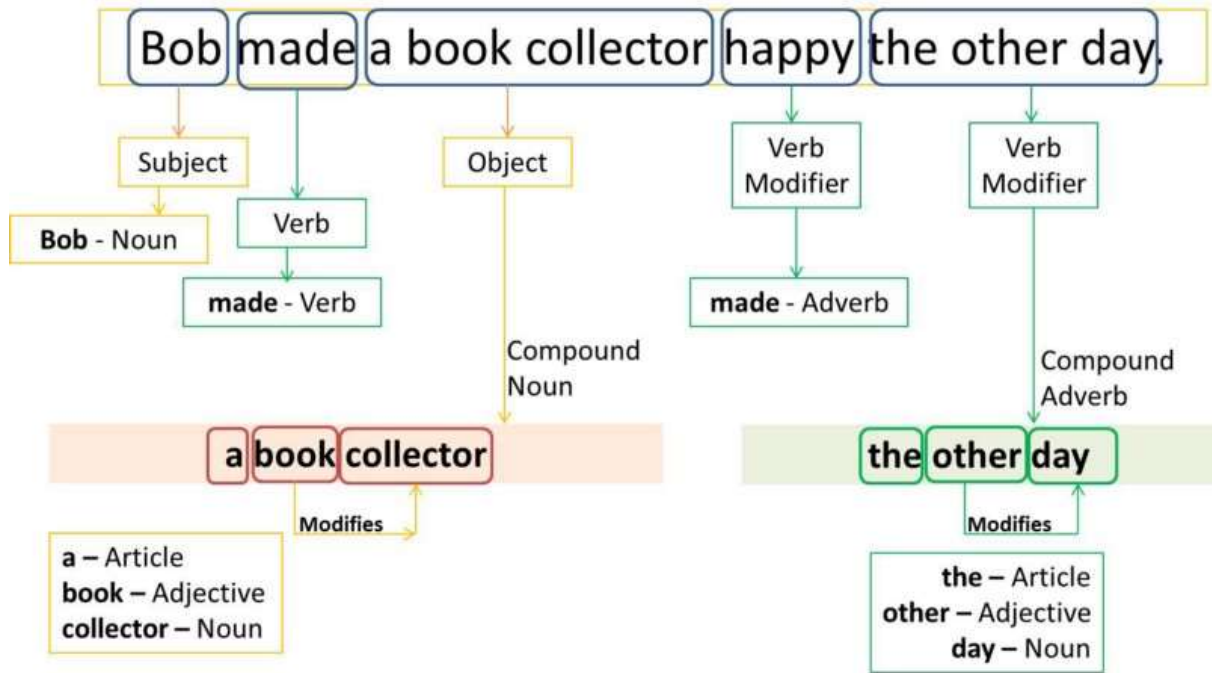


Рисунок 2.6 – Схематичне зображення розмічування частин мови

Зумівши виконати визначення частин мови вручну, це почали робити в контексті обчислювальної лінгвістики, використовуючи алгоритми, які асоціюють дискретні терміни, а також приховані частини мови, відповідно до набору описових позначень. Алгоритми розбору на частини мови поділяються на дві відмінні групи: на основі правил і на стохастичні. Розбірник Е. Брілла, один з перших і найбільш широко використовуваних англійських розбірників, використовує алгоритми на основі правил.

Хоча існує загальна згода щодо основних категорій, ряд крайніх випадків ускладнює вирішення питання про єдиний «правильний» набір позначень, навіть у певній мові, наприклад, англійській. Наприклад, важко сказати, чи є «fire» (укр. вогонь) прикметником чи іменником

the big green fire truck

Другий важливий зразок використання–згадування відмінності, як у наступному прикладі, де «blue» (укр. синій) міг бути замінений словом будь-якою частини мови (збірник позначень Brown Corpus додає суфікс «-NC» у таких випадках):

the word "blue" has 4 letters.

Слова у мові, які відрізняються від тих, що в «основному» тексті, зазвичай позначаються як «іноземні», як правило, на додаток до позначення ролі, яку іноземне слово фактично відтворює в контексті.

В англійській є також багато випадків, коли частини мови та «слова» не мають однозначної відповідності, наприклад:

- as far as;
- david's;
- gonna;

- don't;
- vice versa;
- first-cut;
- cannot;
- pre- and post-secondary;
- look (a word) up.

В останньому прикладі, «look» і «up», можливо, функціують як єдине вербальна одиниця, незважаючи на можливість виникнення між ними інших слів. Деякі набори позначень (наприклад, Penn) розбивають переносні слова, скорочення і присвійні на окремі позначення, таким чином уникаючи деяких, але далеко не усіх таких проблем.

Багато наборів трактують слова «be», «have», та «do» як самостійні (як у Brown Corpus), тоді як деякі вважають їх усі просто дієсловами (наприклад, LOB Corpus та Penn Treebank). Оскільки ці конкретні слова мають більше форм, ніж інші англійські дієслова, і зустрічаються в зовсім різних граматичних контекстах, трактування їх просто як «дієслова» означає, що розмітник буде мати набагато менше інформації. Наприклад, позначення, засноване на ПММ, буде поєднувати кілька рядків і стовпців, які в іншому випадку будуть не тільки окремими, але і зовсім несумісними. Більш складний алгоритм може також враховувати конкретне слово в кожному випадку; але з різними позначеннями, сама ПММ часто може передбачити правильне найточніше позначення навіть для нових варіантів написання, і таким чином забезпечити кращу допомогу для подальшої обробки.

Найбільш популярний «набір позначень» для розбору на частини мови для американської англійської — ймовірно, набір позначень Penn, розроблений в проекті Penn Treebank. Він значною мірою схоже на попередні набори Brown Corpus і LOB Corpus, хоча і набагато менший. У Європі набори з Eagles Guidelines) широко використовуються і включають версії для декількох мов.

Робота з розбору виконується для різних мов, а набір позначень залежить від самої мови. Позначення, як правило, розроблені таким чином, щоб включати явні морфологічні відмінності, хоча це призводить до таких невідповідностей, як наявність позначень для займенників, але відсутність для іменників в англійській, і набагато більшим міжмовним відмінностям. Набір позначень для мов з великим змістом, таких як Грецька та Латинь може бути дуже великим; віртуальний розбір слів у аглютинативні мовах таких як інуїтська може бути зовсім неможливим. З іншого боку, Петров запропонував «універсальний» набір позначень, з 12 категоріями (наприклад, ніяких підтипів іменників, дієслів, пунктуації і т. д.; немає відмінності від «to» як інфінітивного позначення проти прийменника (навіть чи «універсального» збігу), тощо). Який набір вибрати, чи дуже малий набір дуже широких міток, чи набагато більший набір більш точних, залежить від конкретної мети. Автоматичне позначання легше на менших наборах тегів.

**Parsing:** Існує два типи розбору: залежносинтаксичний аналіз (dependency parsing), який з'єднує окремі слова з їх відносинами, і навколосинтаксичний

аналіз (constituency parsing), який ітеративно розбиває текст на підфрази (рис. 2.7) [18].

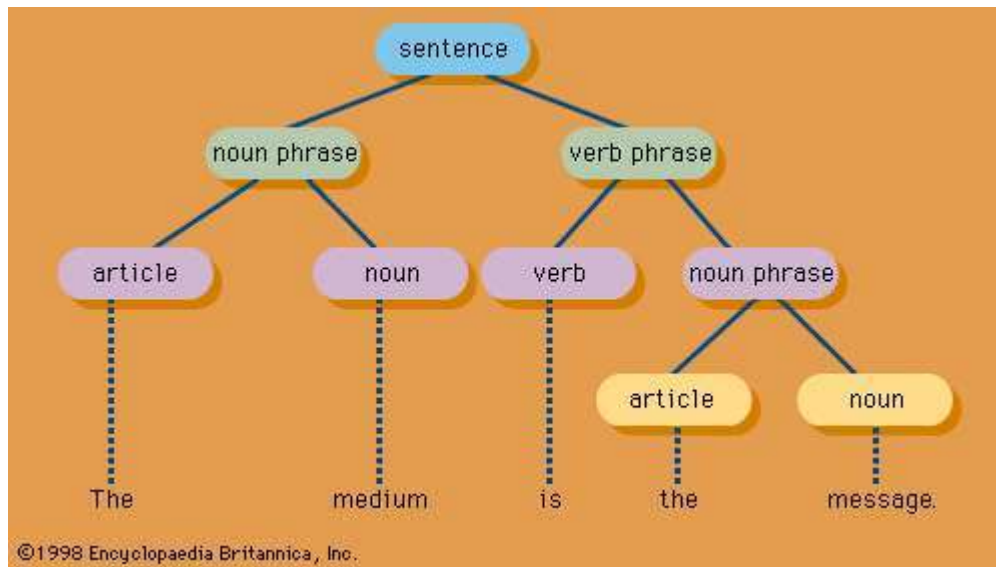


Рисунок 2.7 – Схематичне зображення розбору тексту на підфрази

Синтаксичний розбір або розбір залежності - це завдання розпізнавання речення та присвоєння йому синтаксичної структури. Найбільш широко використовуваною синтаксичною структурою є дерево розбору, яке можна створити за допомогою деяких алгоритмів розбору [19]. Ці дерева розбору корисні в різних програмах, таких як перевірка граматики або, що ще важливіше, вони відіграють критичну роль на етапі семантичного аналізу. Наприклад, щоб відповісти на запитання "Хто є захисником точки LA Laker у наступній грі?" нам потрібно з'ясувати його предмет, предмети, атрибути, щоб допомогти нам зрозуміти, що користувач хоче охоронця точки LA Lakers спеціально для наступної гри.

Тепер завдання синтаксичного розбору є досить складним через те, що дане речення може мати кілька дерев розбору, які ми називаємо двозначностями. Розгляньте речення «Забронюйте цей рейс». які можуть утворювати кілька дерев розбору на основі його неоднозначної частини мовних тегів, якщо ці двозначності не будуть вирішені. Вибір правильного розбору з декількох можливих синтаксисів називається синтаксичним розбором. Алгоритми розбору, такі як алгоритм Кокка-Касамі-Молодшого (СКУ), алгоритму Ерлі або алгоритми розбору діаграм, використовують динамічний підхід програмування для вирішення проблем неоднозначності.

Навколосинтаксичний аналіз має на меті отримати дерево розбору на основі оточення з речення, яке представляє його синтаксичну структуру відповідно до граматики фразової структури. Останні підходи перетворюють дерево розбору в послідовність після обходу першої глибини, щоб можна було застосувати до нього моделі послідовність-до-послідовність (sequence-to-sequence).

**Named entity recognition:** Розпізнавання іменованої сутності (NER) - іноді її називають збиванням, вилученням або ідентифікацією - це завдання ідентифікації та категоризації ключової інформації (сутностей) у тексті [20]. Суб'єктом може бути будь-яке слово або ряд слів, які послідовно посилаються на одне і те ж. Кожна виявлена сутність класифікується на заздалегідь визначену категорію. Наприклад, модель машинного навчання NER (ML) може виявити в тексті слово "Canotic" і класифікувати його як "Компанія" (рис. 2.8).

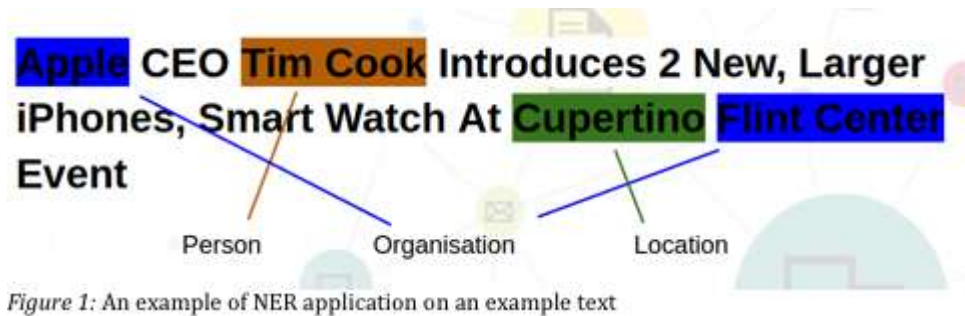


Figure 1: An example of NER application on an example text

Рисунок 2.8 – Розпізнавання сутностей у реченні

В основі будь-якої моделі NER лежить двоступеневий процес:

- виявити названу сутність;
- класифікуйте сутність.

Під цим криється пара речей.

Перший крок передбачає виявлення слова або рядка слів, які утворюють ціле. Кожне слово являє собою маркер: "Великі озера" - це рядок з трьох лексем, що представляє одну сутність. Позначення всередині-поза початком - це звичайний спосіб вказувати, де починаються та закінчуються сутності.

NER підходить для будь-якої ситуації, коли огляд великої кількості тексту корисний на високому рівні. За допомогою програми NER можна, з першого погляду, зрозуміти предмет чи тему тексту та швидко групувати тексти, виходячи з їх релевантності чи подібності.

Деякі відомі випадки використання NER включають:

**Пошукові та рекомендаційні системи:** Підвищити швидкість та актуальність результатів пошуку та рекомендацій шляхом узагальнення описового тексту, оглядів та дискусій

**Класифікація змісту:** Відображає актуальні тенденції визначивши предмет та теми публікацій блогу та новин.

**Semantic role labeling:** При обробці природною мовою семантичне позначення ролей (його також називають дрібним семантичним розбором або заповненням слотів) - це процес, який присвоює мітки словам або фразам у реченні, які вказують на їх семантичну роль у реченні, наприклад, на роль агента, мети, або результат [21].

Він складається з виявлення смислових аргументів, пов'язаних із присудком або дієсловом речення, та їх класифікації на їх конкретні ролі.

Наприклад, враховуючи таке речення, як "Мері продала книгу Джону", завданням було б визнати дієслово "продати" таким, що представляє присудок, "Мері" як представника продавця (агента), "книгу" як представника «Товари» (тема) та "Джон" як реципієнт. Це важливий крок до осмислення сенсу речення. Семантичний аналіз подібного роду знаходиться на нижчому рівні абстрагування, ніж синтаксичне дерево, тобто воно має більше категорій, тому групується менше пропозицій у кожній категорії. Наприклад, "книзі належить мені" знадобляться два ярлики, такі як "одержимий" і "володар", тоді як "книга продана Джону" потребували б ще двох ярликів, таких як "мета" (або "тема") та "приймач" (або "одержувач"), хоча ці два пункти були б дуже схожі на функції "суб'єкт" та "об'єкт" (рис 2.9).

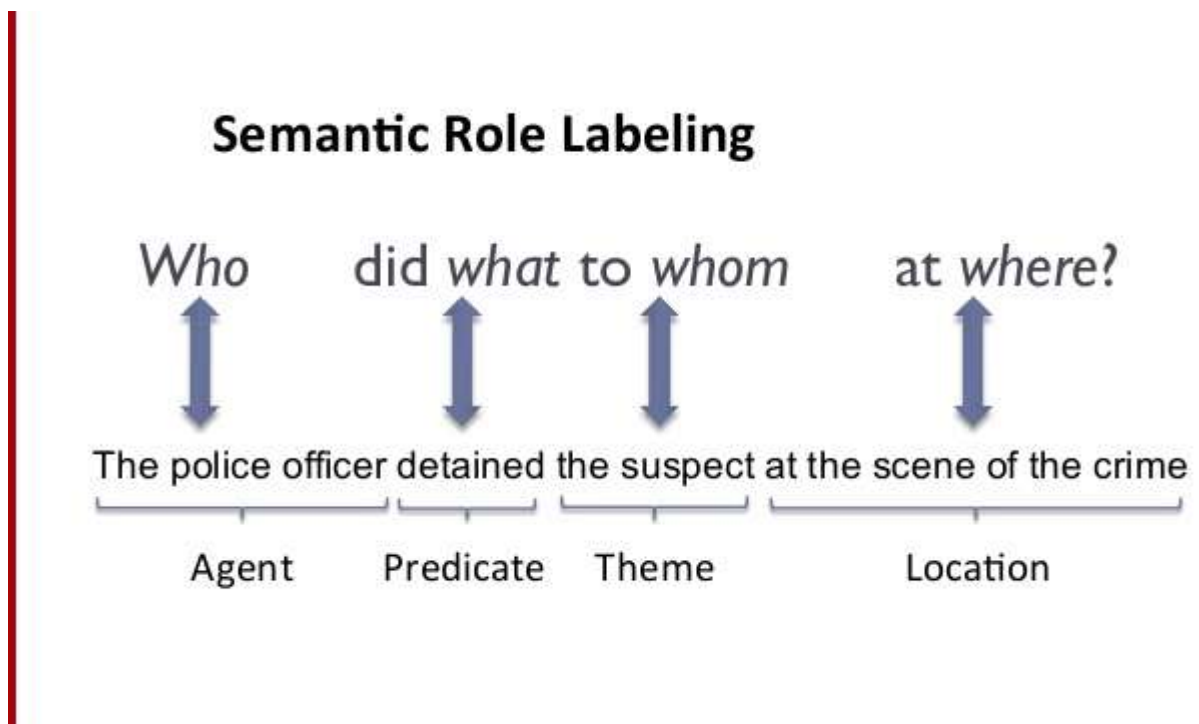


Рисунок 2.9 – Принцип семантичного розподілення ролей

Проект FrameNet створив перший великий обчислювальний лексикон, який систематично описував багато предикатів та відповідні їм ролі. Даніель Гілдеа (Каліфорнійський університет, Берклі / Міжнародний інститут комп'ютерних наук) та Даніель Юрафські (нині викладають в Стенфордському університеті, але раніше працювали в Університеті Колорадо та УК Берклі) розробили першу автоматичну семантичну систему маркування ролей на базі FrameNet. Корпус PropBank додав вручну створені анотації семантичної ролі до текстів Penn Treebank з текстів Wall Street Journal. Багато систем автоматичного семантичного маркування ролей використовували PropBank як навчальний набір даних, щоб навчитися автоматично коментувати нові речення.

**Sentiment Classification:** Аналіз почуттів - це метод аналізу тексту, який виявляє полярність (наприклад, позитивну чи негативну думку) всередині тексту, будь то цілий документ, абзац, пропозиція чи пункт [22].

Розуміння емоцій людей є важливим для бізнесу, оскільки клієнти можуть висловити свої думки та почуття відкритіше, ніж будь-коли раніше. Автоматично аналізуючи відгуки клієнтів, починаючи з відповідей опитувань на бесіди в соціальних мережах, бренди можуть уважно слухати своїх клієнтів та пристосовувати продукти та послуги відповідно до їх потреб.

Наприклад, використання аналізу настроїв для автоматичного аналізу 4000+ відгуків про ваш продукт може допомогти вам дізнатися, чи задоволені клієнти щодо цінних планів та обслуговування клієнтів.

Моделі аналізу почуттів зосереджуються на полярності (рис. 2.10) (позитивна, негативна, нейтральна), а також на почуттях та емоціях (сердитий, щасливий, сумний тощо) і навіть на намірах (наприклад, зацікавлених проти не зацікавлених).

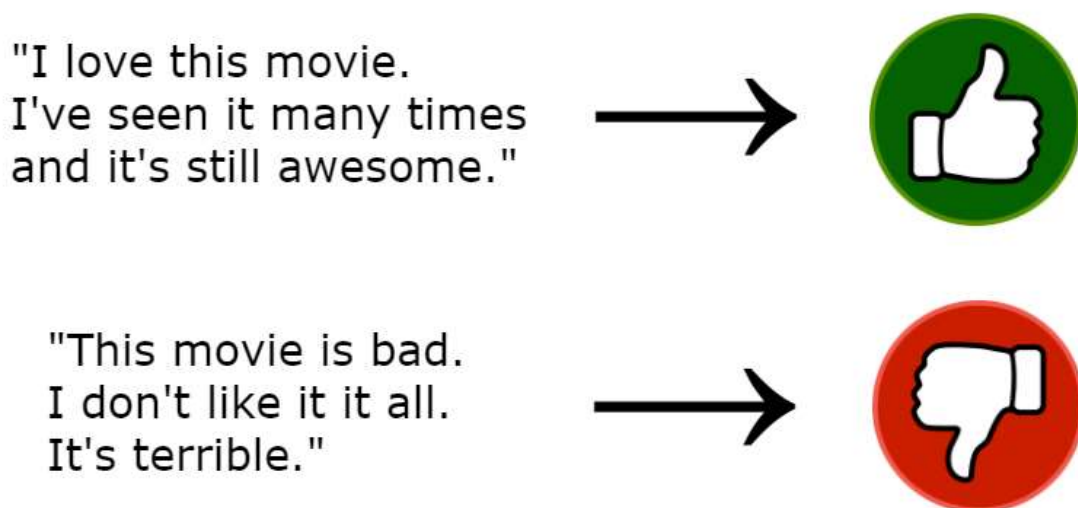


Рисунок 2.10 – Визначення почуття з тексту

За підрахунками, 80% світових даних неструктуровані, тобто неорганізовані. Величезні обсяги текстових даних (електронні листи, квитки на підтримку, чати, бесіди в соціальних мережах, опитування, статті, документи тощо) створюються щодня, але важко проаналізувати, зрозуміти та сортувати, не кажучи вже про трудомісткі та дорогі .

Аналіз почуттів, однак, допомагає компаніям осмислити весь цей неструктурований текст, автоматично позначивши його.

Переваги аналізу настроїв включають:

**Сортування даних у масштабі:** Чи можете ви уявити, вручну сортуючи тисячі твітів, розмови щодо підтримки клієнтів чи опитування? Існує занадто багато даних для обробки вручну. Аналіз почуттів допомагає підприємствам обробляти величезну кількість даних ефективно та економічно.

**Аналіз в реальному часі:** Аналіз почуттів може визначити критичні проблеми в режимі реального часу, наприклад, посилюється PR-криза в соціальних мережах? Чи розлючений клієнт збирається збивати? Моделі аналізу почуттів можуть допомогти вам негайно визначити подібні ситуації та оцінити настрої брэнда, тому ви можете вжити заходів відразу.

**Послідовні критерії:** Підраховано, що люди погоджуються лише приблизно 60-65% часу, коли визначають настрої певного тексту. Позначення тексту за настроями є високо суб'єктивним, на нього впливають особисті переживання, думки та переконання. Використовуючи централізовану систему аналізу настроїв, компанії можуть застосовувати однакові критерії до всіх своїх даних, допомагаючи їм підвищити точність та отримати кращу інформацію.

**Machine Translation:** Машинний переклад (МТ) - це завдання перекласти текст з мови джерела на його аналог цільовою мовою. Існує багато складних аспектів МТ: 1) велика різноманітність мов, алфавітів та граматики; 2) завдання перекласти послідовність (наприклад, речення) у послідовність складніше для комп'ютера, ніж робота лише з числами; 3) немає жодної правильної відповіді (наприклад: перекладаючи з мови без залежних від статі займенників, він і вона можуть бути однаковими) [23].

Машинний переклад - відносно давнє завдання. З 1970-х років існували проекти щодо автоматичного перекладу. За роки виникли три основні підходи:

Машинний переклад на основі правил (RBMT): 1970-1990-ті

Статистичний машинний переклад (SMT): 1990-ті-2010 роки

Нейронний машинний переклад (NMT): 2014-наші дні

Нейронний підхід використовує нейронні мережі для досягнення машинного перекладу. Порівняно з попередніми моделями, НМТ можуть бути побудовані за допомогою однієї мережі замість трубопроводу окремих завдань.

У 2014 році були введені моделі послідовності за послідовністю, що відкриває нові можливості нейронних мереж в НЛП. Перед моделями seq2seq нейронним мережам потрібен спосіб перетворення послідовності введення в готові до комп'ютера числа (однокольорове кодування, вбудовування). За допомогою seq2seq стала можливою можливість навчання мережі з послідовностями введення та виведення.

НМТ з'явився швидко. Після кількох років досліджень ці моделі перевершили ЗПТ. З покращеними результатами багато компаній-постачальників перекладачів змінили свої мережі на нейронні моделі, включаючи Google та Microsoft.

Проблема з нейронними мережами виникає, якщо дані тренувань є незбалансованими, модель не може вчитись з рідкісних зразків, а також з частих. Що стосується мов, це поширена проблема, оскільки є багато рідкісних слів, які використовуються лише кілька разів у всій Вікіпедії, наприклад. Тренувати модель, яка не схильна до частих слів (наприклад, багаторазові зустрічі на всіх сторінках Вікіпедії), може бути складним завданням. Нещодавній документ пропонує рішення, використовуючи крок після обробки для перекладу цих рідкісних слів зі словником.

Нещодавно дослідники Facebook запровадили невідконтрольну МТ модель, яка працює як з SMT, так і з NMT, що вимагає лише великих одномовних корпорацій, а не двомовних. Основним вузьким місцем попередніх прикладів була відсутність великої бази даних з перекладами для навчання. Ця модель показує обіцянку вирішити цю проблему.

Приклади НМТ

- посилання Google Translate (з 2016 року) до мовної команди в Google AI;
- microsoft Translate (з 2016 року) посилання на дослідження МТ в Microsoft;
- переклад у Facebook: посилання на NLP на AI Facebook;
- openNMT: система перекладу нейронних машин з відкритим кодом;
- переваги;
- моделі "до кінця" (відсутність конкретних завдань);
- недоліки;
- потрібен двомовний корпус;
- проблема рідкісних слів.

**Question answering:** Є три основні сучасні парадигми відповіді на питання [24]:

а) Фактоїдне питання на основі ІР-відповіді - відповіді на запитання користувача шляхом пошуку коротких текстових сегментів в Інтернеті чи іншій колекції документів. На етапі опрацювання запитань витягується ряд відомостей із запитання. Тип відповіді визначає тип сутності, з якої складається відповідь (особа, місцезнаходження, час тощо). Запит визначає ключові слова, які слід використовувати для ІР-системи для пошуку документів.

б) Відповідь на запитання на основі знань - це ідея відповіді на питання природної мови шляхом відображення її на запит по структурованій базі даних. Логічна форма питання, таким чином, або у формі запиту, або може бути легко перетворена в одну. База даних може бути повною реляційною базою даних або більш простими структурованими базами даних, як набори трифайців RDF. Системи для відображення текстового рядка в будь-яку логічну форму називаються семантичними аналізаторами. Семантичні аналізатори для відповіді на запитання зазвичай відображають або якусь версію обчислення предиката або мову запиту, як SQL або SPARQL.

с) Використання декількох джерел інформації: система Watson IBM від IBM, яка перемогла під загрозою! виклик у 2011 році - це приклад системи, яка спирається на широкий спектр ресурсів для відповіді на запитання. Перший етап - обробка питань. Система DeepQA виконує розбір, позначає теги сутності та вилучення відношень у питанні. Потім, як і текстові системи, система DeepQA витягує фокус, тип відповіді (також називається лексичним типом відповідей або LAT) і виконує класифікацію запитань та секціонування питань. Далі DeepQA виділяє фокус питання. Нарешті, питання класифікується за типом як питання щодо визначення, множинного вибору, головоломки або заповнення порожнього. Далі йде етап генерації відповідей кандидата за типом



запитання, де оброблене запитання поєднується із зовнішніми документами та іншими джерелами знань, щоб запропонувати багато відповідей кандидата. Ці відповіді кандидата можуть бути витягнуті з текстових документів або зі структурованих баз знань. Потім він проходить через етап оцінки відповіді кандидата, який використовує багато джерел доказів для оцінки кандидатів. Одним з найбільш важливих є тип лексичної відповіді. На етапі злиття та оцінювання остаточної відповіді спочатку об'єднуються відповіді кандидата, які є рівнозначними. Злиття та ранжування насправді виконується ітераційно; спочатку кандидати класифікуються класифікатором, даючи приблизне перше значення для кожної відповіді кандидата, потім це значення використовується для вирішення, який із варіантів імені вибрати як об'єднану відповідь, потім об'єднані відповіді перераховуються.

**Проблеми у відповіді на питання**

Основні проблеми, які постає перед системою відповідей на запитання, описані нижче:

**Лексична прогалина:** У природній мові те саме значення може виражатися по-різному. Оскільки на запитання, як правило, можна відповісти лише у випадку виявлення кожної згаданих концепцій, усунення цього розриву значно збільшує частку питань, на які може відповісти система.

**Неоднозначність:** це явище однієї і тієї ж фрази, що має різні значення; це може бути структурно-синтаксичним (на кшталт «літаючих літаків») або лексичним та семантичним (як «банк»). Один і той же рядок випадково відноситься до різних понять (як у грошовому банку проти річкового берега) та полісемії, де одна і та ж нитка посилається на різні, але пов'язані з цим поняття (як у банку як компанія проти банку як будівлі).

**Багатомовність:** знання в Інтернеті виражаються різними мовами. Хоча ресурси RDF можна описати на кількох мовах одночасно за допомогою мовних тегів, у веб-документах завжди не використовується жодної мови. Крім того, користувачі мають різні рідні мови. Очікується, що система забезпечення якості розпізнає мову та отримає результати на ходу.

**Dialogue Systems:** Діалогічна система, або розмовний агент (CA), - це комп'ютерна система, призначена для спілкування з людиною (рис. 2.11). Діалогові системи використовували один або кілька текстових, мовленнєвих, графічних, хаптичних, жестів та інших режимів для зв'язку як на вхідному, так і на вихідному каналах [25].

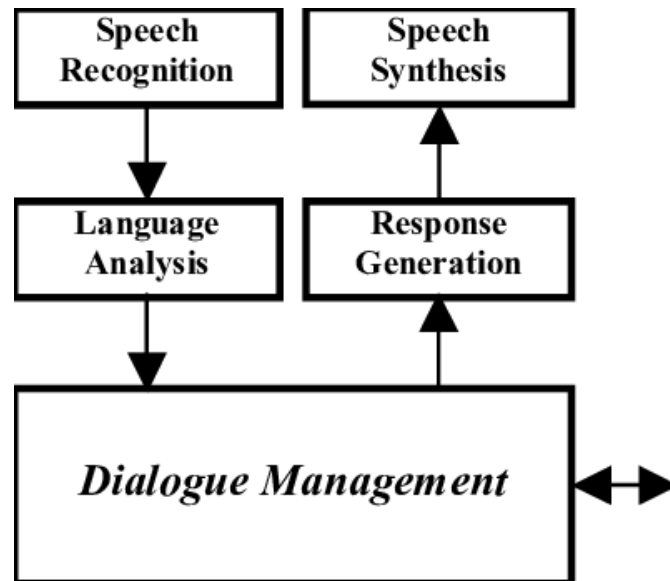


Рисунок 2.11 – Базова схема діалогічної системи

Елементи системи діалогу не визначені, однак вони відрізняються від чатботу. Типовий майстер графічного інтерфейсу вступає у такий собі діалог, але він включає дуже мало загальних компонентів системи діалогу, і стан діалогу є тривіальним.

Після діалогових систем, заснованих лише на письмовій обробці тексту, починаючи з початку шістдесятих років, перша діалогічна система діалогу була випущена Проектом DARPA у США в 1977 році. Після закінчення цього 5-річного проекту деякі європейські проекти видали першу систему діалогу, здатну розмовляти багатьма мовами (також французькою, німецькою та італійською). Ці перші системи використовувались у телекомунікаційній галузі для надання телефонних послуг різних послуг у конкретних областях, наприклад автоматизоване обслуговування порядку денного та поїздів.

Які комплекти компонентів включені в систему діалогу та як вони розділяють обов'язки, відрізняється від системи до системи. Основним для будь-якої діалогічної системи є диспетчер діалогів, який є компонентом, що керує станом діалогу та стратегією діалогу.

**Contextual Embeddings:** Представлення тексту (він же вкладення тексту) - це прорив у вирішенні завдань NLP. На початку вектор одного слова являє собою слово, незважаючи на те, що він має різний зміст серед контексту. Наприклад, "Вашингтон" може бути місцезнаходженням, назвою або штатом. «Університет Вашингтона» [26].

Загальна конструкція полягає в тому, що передача речення Мові символів символів для отримання контекстних вкладень таким чином, що модель маркування послідовності може класифікувати сутність (рис 2.12).

Загальна конструкція полягає в тому, що передача речення Мові символів символів для отримання контекстних вкладень таким чином, що модель маркування послідовності може класифікувати сутність.

Відмінні від класичних вкладень слова, Акбік та ін. оголошує це як контекстуальне вкладення слів. Іншим словом, вбудовані слова охоплюють семантику слова в контексті, так що воно може по-різному представлятися під різними, хоча це одне і те ж слово. Для більш детальної інформації ви можете звернутися до контекстуальних векторів слів (CoVe) та вкладень з мовних моделей (ELMo). Albik та ін. назвали їх вбудовування вкладками контекстуальних рядків.

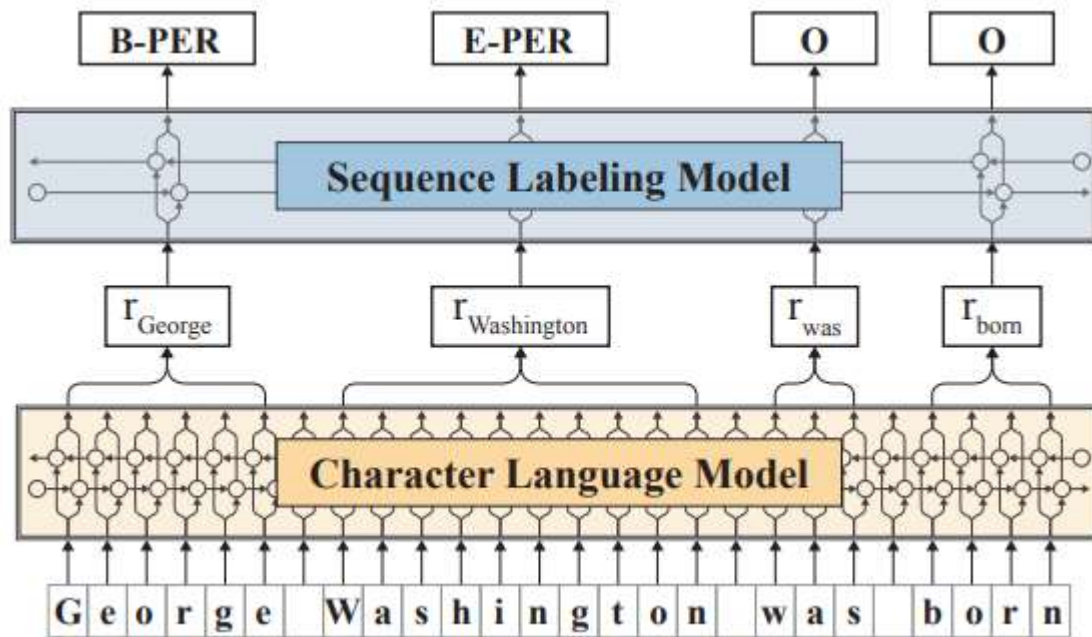


Рисунок 2.12 – Архітектура і дизайн системи

На відміну від інших моделей, вона базується на токенизації рівня символів, а не на токенизації на рівні слів. Іншим словом, він перетворить речення в послідовність символів і пройде через мовну модель, щоб вивчити подання слова.

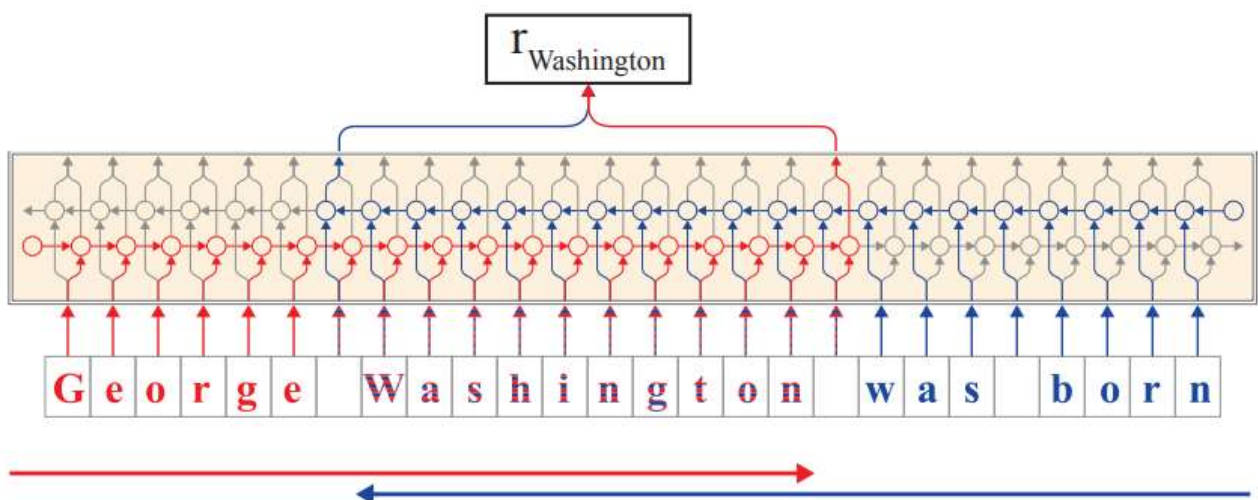


Рисунок 2.13 – Контекстуальні вставки «Вашингтона»

Взявши за приклад "Вашингтон" (рис 2.13), двонаправлена модель LSTM дозволяє "Вашингтону" отримувати інформацію з попереднього слова (тобто Джорджа) та наступних слів (тобто народився) таким чином, щоб він міг обчислити вектори в сентенційному контексті.

Вектори є конкатенатами передньої нейромережі та назад нейтральної мережі. Для нейронної мережі вперед буде вилучено прихований стан після останнього символу слова (тобто "n"). Прихований стан перед першим символом слова (тобто "W") буде витягнутий із зворотної нейронної мережі.

## 2.4 Визначення факторів і відгуків

Проведення дослідження перелічених сервісів аналізу природної мови має на меті визначення ефективності сучасних сервісів в умовах реальних даних. Проводити порівняння було вирішено на типовому для подібних систем наборі даних із бізнес середовища – наборі відгуків. Так як досліджувані системи головним чином на сьогодні найбільш поширені у англомовному середовищі, то для розгляду максимальної ефективності систем були відібрані тестові дані на англійській мові. В якості джерела відгуків необхідно обрати досить масштабний і широковідомий сервіс електронної комерції. Для того, аби дослідити роботу систем якнаймога краще, відібрані тестові дані мають складатися з кількох речень.

Необхідно відсортувати відгуки за наявною оцінкою від самих користувачів, та власною оцінкою проаналізованого тексту відгуків яка ґрунтується на загальному сенсі речення, наявністю слів які надають тонального забарвлення реченню. У результаті необхідно сформувані тестовий набір даних для аналізу тональності тексту який складається з двох вибірок, у кожній із яких міститься по 50 відгуків.

У якості відгуків необхідно створити єдиний формат результату для усіх сервісів, який буде складатися з окремої суми позитивних, нейтральних та негативних результатів перевірки.

Таким чином буде створено зручний набір даних для порівняння ефективності сервісів оцінки тональності тексту.

## 2.5 Планування експерименту

Для проведення експерименту необхідно створити програмне забезпечення на базі якого буде можливо виконати порівняння результатів тестування сервісів.

Тестові дані, які будуть представлені відібраними відгуками користувачів обраного сайту електронної комерції, мають бути збережені у єдиному форматі, для створення можливості використання файлу для кожного з сервісів без модифікації форми зберігання даних.

Необхідно розбити проведення тестування на 3 етапи:

- оцінка змішаних тестових даних;
- оцінка позитивних тестових даних;
- оцінка негативних тестових даних.

Проводити дослідження слід почергово для кожного обраного сервісу на обраному наборі даних.

Отримані результати необхідно записати у порівняльні таблиці, за якими можна буде проводити подальший детальний аналіз.

Таким чином, відповідно до етапів тестування, має бути створено 3 типи таблиць.

Після завершення формування таблиці результатів, створених на змішаних даних, необхідно провести дослідження крайних показників кожного типу настрою: виявити сервіс, який поставив більше за все позитивних оцінок, той, який виявив найбільшу кількість негативних оцінок, і нейтральних. Також необхідно визначити сервіс, оцінки якого є найбільш наближеними до початкових оцінок для тестового набору даних.

На основі результатів зібраних у таблицях позитивних та негативних тестових даних необхідно провести детальний аналіз роботи сервісів. Виявити усі виключні ситуації для кожного досліджуваного сервісу, та провести детальний аналіз аномалій.

Аналіз виключних ситуацій дозволить отримати розуміння основних механізмів оцінки сервісів і типові помилки, які може допускати обраний сервіс.

На основі усіх досліджених даних необхідно буде сформувати загальний висновок за ефективністю сучасних сервісів, та типових недоліків, які можуть бути виявлені.

## **2.6 Програмні засоби для проведення експерименту**

У якості середовища розробки програмного забезпечення для проведення дослідження ефективності сервісів оцінки тональності тексту було обрано Microsoft Visual Studio. Microsoft Visual Studio - лінійка продуктів компанії Microsoft, що включають інтегроване середовище розробки програмного забезпечення і ряд інших інструментальних засобів. Дані продукти дозволяють розробляти як консольні додатки, так і ігри та програми з графічним інтерфейсом, в тому числі з підтримкою технології Windows Forms, а також веб-сайти, веб-додатки, веб-служби як в рідному, так і в керованому кодах для

всіх платформ, підтримуваних Windows, Windows Mobile, Windows CE, .NET Framework, Xbox, Windows Phone .NET Compact Framework і Silverlight.

Visual Studio включає в себе редактор вихідного коду з підтримкою технології IntelliSense і можливістю найпростішого рефакторінга коду. Вбудований відладчик може працювати як відладчик рівня вихідного коду, так і відладчик машинного рівня. Решта вбудовуються інструменти включають в себе редактор форм для спрощення створення графічного інтерфейсу додатку, веб-редактор, дизайнер класів і дизайнер схеми бази даних. Visual Studio дозволяє створювати і підключати сторонні додатки (плагіни) для розширення функціональності практично на кожному рівні, включаючи додавання підтримки систем контролю версій вихідного коду (як, наприклад, Subversion і Visual SourceSafe), додавання нових наборів інструментів (наприклад, для редагування і візуального проектування коду на предметно-орієнтованих мовах програмування) або інструментів для інших аспектів процесу розробки програмного забезпечення (наприклад, клієнт Team Explorer для роботи з Team Foundation Server). [27].

Мовою для розробки програмного забезпечення було обрано C#. C # (вимовляється Сі Шарп) - об'єктно-орієнтована мова програмування. Розроблено в 1998-2001 роках групою інженерів компанії Microsoft під керівництвом Андерса Хейлсберг і Скотта Вільтаумота як мову розробки додатків для платформи Microsoft .NET Framework. Згодом був стандартизований як ECMA-334 і ISO / IEC 23270.

C # відноситься до сім'ї мов з C-подібним синтаксисом, з них його синтаксис найбільш близький до C ++ і Java. Мова має статичну типізацію, підтримує поліморфізм, перевантаження операторів (в тому числі операторів явного і неявного приведення типу), делегати, атрибути, події, змінні, властивості, узагальнені типи і методи, ітератори, анонімні функції з підтримкою замикань, LINQ, виключення, коментарі в форматі XML.

Переїнявши багато від своїх попередників - мов C ++, Delphi, Модула, Smalltalk і, особливо, Java - C #, спираючись на практику їх використання, виключає деякі моделі, що зарекомендували себе як проблематичні при розробці програмних систем, наприклад, C # на відміну від C ++ не підтримує множинне успадкування класів (між тим допускається множинна реалізація інтерфейсів) [28].

Для проведення дослідження ефективності сервісів оцінки тональності текстів необхідно та достатньо виконати розробки консольного програмного забезпечення.

Воно має зчитувати підготовлені дані з файлу у форматі JSON та перетворювати їх у формат, необхідний для роботи з обраними API інтерфейсами.

## **2.7 Висновок до розділу 2**

У ході роботи було проаналізовано та визначено метод для порівняння ефективності роботи досліджуваних методів. Було обґрунтовано вибір факторів та відгуків на експеримент з визначення ефективності сервісів тональності текстів. Було сформовано вимоги до програмного забезпечення для проведення запланованого тестування, а також визначено особливості інструментальних та апаратних засобів.

### **3 АНАЛІЗ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ СЕРВІСІВ ОЦІНКИ ТОНАЛЬНОСТІ ТЕКСТУ ТА НАДАННЯ ПРАКТИЧНИХ РЕКОМЕНДАЦІЙ**

#### **3.1 Формування цілей і постановка задачі для розробки програмного забезпечення**

Головною ціллю розробки програмного забезпечення для порівняння сервісів визначення тональності тексту є створення необхідних інструментальних умов за яких можливе проведення тестування на одноманітному наборі тестових даних.

Для проведення експерименту було прийнято рішення розробити програмне забезпечення на базі якого можливо використати сервіси з обробки тональності тексту і у результаті їх роботи – порівняти їх оцінки, та сформувавши загальний висновок з їх ефективності, дослідити можливі нестандартні ситуації.

Програмне забезпечення для проведення експерименту повинно мати чіткий обмежений набір завдань. Головним завданням є надання повного обсягу необхідних результатів, які потрібні для аналізу досліджуваних сервісів. Воно має бути максимально простим у проектуванні та розробці, і відповідати потребам головного завдання. Для визначення формату розроблюваного ПЗ для проведення експерименту, розглянемо технології, які буде необхідно використати під час розробки.

#### **3.2 Огляд технологій і визначення формату для розробки програмного забезпечення**

API сервіси зручні тим, що не мають залежності від типу платформи, на основі якої данні сервіси планують використовувати. Прикладний програмний інтерфейс (англ. Application Programming Interface, скорочено API) — це сукупність засобів та правил, що вможливають взаємодію між окремими складниками програмного забезпечення або між програмним та апаратним забезпеченням [29].

У царині веб-розробки поняття API охоплює низку засобів програмного коду (як-от методи, події та посилання), що їх розробник може використовувати у власних застосунках задля взаємодії з програмним чи апаратним навкіллям — зі складниками веб-переглядача, іншим програмним чи апаратним забезпеченням комп'ютера користувача, або й навіть сторонніми сайтами чи службами.



Сервіси API розроблюються за стандартами, які дозволяють комунікувати з ними за обмеженим набором правил. Найпоширенішим стандартом є REST. Передача репрезентативного стану (англ. Representational State Transfer, REST) - це набір архітектурних обмежень, які утворюють ефективні, надійні та масштабовані розподілені системи. Система називається RESTful, коли дотримується цих обмежень [30].

Головна ідея REST полягає в тому, що ресурс, наприклад, документ, передається разом зі своїм станом та зв'язками (гіпертекстом) за допомогою чітко визначених, стандартизованих операцій та форматів. Часто API або сервіси називають себе RESTful, коли прямо змінюють тип документа, як протилежність запуску команд в іншому місці.

REST API заснований на URI (уніфікований ідентифікатор ресурсу) та протоколі HTTP. REST API можуть обмінюватися даними у форматі JSON або XML, хоча багато API REST надсилають дані у JSON [31].

URI - послідовність символів, що ідентифікує фізичний або абстрактний ресурс, який не обов'язково повинен бути доступний через мережу Інтернет, причому, тип ресурсу, до якого буде отриманий доступ, визначається контекстом і / або механізмом [32].

Наприклад, перейшовши по `http://example.com` - ми потрапимо на http-сервер ресурсу ідентифікованого як `example.com`.

HTTP - це протокол, що дозволяє отримувати різні ресурси, наприклад HTML-документи. Протокол HTTP лежить в основі обміну даними в Інтернеті. HTTP є протоколом клієнт-серверного взаємодії, що означає ініціацію запитів для серверу самим одержувачем, як правило, веб-браузером. Отриманий підсумковий документ буде (може) складатися з різних піддокументами є частиною підсумкового документа: наприклад, з окремо отриманого тексту, опису структури документа, зображень, відео-файлів, скриптів і багато чого іншого (рис. 3.1) [33].

Важливо знати, що запит складається з чотирьох речей:

- кінцева точка (endpoint);
- метод (GET, POST, PUT, або DELETE);
- заголовки (HTTP заголовки, повний список тут);
- дані (або тіло).

Для використання розглянутих технологій та дотримання основної мети проектування необхідно і достатньо розробити консольне ПЗ.

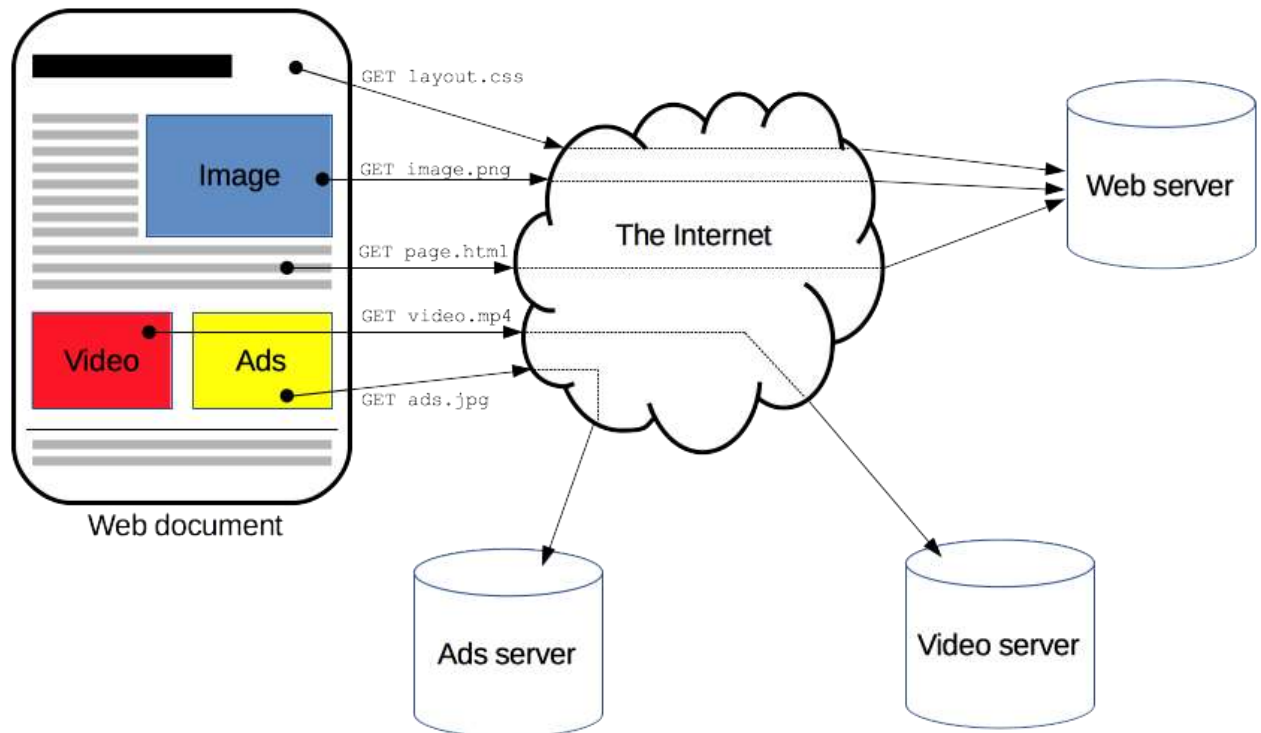


Рисунок 3.1 – Взаємодія за протоколом HTTP

### 3.3 Побудова діаграми варіантів використання

Діаграми варіантів використання призначені для опису функціонального призначення і поведінки програми, яка дозволяє розробникам, замовнику та кінцевим користувачам спільно вирішувати, що повинна робити система, яка проектується, або уже існує [34]. Вона відображає послідовність тих дій, що мають бути виконаними при взаємодії з зазначеним актором.

Основне призначення діаграми наступне:

Основне призначення діаграми наступне:

- визначити загальні межі і контекст предметної області;
- відділити систему від оточення;
- створити базову концептуальну модель для її подальшої реалізації у якості логічних і фізичних моделей.

Головними складовими діаграми варіантів використання є:

- актори;
- прецеденти.

Розроблена діаграма відображає взаємодію акторів або сутностей із системою, які діють за встановленими варіантами використання. В данному випадку актором навизається дійова особа, яка впливає на систему ззовні. Це може бути як людина, так і обчислювальна система. Варіанти використання це

ті ресурси, яка система надає актору. Між акторами встановлюються зв'язки – прецеденти. Вони відображають ті дії які виконуються системою.

Наша система слугує лише для виконання експерименту. У ній є один актор – користувач ПЗ. Розглянемо функції які може використовувати користувач ПЗ (рис. 3.2):

- завантажувати тестові дані;
- проводити тестування;
- переглядати результати тестів.

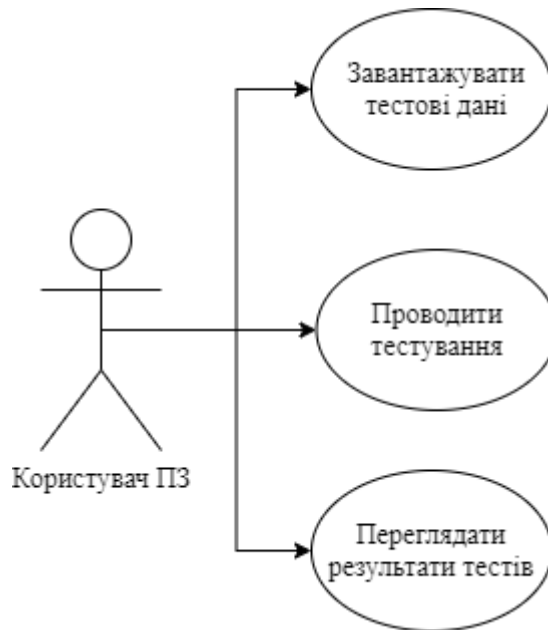


Рисунок 3.2 – Діаграма використання для актора Користува ПЗ

### 3.3 Вимоги до ПО

#### 3.3.1 Функціональні вимоги

Функціональні вимоги до розроблювального ПЗ наступні.

3.3.1.1 Користувач ПЗ повинен мати можливість керувати наступним функціоналом:

- завантажувати тестові дані;
- проводити тестування;
- надавати результати тестів.

### 3.3.2 Нефункціональні вимоги

До нефункціональних вимог відносяться такі:

3.3.2.1 Мінімальні системні вимоги для роботи програми наведені в таблиці 3.1;

Таблиця 3.1 – Конфігурація ПК для роботи програми на стороні клієнта

Операційна система	Microsoft Windows 8,10
Процесор	Pentium G5300 або краще
Об'єм оперативної пам'яті	1 ГБ або більше

3.3.2.2 Постійне стабільне підключення до мережі Інтернет зі швидкістю передачі даних 100мб/с;

3.3.2.3 Наявність монітора, клавіатури, комп'ютерної миші;

3.3.2.4 Форма відображення даних на екрані:

3.3.2.4.1 Вверху консолі повинно відображатись кількість досліджених відгуків користувачів;

3.3.2.4.2 Посередині мають наводитись тональні оцінки визначенні сервісом;

3.3.2.4.3 Внизу має наводитись докладний аналіз оцінки, за умови, що сервіс надає такі дані.

### 3.4 Специфікація варіантів використання

Специфікація варіантів використання містить основну інформацію про варіант використання: назва, учасники, контекст, передумова, постумова, основний та альтернативний (якщо такі наявні) потоки подій [34].

Специфікація варіантів використання виконана у вигляді таблиць з описом прецедентів і сценаріїв і наведена в таблицях 3.2 – 3.5.

Таблиця 3.2 – Сценарій початку обробки відгуків

<b>Ім'я</b>	Start_processing.
<b>Назва</b>	Початок обробки.
<b>Опис</b>	Початок обробки тестового набору даних відгуків.
<b>Передумова</b>	Додаток запущено, тестові данні відгуків завантажено у відповідний файл.

## Продовження таблиці 3.2

<b>Постумова</b>	Додаток починає надсилати данні на до сервісу з метою їх обробки
<b>Основний потік</b>	1 Користувач обирає сервіс для обробки. 2 Користувач підтверджує початок обробки.
<b>Альтернативний потік</b>	1 Користувач обирає сервіс для обробки. 2 Користувач підтверджує початок обробки. 3 Користувачеві виводиться попередження про те що сервіс не доступний

Таблиця 3.3 – Сценарій перегляду оцінок

<b>Ім'я</b>	Show_marks
<b>Назва</b>	Перегляд оцінок.
<b>Опис</b>	Користувач переглядає оцінки тональності оброблених відгуків.
<b>Передумова</b>	1 Користувач розпочав обробку даних. 2 Користувач дочекався успішного завершення обробки.
<b>Постумова</b>	Оцінки відображено на консолі.
<b>Основний потік</b>	1 Користувач переглядає наведені результати обробки

Таблиця 3.4 – Сценарій завантаження запису

<b>Ім'я</b>	Upload_data
<b>Назва</b>	Завантаження даних.
<b>Опис</b>	Користувач завантажує результати тестових даних у файл.
<b>Передумова</b>	1 Користувач успішно отримав результати оцінювання тестових даних.
<b>Потсумова</b>	1 Запис завантажено у файл.
<b>Основний потік</b>	1 Користувач обирає функцію завантаження результатів до файлу.

Таблиця 3.5 – Сценарій повторної обробки даних

<b>Ім'я</b>	Next_processing
<b>Назва</b>	Повторна обробка тестових даних
<b>Опис</b>	Користувач повторно оброблює результати.
<b>Передумова</b>	Користувач провів принаймні одну обробку тестових даних.

## Продовження таблиці 3.5

<b>Постумова</b>	Додаток починає надсилати данні на до сервісу з метою їх обробки
<b>Основний потік</b>	1 Користувач обирає сервіс для обробки. 2 Користувач підтверджує початок обробки.
<b>Альтернативний потік</b>	1 Користувач обирає сервіс для обробки. 2 Користувач підтверджує початок обробки. 3 Користувачеві виводиться попередження про те що сервіс не доступний

### 3.5 Огляд досліджуваних сервісів

Для проведення експерименту було відібрано 4 API сервіси з аналізу тональності тексту. Серед них:

- Text Analytics API від Microsoft;
- Sentiment Analysis від Tweenword;
- Sentiment Analysis від open copcy Text Processing;
- Natural Language API від Google.

Text Analytics API від Microsoft – набір веб-служб для аналізу текстової аналітики, який був створений на основі найкращих систем машинного навчання від Microsoft. Служби з набору Text Analytics API можна використовувати для аналізу неструктурованих текстів та виконання таких завдань, як:

- аналіз тональності тексту;
- пошук ключових фраз;
- визначення мови.

Для використання цього сервісу не потребуються навчальні дані, достатньо лише завантажити на обробку необхідні для дослідження дані. Sentiment Analysis від Tweenword представляє собою простий API сервіс, який може бути використаний для оцінки тональності тексту – позитивний він, або негативний. Даний сервіс має дві кінцеві точки – GET та POST для аналізу наданих користувачем текстових даних.

Sentiment Analysis від Text Processing – це простий сервіс JSON над HTTP для розуміння тексту та обробки природньої мови. Зараз він безкоштовний, та вільний для доступу без аутентифікації, проте розробник залишає за собою право змінити цінову політику у майбутньому. Наразі він пропонує користувачам наступну функціональність:

- Аналіз тональності тексту;
- Штampuвання;
- Фрагментація мовлення;
- Зв'язування іменованих сутностей.

Cloud Natural Language API від Google представляє собою API сервіс засобів обробки природньої мови та анотації тексту. Загалом система засобів може бути використана для комплексного аналізу та розуміння тестових даних.

Аналіз тональності тексту перевіряє і визначає переважаючу емоційну думку у тексті. На офіційному сайті з описом сервісу зазначається, що особливо корисною дана функція може бути для визначення ставлення письменника до висловлювання як позитивного, негативного або нейтрального.

Суб'єктний аналіз проводить перевірку тексту на наявність відомих сутностей. Наприклад, власні іменники, визначні пам'ятки, тощо.

Аналіз настроїв сутності перевіряє текст на наявність відомих сутностей та повертає інформацію про переважну думку сутностей у тексті. Може бути використано для визначення ставлення автора до сутності.

Синтаксичний аналіз витягує основну лексичну інформацію з тексту, розбиваючи текст на набір лексем і речень. Подібна робота може бути використана для подальшого аналізу текстової інформації.

Класифікація вмісту використовується для задач класифікації і повертає категорію контенту відповідно до змісту тексту.

Виклик кожного з перелічених Google API повертає також мову проаналізованого тексту, якщо вона не була попередньо вказана клієнтом на початку запиту.

Ці сервіси є доступними для загального користування, і можуть бути використані задля дослідницьких або бізнес цілей. Варто зауважити, що частина з них безкоштовно надає функціонал лише за умови певних обмежень продуктивності, об'єму оброблених даних за певний період часу та кількості дозволених запитів протягом одного місяця. Наведемо приклади тарифних планів сервісів, які використовувались у ході досліджень:

- Natural Language API від Google безкоштовно надає користувачам 1000 запитів на місяць, які можна здійснювати не частіше ніж один запит на секунду, вартість кожного наступного запиту складає 5 центів [35];
- Sentiment Analysis від Tweepword надає безкоштовно 500 запитів на місяць, без обмеження кількості запитів на секунду. Кожен наступний запит коштує 8 копійок [36].

Таким чином, якщо необхідно використовувати сервіси у бізнес цілях варто звернути увагу на безкоштовні сервіси, наприклад як досліджуваний Sentiment Analysis від open source Text Processing, або обрати наявний платний план відповідно до потреб бізнесу.

### **3.6 Початкові дані**

Проведення дослідження перелічених сервісів аналізу природньої мови мало на меті визначення ефективності сучасних сервісів в умовах реальних даних. Проводити порівняння було вирішено на типовому для подібних систем

наборі даних із бізнес середовища – наборі відгуків. Так як досліджувані системи головним чином на сьогодні найбільш поширені у англomовному середовищі, то для розгляду максимальної ефективності систем були відібрані тестові дані на англійській мові. В якості джерела відгуків необхідно було обрати досить масштабний і широковідомий сервіс електронної комерції. Для такої цілі було обрано онлайн магазин Amazon. Amazon – це американська компанія, одна з наймасштабніших світових організацій з продажу товарів та послуг через Інтернет. Для того, аби дослідити роботу систем якнаймога краще, відібрані тестові дані складаються з кількох речень.

Було знайдено кілька різних категорій популярних товарів з розділу електроніки. Під кожним з товарів було відібрано дві групи відгуків – по 5 позитивних, та по 5 негативних відгуків, на думку досліджувача. Ці відгуки були попередньо проаналізовані та відсортовані. Сортування проводилося за тональністю тестових відгуків користувачів. Було виділено наступні групи відгуків:

- позитивні;
- негативні.

Досліджувач сортував відгуки за наявною оцінкою від самих користувачів, та власною оцінкою проаналізованого тексту відгуків яка ґрунтувалася на загальному сенсі речення, наявністю слів які надають тонального забарвлення реченню. Таким чином, було сформовано тестовий набір даних для аналізу тональності тексту який складається з двох вибірок, у кожній із яких міститься по 50 відгуків.

Перед початком перевірки тестових відгуків сервісами з визначення тональності тексту, усі групи були перемішані в одну загальну групу у випадковій послідовності. Таким чином, після початкової підготовки, експеримент для кожного з названих сервісів проводився на тестових даних загальною кількістю 100 відгуків.

Створений сервіс у якості джерела тестових даних використовує документ у форматі JSON (рис. 3.3).

```
{
  "textList":[
    " text 1 ",
    " text 2 ",
    ...
    " text n "
  ]
}
```

Рисунок 3.3 – Схема формату зберігання тестових даних

Такий формат запису тестових даних дозволив зручно додавати та маніпулювати доданими даними у файлі, і водночас є простим для отримання даних досліджуваними сервісами для аналізу тональності текстових даних.



Він дозволив використовувати єдиний формат зберігання даних для усіх сервісів, хоча їх інтерфейси взаємодії відрізняються.

### 3.7 Тестування сервісів з аналізу тональності тексту

Тестування сервісів для аналізу тональності текстових даних проводилося на основі спроектованого та розробленого програмного забезпечення. Програмна реалізація дозволяє аналізувати тестові дані на кожному обраному сервісі окремо. Таким чином, було сплановано проведення почергового тестування сервісів на єдиному наборі даних з наступним аналізом отриманих результатів.

Під час проведення експериментів декілька разів відбувалися перебої у роботі сервісів. Найбільша кількість помилок спостерігалась у сервіса від Google. Його помилки були пов'язані із обмеженням ліміту обробки даних за одну секунду. Також спостерігалися помилки про перевищення кількості даних для обробки (рис. 3.4).

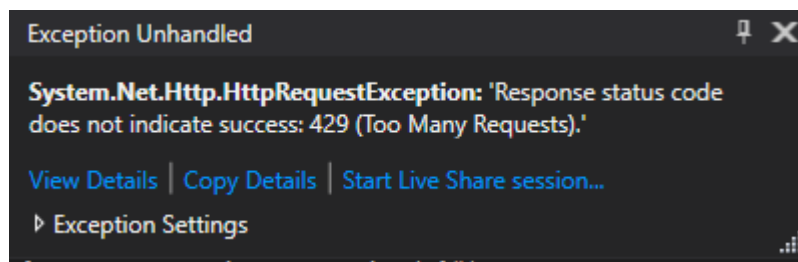


Рисунок 3.4 – Помилка про перевищення кількості запитів від API сервісу Google

Сервіс від Tweenword періодично переставав відповідати на запити, однак це носило тимчасовий характер.

Сервіси Sentiment Analysis від Text Processing, та Text Analytics API від Microsoft за весь час тестувань не створювали жодних виключних ситуацій під час обробки вищезазначених тестових даних.

### 3.8 Формат даних сервісів для оцінки тональності тексту

Розглядувані сервіси мають різний формат видачі даних. Вони різняться за інформативністю і змістом даних. Розглянемо детальніше формати відповідей кожного з API сервісів.

### 3.8.1 Формат даних сервісу Text Analytics API від Microsoft

Text Analytics API від Microsoft надає користувачу детальні результати тестування. Він виводить загальну оцінку (Sentiment) на основі проаналізованих речень тексту дня аналізу серед 4 доступних:

- positive (позитивна);
- negative (негативна);
- neutral (нейтральна);
- mixed (змішана).

Також для тексту наводяться загальні цифрові оцінки від 0.00 до 1.00 для кожної можливої тональності (ConfidenceScores): Positive (позитивної), Negative (негативної) та Neutral (нейтральної).

Більш детальні оцінки сервіс наводить далі для кожного окремо проаналізованого речення. Наводиться текст самого речення, загальна словесна оцінка одного з 4 варіантів, наведених вище, та цифрові оцінки від 0.00 до 1.00 для кожної можливої тональності (ConfidenceScores): Positive (позитивної), Negative (негативної) та Neutral (нейтральної). Приклад відповіді сервісу наведено на рис. 3.5.

```
Document sentiment: Mixed

Text: "I've only had this laptop about 2 1/2 weeks."
Sentence sentiment: Neutral
Positive score: 0.00
Negative score: 0.01
Neutral score: 0.99

Text: "I use it daily for about 4 hours, let it charge for 3-4 hours, then use it again until
Sentence sentiment: Neutral
Positive score: 0.00
Negative score: 0.00
Neutral score: 1.00

Text: "It looks like the battery is draining fast, but it really does last about 4 hours."
Sentence sentiment: Negative
Positive score: 0.25
Negative score: 0.74
Neutral score: 0.01
```

Рисунок 3.5 – Приклад виведення результатів сервісу Text Analytics API від Microsoft

Сервіс Text Analytics API від Microsoft надає також розширений метод для обробки даних. Метод AnalyzeSentimentBatch() з опцією AdditionalSentimentAnalyses.OpinionMining дозволяє отримувати додатковий параметер оцінки MinedOpinion для кожного речення, який відображає проаналізоване ключове слово речення, та пов'язані з ним слова, які характеризують данне слово, надаючи йому тонального забарвлення.

Воно надає текст слова, цифрову оцінку кожному тональному значенню, та текстову оцінку переважаючого значення. Для проаналізованого слова можливе позитивна, або негативна оцінки.

Пов'язані слова мають аналогічну структуру відображення даних, як і для основного слова. Частина розширеної відповіді сервісу наведено на рис. 3.6.

```
Aspect: food, Value: Negative
Aspect positive score: 0.01
Aspect negative score: 0.99
    Related Opinion: unacceptable, Value: Negative
    Related Opinion positive score: 0.01
    Related Opinion negative score: 0.99
```

Рисунок 3.6 – Розширена частина відповіді сервісу Text Analytics API від Microsoft

Задля зручності порівняння ефективності роботи сервісів, розширена версія сервісу не використовувалася під час дослідження.

### 3.8.2 Формат даних сервісу Sentiment Analysis від Tweenword

Sentiment Analysis від Tweenword надає загальну характеристику речення та наводить результати, на основі чого була сформована оцінка наданому тестовому реченню.

Загальний тестовий результат містить тип тексту, що є його словесною оцінкою, з 3 можливих:

- positive (позитивна);
- negative (негативна);
- neutral (нейтральна).

Загальний рахунок (Score) розраховується як сума рахунків усіх ключових слів, поділена на їх кількість. Може бути у діапазоні від -1.00 до 1.00. Після цього йде співвідношення (Ratio).

Після загальної оцінки наводиться список ключових слів (Word) з оцінкою (Score) для кожного з них, та може бути у діапазоні від -1.00 до 1.00. Приклад відповіді сервісу наведено на рис. 3.7.

```
Type: neutral
Score: -0.013464575742989432
Ratio: -0.028633679800640133
Keywords:
  Word: like
  Score: 1.70868868
  Word: horrible
  Score: 0.956923508
  Word: durable
  Score: 0.881294943
```

Рисунок 3.7 - Приклад виведення результатів сервісу Sentiment Analysis від Tweenword

### 3.8.3 Формат даних сервісу Sentiment Analysis від Text Processing

Серед обраних сервісів для оцінки тональності тексту, сервіс Sentiment Analysis від Text Processing надає найменшу кількість додаткової інформації про свою роботу.

Загальний результат наводиться у вигляді текстової оцінки (label). Загальна словесна оцінка одного з 3 можливих варіантів, а саме:

- pos (позитивна);
- neg (негативна);
- neutral (нейтральна).

Далі йдуть цифрові дані оцінки отриманого тексту у цілому. Цифрові оцінки йдуть у межах від 0.00 до 1.00 для кожної можливої тональності: pos (позитивної), neg (негативної) та neutral (нейтральної). Приклад відповіді сервісу наведено на рис. 3.8.

```
Score: {
  "probability": {
    "neg": 0.4854993494217552,
    "neutral": 0.16078613039366044,
    "pos": 0.5145006505782448
  },
  "label": "pos"
}
```

Рисунок 3.8 – Приклад виведення результатів сервісу Sentiment Analysis від Text Processing

### 3.8.4 Формат даних сервісу Natural Language API від Google

Сервіс Natural Language API від Google пропонує користувачеві як загальну так і детальну оцінку наданого для тонального аналізу тексту. Загальна оцінка складається з сумарної магнітуди (Magnitude) (не від'ємне число), та сумарного результату (Score), з діапазону від -1.0 до 1.0, для досліджених речень тексту. Сервіс не наводить словесну оцінку тексту, тож користувач має можливість сам ставити тональну оцінку на основі цифрових даних, з тими межами, які вважає за необхідне.

У детальній оцінці наводиться текст самого речення, а також його особисті магнітуду (Magnitude), та результату (Score), які є рівними за цифровим значенням і знаходяться у діапазоні від -1.0 до 1.0, однак результат може набувати позитивного, або негативного знаку в залежності від відповідного тонального настрою обраного тексту.

Загальна магнітуда тексту рахується як сума магнітуд усіх речень. Загальний результат представляє собою сумму результатів усіх речень, поділений на їх кількість. Приклад відповіді сервісу наведено на рис. 3.9.

```
Magnitude: 1.9
Score: 0.6
  Sentence: I have only had it a week, obviously but love it so far.
  Magnitude: 0.9
  Score: 0.9
  Sentence: It is lighter and slightly smaller.
  Magnitude: 0
  Score: 0
  Sentence: It fits well in my hand and I love it and it is my 6th kindle
  Magnitude: 0.9
  Score: 0.9
```

Рисунок 3.9 – Приклад виведення результатів сервісу Natural Language API від Google

### 3.9 Результати тестування сервісів оцінки тональності тексту

У ході дослідження ефективності сервісів з оцінки тональності тексту всі досліджувані сервіси були протестовані на спільних тестових даних. Заради зручності дослідження, спочатку було протестовано системи на загальних даних, а після цього окремо на наборі позитивних та негативних відгуків, які були відібрані досліджувачем.

Таким чином, було сформовано наступні результати тестів. Перевірка загального набору текстів дала наступні дані, які наведені у таблиці 3.6:

Таблиця 3.6 – Порівняння результатів оцінки загального набору даних

Тип тональності	Microsoft Text Analytics API	Tweenword Sentiment Analysis	Text Processing Sentiment Analysis	Google Natural Language API
Positive	48	54	20	42
Negative	52	24	60	46
Neutral	0	22	20	12
Mixed*	(72)	-	-	-

Слід зазначити, що тональна оцінка для Google Natural Language API була сформована власне автором дослідження на основі цифрової оцінки, яку надає сервіс. У випадку, коли загальна оцінка досліджуваного тексту більше 0, тексту ставиться позитивна оцінка (Positive). Коли оцінка менше за 0, текст оцінюється як негативний (Negative). Коли оцінка дорівнює 0, відповідно, значення прирівнюється до нейтрального (Neutral).

Змішана оцінка (Mixed) була надана лише сервісом Microsoft Text Analytics API. Автором дослідження було прийнято рішення не наводити змішану оцінку для Microsoft Text Analytics API, бо вона не дає змоги зрозуміти загальну тональну оцінку досліджуваного тексту. Таким чином, словесна оцінка для даного сервісу була сформована автором дослідження. У випадку, коли цифрове значення однієї з 3х можливих цифрових оцінок більше, ніж кожна інша окремо взята оцінка, то тексту присвоюється відповідна текстова оцінка. Іншим позитивним явищем такого способу оцінення стала одноманітність тестових результатів та підвищення точності порівняння.

Після цього було проведено тестування на наборі позитивних відгуків. Результат порівняння результатів позитивних відгуків наведено у таблиці 3.7:

Таблиця 3.7 – Порівняння результатів оцінки позитивного набору даних

Тип тональності	Microsoft Text Analytics API	Tweenword Sentiment Analysis	Text Processing Sentiment Analysis	Google Natural Language API
Positive	40	38	20	40
Negative	10	4	24	2
Neutral	0	8	6	8
Mixed*	(38)	-	-	-

Останнім було проведено тестування сервісів на негативному наборі відгуків. Результати цього тестування наведено у таблиці 3.8:

Таблиця 3.8 – Порівняння результатів оцінки негативного набору даних

Тип тональності	Microsoft Text Analytics API	Tweenword Sentiment Analysis	Text Processing Sentiment Analysis	Google Natural Language API
Positive	8	16	0	2
Negative	42	20	36	44
Neutral	0	14	14	4
Mixed*	(34)	-	-	-

### 3.10 Аналіз результатів дослідження сервісів оцінки тональності тексту

Після проведення тестування отримані дані було проаналізовано з метою визначення характерних особливостей для кожного з досліджуваних сервісів, значних відхилень від початкових оцінок та виключних ситуацій. Було проведено як порівняльний аналіз усіх сервісів разом, так і дослідження результатів кожного сервісу окремо.

#### 3.10.1 Аналіз результатів сервісу Microsoft Text Analytics API

Дослідження сервісу Microsoft Text Analytics API дали цікаві результати. Цей сервіс єдиний із усіх досліджуваних сервісів мав додатковий 4й тип текстової оцінки речень – змішаний (Mixed) тип.

Сервіс Microsoft Text Analytics API ставить дослідженому тексту оцінку Змішаний (Mixed) у тому випадку, коли у тексті наявні як позитивні, так і негативні речення. Нейтральні речення також можуть бути у наявності.

У текстових даних, які були використані для дослідження, використовуються відгуки до товарів з розділу електроніки. Основний зміст розглянутих відгуків стосується висловлювань людей на рахунок позитивних та негативних сторін придбаних товарів. Відгуки складаються з декількох речень, також наявні відгуки у яких є більше 10 речень.

На основі результатів тестування можна прийти висновку, що використовувати тестову оцінку сервісу Microsoft Text Analytics API, як основну є некоректним підходом. Адже у випадку наявності у тексті речень протилежної тональності буде надана текстова оцінка Mixed незалежно від цифрової оцінки сервісу. У результаті, за словесними оцінками сервісу, серед розглянутих 100 речень, 72 речення були оцінені як змішані (Mixed).

Грунтуючись на даних висновках було прийнято рішення під час проведення дослідження використовувати цифрові оцінки сервісу, які надаються для кожного можливого тонального значення (позитивна, негативна, нейтральна). Таким чином, словесна оцінка формується на підставі переважаючої цифрової оцінки.

За умови такого підходу сервіс почав надавати оцінки близькі за значенням до оригінальних оцінок.

Сервіс визначив позитивними 40 відгуків з 50. Під час оцінки негативних відгуків, було визначено 42 з 50 речень. Сформовані результати не містили нейтральних оцінок.

Розглянемо текст, оцінка якого відрізняється від початкової:

1) I took these up the El Capitan in Yosemite climbing. 2) The third day we had to replace the batteries (so if communication is critical, I would replace them every two days). 3) Around '1200 I dropped mine on accident and someone at the base found it the next day and radioed us, and he found my car so it could be returned to me. 4) Apparently they are durable considering it survived a quarter mile fall with little more than a scratch. 5) I like the simple lay out with the buttons, and the buttons are not horribly sensitive like I have experienced with other radios. 6) I found them to be very effective for climbing.

У даному випадку система оцінила речення номер 2 і 3 як негативні. Позитивну оцінку поставили лише реченню номер 5. Решта речень були оцінені нейтральними.

Якщо ми поглянемо на речення номер 2, то побачимо, що воно має нейтральний характер та використовується лише для опису ситуації. Однак, через наявність словосполучення «комунікація критична», сервіс визначив значення слова «критична» як негативне, а не як синонім слова «важлива». Помилка відбулась через помилку при аналізі ключових слів.

Таким чином, при загальному нейтральному настрої тексту, через помилку у аналізі сервіс поставив йому негативну оцінку. Помилки такого роду одна з проблем сучасних сервісів з обробки природньої мови. Роботи над усуненням подібних помилок активно проводяться у різних дослідницьких центрах.

Загальний аналіз сервісу доволі зрозумілий та логічний.



### 3.10.2 Аналіз результатів сервісу Tweenword Sentiment Analysis

Дослідження сервісу Tweenword Sentiment Analysis дозволило зрозуміти, що його оцінки більше орієнтовані на позитивні результати. Під час дослідження загальної кількості тестових даних зі 100 відгуків сервіс оцінив позитивними 54 відгука, а негативними лише 24.

Якщо розглянути ситуацію оцінювання детальніше, то можна побачити, що під час оцінювання позитивних відгуків сервіс оцінив позитивними 38 відгуків з 50, 8 відніс до нейтральної категорії і 4 відгука до негативної категорії. Однак під час оцінки негативних відгуків точність, у порівнянні з початковою оцінкою, стала нижче ніж під час оцінки позитивних результатів. Негативними було оцінено лише 20 відгуків з 50. Проте 16 відгуків з набору негативних отримали позитивну оцінку і 14 – нейтральну.

Для того, щоб дослідити методологію оцінки текстів за допомогою сервісу Tweenword Sentiment Analysis, розглянемо один з виключних прикладів детальніше. Для дослідження відберемо негативний відгук, який був оцінений сервісом як позитивний:

I am returning these. I'm sure they're great if your kids just want to use them around the house-- with each other. However, we bought them to use with a few other kids in the neighborhood. The description says, Good up to 16 miles. So I thought I was good to go. What it fails to mention in the Amazon description, however, is that it's 16 miles from mountain to valley, and only good up to 1/2 mile in the neighborhood. The box very clearly states that, but the online description is misleading.

Цей відгук є негативним за своєю природою і несе у собі критичний відгук. Сервіс оцінив його як позитивний. Розглянемо детальніше аргументацію, яку надає сервіс до власної оцінки (рис 3.10).

Type: positive	Word: return
Score: 0.15651759683673472	Score: 0.128641593
Ratio: 0.46893234353156843	Word: around
Keywords:	Score: 0.117922934
Word: good	Word: fail
Score: 2.830162047	Score: -0.726523694
Word: clearly	Word: misleading
Score: 1.110796512	Score: -0.666762836
Word: great	Word: other
Score: 0.797954407	Score: -0.287832858
Word: up	Word: description
Score: 0.28003636	Score: -0.214599066
Word: use	Word: kid
Score: 0.26849426	Score: -0.143066044
Word: mention	Word: neighborhood
Score: 0.238216828	Score: -0.143066044
Word: very	Word: only
Score: 0.192192596	Score: -0.096144741
Word: go	Word: sure
Score: 0.152059727	Score: -0.056871358

Рисунок 3.10 – Аналіз оцінки від сервісу Tweenword Sentiment Analysis

Як можна побачити, основний акцент зроблений на загальну кількість ключових слів, представлених у досліджуваному тексті. Проте, якщо проаналізувати представлений список слів, можна побачити одну важливу особливість. Деякі слова проаналізовані у некоректному контексті. Так, англійське слово «kid» має кілька перекладів, серед яких нейтральні іменники, як «дитина», так і негативні дієслова, наприклад «висміювати», «обманювати». У досліджуваному контексті мова йде про дітей, які користувалися товаром із відгуки. У контексті речень, у яких згадуються слова «діти», також не міститься негативних слів.

Таким чином, можна дійти висновку, що сервіс не проводить аналіз речення на його зміст, а аналізує лише ключові слова не зважаючи на їх контекст. Іншим припущенням може бути некоректний аналіз слів.

На сьогоднішній день доволі активно розвиваються системи які розглядають слова не як окрему сутність, а як частину тексту, разом з контекстом. Такі системи показують більш точні результати оцінок, завдяки розпізнаванню сенсу тексту.

Аналіз сервісу містить помітні помилки, тож точність оцінок, зроблених сервісом Tweenword Sentiment Analysis може бути взята під сумнів, у контексті досліджуваного набору даних.

### 3.10.3 Аналіз результатів сервісу Text Processing Sentiment Analysis

Сервіс Text Processing Sentiment Analysis з наявних 100 тестових даних оцінив позитивними лише 20 відгуків у той час як негативними – 60. Це є доволі цікавим випадком і відрізняється від характеру оцінок попередньо розглянутого сервісу у протилежний бік.

Щоб дослідити детальніше особливості оцінки сервісу, розглянемо його оцінки для позитивного та негативного набору даних окремо.

Під час аналізу набору позитивних відгуків сервіс оцінив позитивними лише 20 відгуків. У той час, 24 відгуки з 50 позитивних, сервіс оцінив негативними. Решта 6 відгуків, відповідно, отримали нейтральні оцінки.

Аналіз негативних результатів приніс зовсім інші результати. З 50 негативних відгуків жоден не отримав позитивної оцінки. 36 відгуків отримали негативну оцінку і 14 – нейтральну.

Нажаль, сервіс не надає детальних пояснень до своїх оцінок. Тож не має можливості проаналізувати спосіб оцінювання сервісу напряду. Спираючись на досліди проведені для інших сервісів, можна зробити припущення, що сервіс не проводить аналіз речення на його зміст, а аналізує лише ключові слова не зважаючи на їх контекст. Іншим припущенням можуть бути помилки під аналізу пов'язаних слів.

Використання методів для оцінки ключових слів разом з контекстом речень, теоретично могли б підвищити точність оцінювання для данного сервісу.

### 3.10.4 Аналіз результатів сервісу Natural Language API від Google

Natural Language API від Google у ході дослідження виявив результати, достатньо близькі до початкових. Зі 100 тестових даних 42 були оцінені як позитивні і 46 як негативні.

Більш детальні результати також демонструють високу точність оцінок. Під час дослідження позитивного набору даних, сервіс оцінив позитивними 40 відгуків, і лише 2 – негативними. При дослідженні негативного набору даних, 44 відгуків отримало негативну оцінку і 2 – позитивну.

Похибка оцінювань, у порівнянні із початковою оцінкою, знаходиться на дуже низькому рівні. Для того, аби упевнитись у коректності роботи сервісу, розглянемо виключні ситуації оцінювання.

Даний приклад демонструє позитивний відгук, який був оцінений сервісом як негативний:

I so miss my old clunky early model Kindle Fire. Reading on the Paperwhite is great but from the main screen and in the Kindle Store, I have trouble reading the black and white on the book covers. Also, I am very disappointed that customer

reviews are not viewable on the Kindle Store any longer. This means I will not be shopping for books on my Kindle, which means I will buy a lot fewer books over all. Not by choice but by lack of information and convenience. I hope this will change in the future. I'm still giving the sweet little Paperwhite 4 stars but am still learning all the features. EVERYTHING is different on it which was unexpected.

Досліджений відгук був віднесений автором дослідження до позитивних відгуків, бо автор відгуку вказав, що ставить позитивну оцінку обраному товару. Проте сам тест відгуку несе у собі основний об'єм критичних відгуків. Сервіс наводить оцінку для кожного речення проаналізованого тексту (рис. 3.11). Це дає нам змогу зрозуміти процес оцінки. Згідно з результатами даного оцінювання, можна погодитись, що відгук містить чимало негативних аспектів, тож він може бути оцінений як негативний.

```
Magnitude: 3
Score: -0.1
Sentence: I so miss my old clunky early model Kindle Fire.
Magnitude: 0
Score: 0
Sentence: Reading on the Paperwhite is great but from the main screen and in the Kindle Store, I have trouble reading
Magnitude: 0.4
Score: -0.4
Sentence: Also, I am very disappointed that customer reviews are not viewable on the Kindle Store any longer.
Magnitude: 0.8
Score: -0.8
Sentence: This means I will not be shopping for books on my Kindle, which means I will buy a lot fewer books over all.
Magnitude: 0.1
Score: -0.1
Sentence: Not by choice but by lack of information and convenience.
Magnitude: 0.6
Score: -0.6
Sentence: I hope this will change in the future.
Magnitude: 0.1
Score: -0.1
Sentence: I'm still giving the sweet little Paperwhite 4 stars but am still learning all the features.
Magnitude: 0.7
Score: 0.7
Sentence: EVERYTHING is different on it which was unexpected.
Magnitude: 0
Score: 0
```

Рисунок 3.11 – Аналіз речення сервісом Natural Language API від Google

Таким чином, Natural Language API від Google складає враження добре налаштованого сервісу, придатного до використання на текстових даних, подібних до досліджуваного набору.

### 3.10.5 Загальний порівняльний аналіз результатів сервісів оцінки тональності тексту

Зробимо загальний порівняльний огляд, як сервіси оцінили надані тестові дані. У данному розділі не буде аналізуватись кількість змішаних результатів, вони окремо були проаналізовані для сервісу Microsoft Text Analytics API.

Найбільше позитивних оцінок поставив сервіс Tweenword Sentiment Analysis. Він оцінив позитивними 54 відгука. Найменшу кількість позитивних

оцінок зробив Text Processing Sentiment Analysis, оцінивши позитивними лише 20 відгуків.

Якщо поглянути на негативні оцінки, то лідером став Text Processing Sentiment Analysis. Його результат – 60 негативних оцінок. Найменшу кількість негативних оцінок знову надав Tweenword Sentiment Analysis. Він оцінив негативними лише 24 з досліджених відгуків.

Кількість нейтральних оцінок досягає найбільших значень для Tweenword Sentiment Analysis – 22 результата. Microsoft Text Analytics API не оцінив жоден з досліджених тестових відгуків як нейтральний.

На розглянутому тестовому наборі даних, найбільш близькі до початкових результати продемонстрували сервіси Microsoft Text Analytics API та Natural Language API від Google. Аналіз результатів оцінок даних сервісів дає розуміння про процес оцінки даних сервісів. Вони аналізують зміст досліджуваних речень за їх сенсом. Це дозволяє проводити точне оцінювання, та зменшує кількість помилок.

У той же час у сервісів Tweenword Sentiment Analysis та Text Processing Sentiment Analysis оцінки для тестового набору суттєво відрізняються від початкових оцінок. Аналіз сервісу Tweenword Sentiment Analysis дозволяє зрозуміти, що сервіс не проводить оцінку речень за їх сенсом, а лише спирається на оцінку ключових слів. Такий піхід у сучасних сервісах є застарілим і точність оцінювання через це нижче ніж у конкурентів. Text Processing Sentiment Analysis не надає детального аналізу до своїх оцінок, проте дозволяє зробити припущення, що також робить оцінку текстових даних лише на основі ключових слів.

Таким чином, дослідження виявило той факт, що останні розробки у сфері обробки природньої мови забезпечують високі результати аналізу текстових даних. Проте сервіси, які використовують застарілі технології, поступаються у своїй ефективності.

### **3.11 Рекомендації щодо використання результатів дослідження**

У ході роботи було проаналізовано 4 сервіси для оцінки тональності тексту, які дослупні для використання широким загалом, та можуть бути використані для навчальних або бізнес цілей.

Тестування проводилось на реальних даних, які були взяті з відомого сайту електронної комерції Amazon. Дані представляють з себе відгуки, які були розділені на 2 категорії – позитивні та негативні. Загальна кількість тестових даних становить 100 відгуків.

Таким чином, дане дослідження можна вважати дослідженням у реальних умовах, його результати можуть бути корисні для власників сайтів, які прагнуть оптимізувати бізнес-процеси та підвищити ефективність свого бізнесу.

У результаті тестування було встановлено, що оцінка двох із досліджуваних сервісів досить близька до початкової. Результати аналізу оцінок підтвердили логічність оцінки тестових даних. У підсумку можна зазначити, що сервіси Natural Language API від Google та Microsoft Text Analytics API можуть бути використані для оцінки тональності відгуків з високою точністю результатів.

Решта два сервіси під час тестування виявили істотну розбіжність результатів з початковими оцінками. Аналіз оцінок сервісів виявив помилки в оцінках тональності речень. Грунтуючись на результатах оцінок, сервіси Tweenword Sentiment Analysis та Text Processing Sentiment Analysis можуть давати суттєві похибки під час оцінки тональності відгуків.

Варто взяти до уваги сам формат відгуків. Усі надані оцінки є коректними для наборів даних, схожих до тестового набору даних, який був досліджений у ході роботи. Не виключено, що точність оцінки даних іншого типу можуть відрізнятися для даних сервісів.

Під час вибору сервісу для оцінки тональності тексту, варто враховувати специфіку даних, та обов'язково перевіряти точність роботи обраних сервісів на власних даних.

Необхідно враховувати об'єми даних які необхідно обробити. Більшість існуючих сервісів є платними, та пропонують різні тарифні плани в залежності від кількості запитів за певний період часу.

Уважний підхід до вибору допоміжних сервісів оцінки тональності тексту може підвищити ефективність контролю та забезпечити розвиток для сайту.

### **3.12 Висновки до розділу 3**

У даному розділі наводяться дані про проектування програмного забезпечення, необхідного для дослідження сервісів з оцінки тональності тексту, наводяться результати проведеного дослідження та аналіз отриманих результатів. Наприкінці розділу наводяться практичні рекомендації з використання отриманих результатів.

Для проведення дослідження було спроектоване та розроблене консольне програмне забезпечення на мові програмування C#. Програмне забезпечення має інтерфейс для вибору одного з досліджуваних сервісів для оцінки тональності текстових даних. Текстові дані зберігаються у єдиних для усіх сервісів форматах, можуть легко змінюватися та редагуватися. Створено єдиний інтерфейс для відображення результатів роботи сервісів.

Дослідження було проведено почергово для усіх досліджуваних сервісів. Сервіси тестувалися на позитивному, негативному та об'єднаному наборах тестових даних. Було розглянуто моделі відповідей сервісів.

Отримані тестові дані були проаналізовані. Було проаналізовано загальні оцінки для кожного з сервісів, та розглянуто виключні результати. Дослідження

виключних ситуацій дозволило зрозуміти особливості формування оцінок сервісів. Було проведено загальне порівняння усіх сервісів за їх результатами.

На основі досліджених особливостей сервісів було сформовано практичні рекомендації, щодо використання подібних сервісів.

## ВИСНОВКИ

У ході роботи над атестаційною роботою магістра було досліджено ефективність сучасних комерційних сервісів з оцінки тональності тексту.

Проводиться аналіз сучасного стану проблеми оцінки тональності текстів, відбувається постановка задач для дипломної роботи з дослідження сервісів оцінки текстової інформації.

Для дослідження сучасного стану проблеми розглядаються напрями обробки природної мови та її підвиду оцінки тональності тексту. Проводиться огляд ряду задач, які вирішуються застосуванням сучасних сервісів, побудованих на алгоритмах зазначених вище технологій. На сьогоднішній день, системи активно використовуються для бізнес задач та забезпечують значне підвищення ефективності розвитку бізнесу.

Проводиться огляд основних критичних моментів, які негативно впливають на рівень точності сучасних систем оцінки тональності тексту. У цих напрямках активно проводяться дослідницькі роботи, однак, у повному обсязі дані проблеми досі лишаються не вирішеними.

Було обгрунтовано вибір факторів та відгуків на експеримент з визначення ефективності сервісів тональності текстів. Було сформовано вимоги до програмного забезпечення для проведення запланованого тестування, а також визначено особливості інструментальних та апаратних засобів.

Для проведення дослідження було спроектоване та розроблене консольне програмне забезпечення на мові програмування С#. Програмне забезпечення має інтерфейс для вибору одного з досліджуваних сервісів для оцінки тональності текстових даних. Текстові дані зберігаються у єдиних для усіх сервісів форматах, можуть легко змінюватися та редагуватися. Створено єдиний інтерфейс для відображення результатів роботи сервісів.

Дослідження було проведено почергово для усіх досліджуваних сервісів. Сервіси тестувалися на позитивному, негативному та об'єднаному наборах тестових даних. Було розглянуто моделі відповідей сервісів.

Отримані тестові дані були проаналізовані. Було проаналізовано загальні оцінки для кожного з сервісів, та розглянуто виключні результати. Дослідження виключних ситуацій дозволило зрозуміти особливості формування оцінок сервісів. Було проведено загальне порівняння усіх сервісів за їх результатами.

З'ясовано, що останні розробки дозволяють проводити оцінку з високим ступенем якості. Однак не усі представлені на ринку сервіси базуються на нових методиках аналізу, у результаті у їх оцінках спостерігаються значні похибки.

На основі досліджених особливостей сервісів було сформовано практичні рекомендації, щодо використання подібних сервісів.



## ПЕРЕЛІК ПОСИЛАНЬ

1. Аналіз тональності тексту [Електронний ресурс] – Режим доступу до ресурсу: [https://en.wikipedia.org/wiki/Sentiment\\_analysis](https://en.wikipedia.org/wiki/Sentiment_analysis)
2. Natural language processing, Towardsdatascience – [Електронний ресурс]. – Режим доступу: <https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1>
3. Natural language processing, Wikipedia – [Електронний ресурс]. – Режим доступу: [https://en.wikipedia.org/wiki/Natural\\_language\\_processing](https://en.wikipedia.org/wiki/Natural_language_processing)
4. Основы Natural language processing– [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/Voximplant/blog/446738/>
5. Natural language processing, Towardsdatascience – [Електронний ресурс]. – Режим доступу: <https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1>
6. 5 Use Cases for Natural Language Processing – [Електронний ресурс]. – Режим доступу: <https://customerthink.com/5-use-cases-for-natural-language-processing-application-in-marketing/>
7. How to Use Sentiment Analysis to Manage Your Brand Reputation & Win Customers – [Електронний ресурс]. – Режим доступу: <https://brandmentions.com/blog/sentiment-analysis/>
8. Sentiment Analysis in Marketing – [Електронний ресурс]. – Режим доступу: <https://www.cmswire.com/digital-marketing/sentiment-analysis-in-marketing-what-are-you-waiting-for/>
9. Marketing Research: The Role of Sentiment Analysis – [Електронний ресурс]. – Режим доступу: <http://wps.fep.up.pt/wps/wp489.pdf>
10. Sentiment Analysis: Types, Tools, and Use Cases – [Електронний ресурс]. – Режим доступу: <https://www.altexsoft.com/blog/business/sentiment-analysis-types-tools-and-use-cases/>
11. Sentiment Analysis of Product Reviews Cases – [Електронний ресурс]. – Режим доступу: <https://monkeylearn.com/blog/sentiment-analysis-of-product-reviews/#:~:text=Sentiment%20analysis%20is%20the%20automated,%3A%20Positive%2C%20Neutral%2C%20Negative.&text=Compare%20your%20product%20reviews%20with%20those%20of%20your%20competitors.>
12. Four Pitfalls of Sentiment Analysis Accuracy – [Електронний ресурс]. – Режим доступу: <https://www.toptal.com/deep-learning/4-sentiment-analysis-accuracy-traps>
13. How to: Sentiment analysis and Opinion Mining – [Електронний ресурс]. – Режим доступу: <https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/how-to/text-analytics-how-to-sentiment-analysis?tabs=version-3-1>
14. Sentiment Analysis – [Електронний ресурс]. – Режим доступу: <https://www.twinword.com/api/sentiment-analysis.php>
15. Interpreting The Score And Ratio Of Sentiment Analysis – [Електронний ресурс]. – Режим доступу: <https://www.twinword.com/blog/interpreting-the-score->

[and-ratio-of-sentiment/](#)

16. Uplab – [Електронний ресурс]. – Режим доступа: <https://www.uplab.ru/blog/artificial-intelligence/>
17. Wikipedia – [Електронний ресурс]. – Режим доступа: [https://en.wikipedia.org/wiki/Part-of-speech\\_tagging](https://en.wikipedia.org/wiki/Part-of-speech_tagging)
18. Tom Young. Recent Trends in Deep Learning Based Natural Language Processing, 2018. – 32с.
19. Medium – [Електронний ресурс]. – Режим доступа: <https://medium.com/@5hirish/dependency-parsing-in-nlp-d7ade014186>
20. Medium – [Електронний ресурс]. – Режим доступа: <https://medium.com/mysuperaai/what-is-named-entity-recognition-ner-and-how-can-i-use-it-2b68cf6f545d>
21. Wikipedia – [Електронний ресурс]. – Режим доступа: [https://en.wikipedia.org/wiki/Semantic\\_role\\_labeling](https://en.wikipedia.org/wiki/Semantic_role_labeling)
22. Monkeylearn – [Електронний ресурс]. – Режим доступа: <https://monkeylearn.com/sentiment-analysis/>
23. Towardsdatascience – [Електронний ресурс]. – Режим доступа: <https://towardsdatascience.com/machine-translation-a-short-overview-91343ff39c9f>
24. Medium – [Електронний ресурс]. – Режим доступа: <https://medium.com/lingvo-masino/question-and-answering-in-natural-language-processing-part-i-168f00291856>
25. Wikipedia – [Електронний ресурс]. – Режим доступа: [https://en.wikipedia.org/wiki/Dialogue\\_system](https://en.wikipedia.org/wiki/Dialogue_system)
26. Towardsdatascience – [Електронний ресурс]. – Режим доступа: <https://towardsdatascience.com/contextual-embeddings-for-nlp-sequence-labeling-9a92ba5a6cf0>
27. Microsoft Visual Studio – [Електронний ресурс]. – Режим доступа: [https://ru.wikipedia.org/wiki/Microsoft\\_Visual\\_Studio](https://ru.wikipedia.org/wiki/Microsoft_Visual_Studio)
28. C# – [Електронний ресурс]. – Режим доступа: [https://ru.wikipedia.org/wiki/C\\_Sharp](https://ru.wikipedia.org/wiki/C_Sharp)
29. API (Прикладний програмний інтерфейс) [Електронний ресурс] – Режим доступа до ресурсу: <https://developer.mozilla.org/uk/docs/Glossary/API>
30. Передача репрезентативного стану (REST) [Електронний ресурс] – Режим доступа до ресурсу: <https://developer.mozilla.org/uk/docs/Glossary/REST>
31. Понимание основ работы API и REST API [Електронний ресурс] – Режим доступа до ресурсу: <https://sebweo.com/ru/ponimanie-osnov-raboty-api-i-rest-api-kratkoe-vvedenie/>
32. URI [Електронний ресурс] – Режим доступа до ресурсу <https://habr.com/ru/post/232385/>
33. Обзор протокола HTTP [Електронний ресурс] – Режим доступа до ресурсу <https://developer.mozilla.org/ru/docs/Web/HTTP/Overview>
34. Діаграма варіантів використання : [Електронний ресурс] / Режим доступа: [www.URL:https://uk.wikipedia.org/wiki/Діаграма\\_прецедентів](http://www.URL:https://uk.wikipedia.org/wiki/Діаграма_прецедентів)

35. Специфікація варіантів використання: [Електронний ресурс] / Режим доступу: [http://life-prog.ru/ukr/1\\_4643\\_variant-vikoristannya.html](http://life-prog.ru/ukr/1_4643_variant-vikoristannya.html)
36. Google Text Analysis pricing: [Електронний ресурс] / Режим доступу: <https://rapidapi.com/insights-ml-insights-ml-default/api/google-text-analysis/pricing>
37. Twinword Text Analysis pricing: [Електронний ресурс] / Режим доступу: <https://rapidapi.com/twinword/api/sentiment-analysis/pricing>

## ДОДАТОК А

### Лістинг коду ПЗ для дослідження сервісів оцінки тональності тексту Program.cs

```

using System;
using TestSentimentalAnalysisAPI.Interfaces;
using TestSentimentalAnalysisAPI.Services;

namespace TestSentimentalAnalysisAPI
{
    class Program
    {
        static void Main(string[] args)
        {
            AnatysisTestData();

            public static void AnatysisTestData()
            {
                int selectedApiNumber;
                Console.WriteLine("Choose sentimental analytics API: \n 1 -
MicrosoftTextAnalyticsApi, " +
                    "\n 2 - GoogleCloudNaturalLanguageApi, \n 3 -
TwinwordSentimentAnalysisApi, " +
                    "\n 4 - TextProcessingApi \n");
                selectedApiNumber = int.Parse(Console.ReadLine().ToString());
                Console.WriteLine();

                IAnalysisText analysisService;

                switch (selectedApiNumber)
                {
                    case 1:
                        analysisService = new MicrosoftTextAnalyticsService();
                        analysisService.AnalysisText(); break;
                    case 2:
                        analysisService = new
GoogleCloudNaturalLanguageService();
                        analysisService.AnalysisText(); break;
                    case 3:
                        analysisService = new TwinwordTextService();
                        analysisService.AnalysisText(); break;
                    case 4:
                        analysisService = new TextProcessingService();
                        analysisService.AnalysisText(); break;
                    default: Console.WriteLine("Choose correct number");
                }
                break;

                Console.WriteLine("Do you want to test another API? Y-Yes, N-
No. Default - No \n");
                var choise = Console.ReadLine().ToString();
                if (choise == "Y" || choise == "y")
                    AnatysisTestData();
                Console.WriteLine();
            }
        }
    }
}

```

## TestDataReaderService.cs

```

using Newtonsoft.Json;
using System.IO;
using TestSentimentalAnalysisAPI.Models;

namespace TestSentimentalAnalysisAPI.Services
{
    public class TestDataReaderService
    {
        public static TextDataModel ReadTestData()
        {
            if
(File.Exists("C:\\PracticeProjects\\DiplomaProjects\\TestSentimentalAnalysisAPI\\
TestDataMicrosoftSentimentalApi.json"))
            {
                //string json =
File.ReadAllText("C:\\PracticeProjects\\DiplomaProjects\\TestSentimentalAnalysis
API\\TestDataMicrosoftSentimentalApi.json");
                //string json =
File.ReadAllText("C:\\PracticeProjects\\DiplomaProjects\\TestSentimentalAnalysis
API\\NegativeText.json");
                string json =
File.ReadAllText("C:\\PracticeProjects\\DiplomaProjects\\TestSentimentalAnalysis
API\\PositiveText.json");
                TextDataModel text =
JsonConvert.DeserializeObject<TextDataModel>(json);

                return text;
            }

            return new TextDataModel();
        }
    }
}

```

## MicrosoftTextAnalyticsService.cs

```

using Azure;
using Azure.AI.TextAnalytics;
using System;
using System.Collections.Generic;
using TestSentimentalAnalysisAPI.Interfaces;

namespace TestSentimentalAnalysisAPI.Services
{
    public class MicrosoftTextAnalyticsService: IAnalysisText
    {
        private static readonly AzureKeyCredential credentials = new
AzureKeyCredential("c38a39a0e15745fd87069132c70c844b");
        private static readonly Uri endpoint = new Uri("https://movie-
feedbacks.cognitiveservices.azure.com/");
        private int positiveCount = 0, negativeCount = 0, neutralCount =
0, mixedCount = 0;

        public void AnalysisText()
        {
            var client = new TextAnalyticsClient(endpoint, credentials);

            var textList = TestDataReaderService.ReadTestData();

```

```

        SentimentAnalysis(client, textList.TextList);

        Console.WriteLine("Press any key to exit.");
        Console.ReadKey();
    }

    private void SentimentAnalysis(TextAnalyticsClient client,
List<string> textList)
    {
        foreach (var text in textList)
        {
            DocumentSentiment documentSentiment =
client.AnalyzeSentiment(text);
            Console.WriteLine($"Document sentiment:
{documentSentiment.Sentiment}\n");
            Console.WriteLine($"Sentence confidence score:
{documentSentiment.ConfidenceScores.Positive}");
            Console.WriteLine($"Sentence confidence score:
{documentSentiment.ConfidenceScores.Negative}");
            Console.WriteLine($"Sentence confidence score:
{documentSentiment.ConfidenceScores.Neutral}");

            foreach (var sentence in documentSentiment.Sentences)
            {
                Console.WriteLine($"Text: \"{sentence.Text}\"");
                Console.WriteLine($"Sentence sentiment:
{sentence.Sentiment}");
                Console.WriteLine($"Positive score:
{sentence.ConfidenceScores.Positive:0.00}");
                Console.WriteLine($"Negative score:
{sentence.ConfidenceScores.Negative:0.00}");
                Console.WriteLine($"Neutral score:
{sentence.ConfidenceScores.Neutral:0.00}\n");
            }

            if (documentSentiment.ConfidenceScores.Positive >
documentSentiment.ConfidenceScores.Negative &&
documentSentiment.ConfidenceScores.Positive >
documentSentiment.ConfidenceScores.Neutral)
                positiveCount++;
            if (documentSentiment.ConfidenceScores.Negative >
documentSentiment.ConfidenceScores.Positive &&
documentSentiment.ConfidenceScores.Negative >
documentSentiment.ConfidenceScores.Neutral)
                negativeCount++;
            if (documentSentiment.ConfidenceScores.Neutral >
documentSentiment.ConfidenceScores.Positive &&
documentSentiment.ConfidenceScores.Neutral >
documentSentiment.ConfidenceScores.Negative)
                neutralCount++;
        }

        Console.WriteLine("***** Total result ***** \n");
        Console.WriteLine($"Positive sentences: {positiveCount*2}");
        Console.WriteLine($"Negative sentences: {negativeCount*2}");
        Console.WriteLine($"Neutral sentences: {neutralCount*2}");
        Console.WriteLine($"Mixed sentences: {mixedCount*2}");
    }
}
}

```

## TextDataModel.cs

```
using System.Collections.Generic;

namespace TestSentimentalAnalysisAPI.Models
{
    public class TextDataModel
    {
        public List<string> TextList { get; set; }
    }
}
```

## IAnalysisText.cs

```
using System;
using System.Collections.Generic;
using System.Text;

namespace TestSentimentalAnalysisAPI.Interfaces
{
    public interface IAnalysisText
    {
        void AnalysisText();
    }
}
```

## ДОДАТОК Б

### Екранні форми програмного забезпечення

```
Choose sentimental analytics API:  
1 - MicrosoftTextAnalyticsApi,  
2 - GoogleCloudNaturalLanguageApi,  
3 - TwinwordSentimentAnalysisApi,  
4 - TextProcessingApi
```

Рисунок Б.1 – Екранна форма вибору сервісу для перевірки

```
Document sentiment: Mixed  
  
Text: "I've only had this laptop about 2 1/2 weeks."  
Sentence sentiment: Neutral  
Positive score: 0.00  
Negative score: 0.01  
Neutral score: 0.99  
  
Text: "I use it daily for about 4 hours, let it charge for 3-4 hours, then use it again until  
Sentence sentiment: Neutral  
Positive score: 0.00  
Negative score: 0.00  
Neutral score: 1.00  
  
Text: "It looks like the battery is draining fast, but it really does last about 4 hours."  
Sentence sentiment: Negative  
Positive score: 0.25  
Negative score: 0.74  
Neutral score: 0.01
```

Рисунок Б.2 – Екранна форма виводу деталей дослідження

```
***** Total result *****  
  
Positive sentences: 20  
Negative sentences: 24  
Neutral sentences: 6
```

Рисунок Б.3 – Екранна форма виводу результату дослідження